

An Introduction to Metaphysics

Identity, Free Will, and Time

By David W. Agler

This is a copyright page for my handout packet.
www.davidagler.com

Copyright © 2026 by David W. Agler

All rights reserved. No part of this publication may be reproduced, distributed, or transmitted in any form or by any means, including photocopying, recording, or other electronic or mechanical methods, without the prior written permission of the author, except in the case of brief quotations embodied in critical reviews and certain other noncommercial uses permitted by copyright law.

Agler, David W., 1982- An Introduction to Metaphysics : time, identity, and free will / David W. Agler. pages cm Includes bibliographical references and index. ISBN 978-0-9000000-0-0 1. Metaphysics. I. Title.

Printed in the United States of America

1 2 3 4 5 6 7 8 9 10

To my daughter Renna and my wife Liz.

Contents

I	Introduction	
II	Objects and Identity	
1	What exists?	5
2	Does God exist?	17
3	Do ordinary objects exist?	73
4	Do persons exist?	75
5	How do we exist through time?	87
6	Are there intermittent objects?	175
III	Freedom	
7	Are we free?	187
8	Is free will possible?	207
9	Can we be free if we are determined?	229
10	Libertarianism	245
11	Why deliberate?	259
12	Should we punish criminals?	267
IV	Time	
13	What do we believe about time?	279
14	What is a theory of time?	285
15	Does only the present exist?	301
16	The Moving Spotlight Theory	325
17	Does time grow?	345
18	Do all times exist?	357
19	Is backward time travel possible?	403
V	Back Matter	
	Notes	417

PREFACE

0

WHAT THIS BOOK IS ABOUT?

The primary goal of this text is to be an introduction to metaphysics. Unlike other introductions to metaphysics that provide a broad overview of the discipline, the goal of this text is to introduce metaphysics through three topics: time, identity, and free will.

WHO IS THIS BOOK FOR?

This book is written with the non-philosophy student in mind. Because this is my target audience, there are a few points to keep in mind.

Point 1: There are places where the text could express things more precisely or accurately. In some places, I sacrifice both of these virtues for making either the theory or the arguments more understandable for my target audience.

Point 2: I try to avoid (or keep to a minimum) the use of logic throughout this text. Several texts introduce logical toolkits at the beginning of a text, but I have found that this is only suitable for readers with a preexisting technical background.

Point 3: Since my audience is a non-philosophy student, I sometimes will give up on the back-and-forth (point and counterpoint) too soon. This is largely because, at some point, further discussion of a theory requires reformulating the theory with more precision (point 1), introducing formal notation (point 2), or I fear the reader will grow weary with the theory and will want to move onto something more exciting.

THE STATE OF THIS TEXT

Currently, this text is draft form. In many places, it is in rough shape. If you are reading this and would like to suggest improvements to the text, or you just want to say hello, please feel free to contact me at dwa132@psu.edu or through my YouTube Channel <https://www.youtube.com/davidagler>.

SOME CREDITS

My initial interest in metaphysics was largely sparked by one person: Michael B. Burke. I took his course on metaphysics as an undergraduate majoring in English at IUPUI and then as a graduate student in Philosophy. My contin-

ued interest in the topic is largely due to the fact that I've had the opportunity to teach the course several times at The Pennsylvania State University.

Since this text is written in L^AT_EX, writing this book would not be possible without the many, many, many packages and people who develop and maintain packages and upload them to CTAN or who post answers to all of my questions on the T_EX StackExchange. Some notable ones include:

1. Peter R. Wilson who created and Lars Madsen who maintains the `memoir class`
2. Nils Fleischhacker who created `tikzpeople`
3. Emiliano Heyns who created the `Zotero Better Bibtex` extension
4. Brian Dunn who created the `lwarp` package.

Part I

Introduction

Part II

Objects and Identity

WHAT EXISTS?

1

1.1 INTRODUCTION

What exists and how do the things that exist relate to each other? Let's consider this question by proposing some general answers.

1.2 EVERYTHING EXISTS

One possible answer to the question of 'what kinds of things exist' is *everything*. What do we mean by *everything*? One formulation of the view that everything exists involves what is known as the 'principle of plenitude'. This is the principle that the universe contains every possible kind (or form) of being.

Definition 1.1: principle of plenitude (static form)

Every kind of thing (form of being) that can exist does exist.[73, p. 154]

It is important to note that the principle does not say that every possible object exists. For example, the fact that I do not have a twin brother (call him 'Mavid') does not show that the principle is false. This is because while the principle would entail that dogs, cats, humans, and dolphins exist since these are *kinds* of things, it would not imply that Mavid exists since *being David's twin* would not be a kind of a thing.

Exercise 1.1

Draw an object that represents a kind of thing that could exist but does not exist.

1.2.1 The principle of plenitude

At first glance, the principle of plenitude seems false. There are many kinds of things that do not exist. For example, there are dogs and dolphins, but there is no kind of thing that is a hybrid between a dog and a dolphin: a *dogphin*.

Since the principle of plenitude seems false at first glance, perhaps there is an argument that shows that our intuitions are false and that the principle is, in fact, true. Consider the following argument (attributed to Spinoza):[73, pp. 152–3]

- P1: Either an object (e.g., a triangle) necessarily exists or it is impossible.
- P2: A thing necessarily exists if there is no cause (or reason) that would make its existence impossible.
- C: Therefore, anything that is possible (can exist) necessarily exists.

Before evaluating the above argument, let's get a better sense of the meaning of both of the premises. To understanding P1, it is helpful to distinguish two different sentences:

P1 Objects necessarily exist or they necessarily do not exist.

P1* Objects necessarily exist or they do not necessarily exist.

It might not be obvious, but (P1) and (P1*) are different. P1 says things either must exist or must not exist. This implies that there are no objects whose existence is contingent. An object exists contingently if it exists but it might not have existed. For example, suppose you bought a new bike. If that bike's existence is contingent, then while it exists, it might not have existed. For example, it might not have been assembled.

In contrast P1* is consistent with objects existing contingently. It says that an object must exist or it not the case that it must exist, e.g., it can exist but it might not have existed.

P2 says that a triangle (or any object) necessarily exists unless its existence would be a contradiction. Since there is nothing contradictory about a triangle's existence, it is not impossible. Therefore, it necessarily exists. Similarly, since there is nothing contradictory about saying I exist, it is not impossible that I exist. Therefore, I exist.

Now that we have clarified both P1 and P2, let's consider some objections to the above argument.

First, P2 is false. There is no cause that prevents humans with horns on their head from existing. But, they don't exist. More generally, there are many things that could exist (there is nothing that would make their existence impossible), but do not exist.

Second, suppose P2 is true. There is nothing that would make the *continued existence* of the Dodo bird (a flightless bird native to Mauritius that went extinct around 1662) impossible. Therefore, according to P2, the Dodo bird exists. But, the Dodo bird does not exist. Therefore, P2 is false.

Similarly, consider the coywolf (also known as the "wolfote"). The coywolf is a hybrid mammal that descended from coyotes, eastern (and gray) wolves,

and dogs. It emerged as a new species roughly 100,000 years ago. But now consider the state of the universe 200,000 years ago. If P2 is true, then the coywolf exists 200,000 year ago since there is nothing impossible about it existing before it evolved. But, the coywolf did not exist 200,000 years ago. Therefore, P2 is false.

In general, P2 implies that objects never go out of existence via extinction and new objects never come into existence via evolution. Whatever exists, always exists. But, this is false given that science tells us that objects do go extinct and new species do evolve. Therefore, P2 is false.

Third, there are religious reasons to reject P1. Let's suppose God exists, is omnipotent, and P1 is true. If every kind of thing that can exist does exist, then God cannot create new kinds of things. Therefore, either God exists, or is not omnipotent, or the principle of plenitude is false. Someone who accepts God's existence and omnipotence has a reason to reject P1.

Fourth, there are moral reasons to reject the principle. If the principle were true and God created the world, then not only every good exists but every evil exists as well. If every evil exists and these evils are permanent, then there is no hope that these evils can be eradicated. But, there is hope that these evils can be eradicated, e.g., the removal of a cruel dictator. Therefore, the principle of plenitude is false.[73, p. 245]

Fifth, we saw that P1 entails that there are no contingent objects. Any object that exists does so necessarily. And, any object that does not exist, necessarily does not exist. One problem with P1 is that it is not self-evident (it is not intuitively true). Since it is not self-evident, someone who accepts P1 ought to provide a reason in support of it. Let's consider a possible reason to accept P1.

One reason to accept P1 is theological. Suppose that God necessarily exists, God created the universe, and the universe is rational. If the universe is rational, then the kinds of things in the universe are not arbitrarily-selected. God didn't randomly create objects. Instead, God choose what objects to pick based upon God's perfect nature. If God necessarily exists and God's nature determines what things exist, then the objects in the universe necessarily exist and the objects not in the universe necessarily do not exist. Therefore, the objects in the universe either exist necessarily or they necessarily do not exist.

There are several problems with the above argument. Let's briefly mention two. First, the above argument relies upon the controversial premise that God exists. One objection then is that God does not exist. Second, even if we

assume that God does exist, the argument assumes that God's decision to create one kind of thing rather than another is necessitated by God's nature. This assumes that God lacks free will.

As we have seen, there are several reasons for rejecting the principle of plenitude and therefore thinking that 'everything exists' is false. Before abandoning this answer, let's consider a revised version of the principle of plenitude.

1.2.2 The dynamic principle of plenitude

Recall that in an earlier section, one objection to the principle of plenitude was from science. Some kinds of things go extinct (the Dodo bird) and some kinds of things come into existence (the coywolf). One possible response to this problem is to revise the principle of plenitude.

Definition 1.2: the dynamic principle of plenitude

Any kind of thing that can exist either has existed, currently exists, or will exist.

The dynamic principle of plenitude does not assert that the universe is maximally full at any given moment. Rather, it asserts that every kind of being that can exist will exist at some point in time. As Lovejoy[73, p. 244] puts it, '[i]t is only of the universe in its entire temporal span that the principle of plenitude holds good.'

Exercise 1.2

Is the dynamic principle of plenitude true? Give a reason for your view.

1.3 NOTHING EXISTS

Another possible answer to the question of "what exists?" is that *nothing exists*. This view is known as *metaphysical nihilism*. Typically, when someone calls a theory "nihilistic", they mean that a certain kind of thing does not exist. For example, if I am nihilistic about mathematical objects (e.g., numbers), then I deny the exist of mathematical objects.

Let's call our version of nihilism (one that denies the existence of any object) "extreme metaphysical nihilism".

Definition 1.3: extreme metaphysical nihilism

Absolutely nothing exists.

First, let's clarify extreme metaphysical nihilism. In saying that "absolutely nothing exists", this view denies the existence of entities where there is debate about whether that entity exists (e.g., aliens, unicorns) but also entities that people intuitively believe do exist. According to extreme metaphysical nihilism, tables, chairs, cars, cell phones, trees, and human beings do not exist. Even further, the theory entails that you do not exist and the world does not exist.

Next, let's evaluate extreme metaphysical nihilism. Extreme metaphysical nihilism has at least three problems. First, the theory is against common sense and so the theory bears the burden of proof. Second, while there may be plausible arguments for certain versions of nihilism, it is not clear whether there are *any* coherent arguments for *extreme* metaphysical nihilism. Third, many are convinced that there are extremely strong arguments for at least some things existing. A classic example is to claim that while things may not be how they appear (e.g., a tall building may look small from a distance), the fact that we have any experience at all implies not only (1) that we exist and (2) at least one other thing exists (the external world).

In this section, we considered extreme metaphysical nihilism.

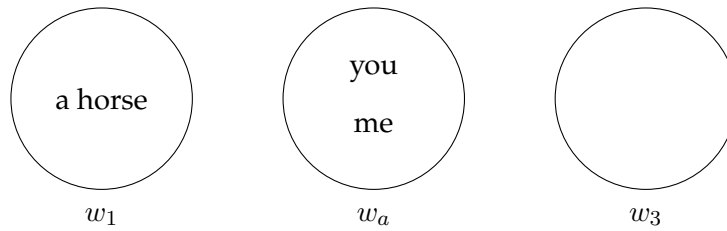
1.3.1 *It is possible that nothing exists*

Let's put extreme metaphysical nihilism to the side. A more popular version of nihilism is what is more known as "metaphysical nihilism". In contrast to extreme metaphysical nihilism that says absolutely nothing exists, plain-old metaphysical nihilism is the theory that "there could have been no concrete beings" or "there could have been no contingent objects." [19, p. 193]

Definition 1.4: metaphysical nihilism

There is at least one possible world w where there are no concrete beings.

Two important points. First, someone who accepts metaphysical nihilism might say that while our world (the actual world) contains many concrete things (e.g., humans, plants, computers), there is some possible world (different from ours) that is empty: it does not contain a single, spatiotemporal being. So, it is a world empty of humans, plants, tables, and computers.



Second, consider the classic question “why is there something rather than nothing?” What this question asks is why, in our world, are there things and stuff rather than nothing at all? Why isn't our world completely empty? If metaphysical nihilism were false, then one could give an easy answer to these questions: There is something rather than nothing because there cannot be nothing. In our world, there are things and stuff because there has to be things and stuff. Our world is not completely empty because there can be no empty worlds.

Let's consider an argument in favor of metaphysical nihilism. This is the “Subtraction Argument”[108][6] In what follows, we will consider a somewhat simplified version of this argument. First, let's start with a few propositions:

- A1: There might be a world w with a finite number of concrete objects.
- A2: For each concrete object in the world, that object might not exist.
- A3: The non-existence of an object O_i does not necessitate the existence of some other object O_j .

Let's briefly explain A1-A3. First, we do not know if our world has an infinite number or a finite number of concrete objects. But suppose our world is infinite because either the universe is physically unbounded (infinitely big) or because there is no smallest particle (infinitely divisible). A1 says that there is some possible world w relative to our world that is finite (not infinitely big and infinitely divisible).

Second, A2 says that for some concrete object O in that finite world, there is some other world where that object O does not exist. For example, suppose you are wearing a watch in this world (and this world is finite). A2 says there is a possible world where that watch does not exist. In sum, we can take the world w with the watch and get another world w^* without the watch. Even shorter, we can *subtract* objects from worlds.

Third, A3 says that the subtraction of objects from a finite world does not necessitate the addition of some other. Suppose there is an object at the edge of the physical space. A3 says that if we were to remove that object from

world, getting another world without that object, it isn't necessary that some other object takes its place.

With (A1)-(A3) clarified, consider the following argument in support of metaphysical nihilism:

- P1: Assume our world w_a (which may be finite or infinite) exists.
- P2: By A1, there is a world w relative to our world w_a that has a finite number n of concrete objects.
- P3: By A2, There is a possible world w^* like our world but with one fewer object: $n - 1$.
- P4: By A2, we can sequentially consider worlds relative to each possible world w with one fewer objects until we arrive at a world w_{min} that contains a single object $n = 1$. Let's call this object O_i .
- P5: By A3, the non-existence of O_i does not imply the existence of some other object O_j .
- IC: Therefore, by A2 and A3, there is a world w_{nil} that lacks the single concrete object that was in w_{min} .
- C: Therefore, since w_{nil} is a possible world containing zero concrete objects, metaphysical nihilism is true.

Let's rephrase the above argument in an easier-to-understand way.

Our world exists. We don't know whether it is finite or infinite. But there is a possible world relative to ours that is finite. Suppose the number of concrete objects in that world is n . In that world, we can imagine a smaller world that has one fewer object. It has $n - 1$ objects. We can repeat that process until we are imagining a world that has no objects. Since that world is empty, metaphysical nihilism is true.

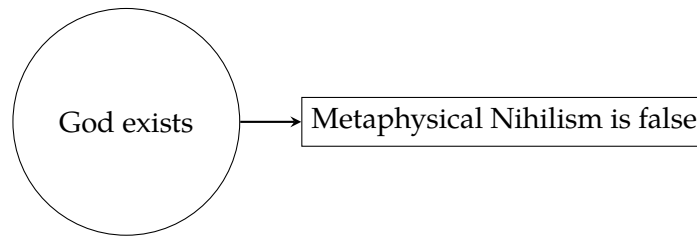
Exercise 1.3

Suppose you wanted to teach the "Subtraction Argument" to a child. Suppose you had a pile of blocks and built a "world" with these blocks. How might you use those blocks to explain the argument to that child?

Now that we have considered an argument in support of metaphysical nihilism, let's consider a few arguments against the view.

First, consider a theological argument. Suppose one can prove that God exists. If God is a concrete and necessary being, then this would mean that God not only exists in our world but in every possible world. And so, it would fol-

low that it is not possible for there to be no concrete beings since God exists in every possible world.[see 145]



The most controversial part of this argument, however, is the assumption that one can prove that God exists. We'll consider arguments for and against God's existence in another chapter.

Second, one might argue that at least one object must exist from the definition of a possible world.

- P1: A possible world is a connected spatiotemporal object.
- P2: If a possible world exists, then at least one concrete object exists (the world itself).
- C: Therefore, it is not possible for there to be a world without a concrete object.

1.3.2 †Argument from abstract objects

Let's consider an argument against metaphysical nihilism from the existence of abstract objects. This argument will claim that because abstract objects exist, concrete objects exist, and so metaphysical nihilism is false.

- P1: There are necessary truths that require the existence of abstract objects.
- IC: Therefore, there are abstract objects. (From P1)
- P2: Necessarily, if there are abstract objects, then there are concrete objects.
- C: Therefore, necessarily, there are concrete objects. (From IC and P2)

Let's briefly unpack this argument. P1 simply asserts that there are necessary truths. Examples include: mathematical claims like " $2+2=4$ " or sentences that are true analytically (by definition) like "all bachelors are unmarried men". In calling these truths "necessary", what is being said is that they are true in every possible world. So, no matter how the world is, these sentences are true in those world.

Next, IC contends that these necessary truths presuppose the existence of abstract objects. For example, if it is necessary that $2+2=4$, and $2+2=4$ can

only be true if there are numbers (abstract objects), then numbers (abstract objects) must exist.

The most controversial and unclear premise is P2. Before arguing for P2, let's point out something about sets. First, a set is defined in terms of its members. For example, if set $A = \{1\}$ and $B = \{1\}$, then A and B are the same set (as they have the same members). Similarly, if A is a set of "creatures with a heart" and B is a set of "creatures with a kidney", then if A and B have the same members, then they are the same set. Let's capture this idea by saying that a set's existence depends upon its members. That is, a set *existentially depends* upon its members.

Next, it is claimed that abstract objects can be *reduced to* sets and universals. For example, numbers can be expressed as sets in the following way:

- Let $1 = u$
- Let $n + 1 = n \cup \{n\} = \{n, \{n\}\}$
- Therefore, $2 = \{u\}$
- Therefore, $3 = \{u, \{u\}\}$
- Therefore, $4 = \{u, \{u\}, \{u, \{u\}\}\}$

The idea then is that that numbers are just fancy ways of expressing sets.

Now let's consider an argument for P2 in the above argument:

- P2.1: All abstract objects can be reduced to sets and universals.
- P2.2: The existence of a set depends on its members while the existence of a universal depends on its instances.
- P2.3: The existential dependence of an entity must "bottom out" somewhere, viz., the existence of a set cannot keep depending on the existence of another set.
- IC1: Therefore, there are things other than sets or universals upon which sets and universals existentially depend.
- IC2: These objects cannot be sets or universals. (P1).
- C: Therefore, there must be something concrete upon which sets or universals depend.

One objection that might be raised is that (1) while it is true that all sets existentially depend upon their members, it is also true that (2) there is some set that does not have any members: the empty set: $\emptyset$. And, if there is a set that does not have any members, then it does not depend upon any concrete thing. Thus, P2 is false: it is not necessarily true that if there are abstract objects, then there are concrete objects.^{[19]¹}

Exercise 1.4

Is it at least possible for there to be a world where nothing concrete exists?

1.4 SOME THINGS EXIST

Let's assume that it is not the case that *everything exists* and it is not the case that *nothing exists*. This seems to leave us with the common-sense conclusion that *some things* exist and *some things* do not.

Having this conclusion immediately raises the following question: *which things exist?* and *which things do not?* One ancient idea that remains popular is the "The Great Chain of Being" (or "Chain" for short). The Chain details not only what objects exist but also the order they exist in relation to each other. Namely, the Chain says that objects exist in a hierarchical relationship, where objects that are higher up in the Chain have more (or more perfect or powerful) attributes than those lower down in the chain.

1. At the top: God, a perfect being that is omnipotent, omnipresent, and omniscient.
2. Angels, wholly spiritual beings that are unchangeable (permanent)
3. Persons, beings that are a combination of spirit and matter, which are changeable and impermanent.
4. Animals (Beasts), wholly material beings that are capable of movement, have sensory organs, governed by instincts.
5. Plants, lack sense organs, limited movement, but capable of growing and reproduction
6. Minerals, no movement, growth, or reproduction.

What distinguishes different levels in the Chain are that higher beings in the chain have more *powers* than those below it. As Schumacher illustrates this feature,

"No one has any difficulty recognizing the astonishing and mysterious difference between a living plant and one that has died and has thus fallen to the lowest Level of Being, inanimate matter. What is this *power* that has been lost? We call it "life". "[113, p. 16]

At some of the levels of the Chain, there are subdivisions. For example, among angels, there are (in the order of rank): Seraphim, Cherubim, and Thrones. Similarly, among humans there are:

1. Kings (Monarchs)

2. Nobles
3. Peasants, working class

Using more contemporary categories, the Chain might organize human beings with the super rich, ruling elites at the top, followed by the working class in the middle, and, at the bottom, the poor.

The ruling elite, dictators, corporate moguls, super wealthy

Middle-lower working class

Homeless, lower class, poor people

Primarily, the Great Chain of Being is a theory about what exists and how objects exist in relation to each other. However, the theory also tends to come with some further claims. Some of these include:

1. God created the universe.
2. Each object that God created in the Chain is in its proper place.
3. The purpose of an object is primarily determined by its place in the chain, e.g., Kings are intended by God to rule (this is called "the divine right of Kings").
4. Your position in the Chain is determined at birth.
5. You cannot move up or down in the Chain.
6. Trying to act against your purpose is to act against both God and Nature. A peasant trying to overthrow the King would not only be against the King's laws, but it would also upset the natural order of the world that is set by God (the divine order).

Exercise 1.5

Do you agree with any part of the Great Chain of Being? If so, what parts?

1.5 WHAT EXISTS

In this chapter, we considered three main answers to the question of *what exists?*:

1. Everything exists: principle of plenitude
2. Nothing exists: extreme metaphysical nihilism

3. Some things exist and some things don't

In addition, we outlined one potential specification of our third answer via the Great Chain of Being. In the next chapter, we will consider whether the most important object in the Chain exists: God.

What exists? In the United States, while belief in God has decreased over time, most Americans are religious (69%) and most (81%) believe that God exists. Belief in God is a fundamental part of a broader set of religious beliefs. For many people, these religious beliefs have significant social, economic, ethical, and political implications. Religious beliefs influence how people vote, what policies they support, how they treat others, what products they buy, how they raise their children, and how they spend their free time. Despite being a popular belief in the United States, practical implications flowing from religious beliefs are controversial. Given that a fair share of the population does not believe in God (either because they believe God does not exist or they don't have a stance on the topic), the debate about whether God exists is not simply an academic matter.

This chapter aims at covering a broad number of arguments both for and against God's exist. Its aim is *breadth* rather than *depth*.

2.1 GOD EXISTS

People believe in God, but is this belief rational? If God does not exist, then while the belief may have some practical benefit (e.g., motivate people to do good), it is not rational. In this section, several arguments for God's existence are considered.

2.1.1 From Personal Experience

People claim to see God, speak to God, hear God, or feel God. For example, the Bible (Acts 9) says that the Apostle Paul heard the voice of Jesus on his road to Damascus. Acts 9 says that Saul fell to the ground and heard a voice saying ``Saul, Saul, why are you persecuting Me? [...] I am Jesus, whom you are persecuting. It is hard for you to kick against the goads [resist what is inevitable]". The argument from direct personal experience cites the direct personal experience have of God as evidence of God's existence.

- P1: Some people have direct, personal experiences of God, e.g., speaking to God or hearing God.
- P2: If people have an experience of God, then there is *prima facie* evidence of God's existence.¹

¹``prima facie evidence for God's existence" means that there is evidence that on first appearance establishes God's existence. What is meant is that there is evidence that unless that evidence is nullified is sufficient (enough) to establish God's existence.

- C: Therefore, there is prima facie evidence of God's existence.

There are several problems with the argument from direct personal experience. First, one of the biggest problems with claims about direct encounters with God is that personal experiences are private and thus cannot be examined (witnessed) by others. A convenience store clerk may claim that Tek robbed him late at night. His claim is supported by his personal experience but is also (at least in principle) capable of being witnessed by others or in other ways, e.g., customers, cameras. When there is reason to doubt the clerk's testimony, we seek these sources to remove said doubt. In contrast, most personal religious experiences tend to be private affairs that are beyond critical scrutiny.

Second, it is possible that people are hallucinating. Normally, we take our experiences of objects to be veridical. If I am staring at a red cup on a table, I believe that my experience is caused by the red cup. However, we also know that some people hallucinate. If I am hallucinating that there is a red cup on the table, then I really do see a red cup but the red cup is not really there. The experience of the red cup is caused by drug withdrawal (chemical imbalance), a medication I am taking, schizophrenia, or some other neurological condition. In this case, we might say that people really do hear voices or see visions, but these visions are not of God. For example, in July of 2025 a Missouri woman was arrested in Pennsylvania after driving her car around 120 mph and strikes multiple vehicles. The woman allegedly told the police "God told her to drive fast" (officers found THC products and psilocybin mushrooms in her car).

Third, there is an interpretive / inferential objection. Suppose Tek is studying in his room and he hears a voice in the next room. He believes that the voice is his sister's voice. However, the voice is not his sister's voice but his mother's voice. In this scenario, Tek wrongly interprets (or infers) the voice he hears to be his sister's.

Relatedly, a person may have experiences that they *wrongly interpret* (infer) to be from God, but, in reality, come from another source. For example, suppose Tek hears a voice in his head instructing him to give to the poor. Tek may interpret that this voice is the voice of God rather than his own voice (conscience).

Fourth, direct experiences with God are not consistently reproducible. Tek may claim to have spoken to God. When Renna asks how to have that same private experience, Tek may outline the steps he took. However, when she follows those steps, no such experiences are had by Renna. The non-reproducibility

Fifth, among those that claim to have received messages from God, those messages are inconsistent. Some people claim that God detests gay marriage, while others claim that God celebrates it.^[93] Some claim that God supports women priests, while others contend that God is against it.

2.1.2 *From Religious Texts*

Religious texts are often used as evidence that God exists. These texts typically (1) assert that God exists and (2) are taken to be reliable sources of information.

- P1: Religious texts say God exists.
- P2: Religious texts are reliable.
- C: Therefore, God exists.

There are several problems with the argument from religious texts for God's existence. First, P2 is controversial. What makes a religious text a reliable source of information? Some supporters of this argument point to the fact that many of the events in religious texts are historically accurate. For example, most historical scholars agree that Jesus was a historical figure, that he was baptized, and that he was crucified.¹ However, the problem with this argument is that while *parts* of religious texts may be historically accurate, this does not mean that *the whole* of the text is historically accurate. For example, even if there is historical agreement that Jesus existed, there is no historical agreement that he performed miracles, rose from the dead, or was the Messiah.²

A second way that P2 is defended is by claiming that it is reliable because God says it is reliable. This argument is, of course, circular. We know God exists because the text says so (and is reliable). It reasons as follows: We know the text is reliable because God says so.²

A second problem with the argument from religious texts concerns disagreement among the texts. The New Testament says that Jesus is the son of God, but the Quran, while acknowledging Jesus (Isa) as a prophet (messenger of God), denies his divinity as it (1) says there is no Trinity as God is One (Surah 5:73) and (2) has Jesus saying that he is not a god (Surah 5:116).

2.1.3 *From Prayers*

In this section, we consider the argument from prayers.

²The reliability of the text is grounded in God's existence and God's existence is grounded in the reliability of the text.

2.1.3.1 *Kinds of prayer*

First, a few definitions.

Definition 2.1: prayer

A 'prayer' is a solemn (serious) request to God, some supernatural power, or object of worship.

There are several different types of prayers.[51] Two examples are petition prayers and intercessory prayers.

Definition 2.2: petition prayer

A prayer petitioning God for something you desire or need.

Examples of petition prayers include: asking God to restore your health when you are sick, for material goods, for success, for courage, or for guidance. The Serenity prayer is an example of a petition prayer.

Serenity Prayer: God, grant me the serenity to accept the things I cannot change, courage to change the things I can, and wisdom to know the difference.

One type of petition prayer is an intercessory prayer.

Definition 2.3: intercessory prayer

A prayer requesting that God help someone else.

Examples of intercessory prayers include requesting that God help a sick person, save a child, give people courage, etc.

2.1.3.2 *Context on prayer*

Many people pray. Around 90% of all Americans occasionally pray while 55% pray daily.[41] Even people not associated with any particular religion (26%), agnostic (9%), and atheistic (1%) pray.[25] 70% of all Americans claim that God can *hear* their prayers. 42% of all Americans think that not only can God hear their prayers, but God acts based on those prayers. When making major life decisions, while most people say they rely upon their "own research", highly religious people also say that they rely "a lot" on prayer and religious reflection.[26]

Americans are evenly divided about how prayer should be integrated into society.

Legally and symbolically, prayer is recognized as important to Americans. In 1952 a law was placed making the first Thursday of May "National Day of Prayer". On that day, the president is required by law to sign a proclamation indicating that this is a day "on which people of the United States may turn to God in prayer and meditation at churches, in groups, and as individuals" (36 U.S.C. §119). In 2025 when President Donald Trump signed the proclamation, he wrote "From the earliest days of our Nation's journey, America has been guided by the grace of Almighty God."^[134] The [Freedom From Religion Foundation](#) challenged the constitutionality of National Prayer Day on the grounds that the day amounts to an "establishment" of religion, but their challenge was dismissed on the grounds that the group had not suffered any injury and had no standing to challenge the law.

One contentious issue is the place of prayer in venues where the government exerts some control. When put into these terms, individuals argue that this violates the First Amendment. In the United States, the First Amendment to the Constitution reads as follows:

"Congress shall make no law respecting an establishment of religion, or prohibiting the free exercise thereof; or abridging the freedom of speech, or of the press; or the right of the people peaceably to assemble, and to petition the Government for a redress of grievances."

The First Amendment is said to have two clauses concerning religion: (1) the Establishment Clause and (2) the Free Exercise Clause.

The *Establishment Clause* refers to this part of the First Amendment: "Congress shall make no law respecting an establishment of religion". This clause is understood as: (1) preventing any action on the part of the government to establish a religion and (2) requiring neutrality from the government when dealing with different religions. The *Free Exercise Clause* refers to this part of the First Amendment: "Congress shall make no law [...] prohibiting the free exercise [of religion]" . This clause requires the government to protect citizens' right to practice their religion so long as it does not run into conflict with either (1) public morals or (2) a "compelling" government interest (cf. [Sherbert v. Verner \(1963\)](#)).

The scope of the Establishment Clause was refined over the years following several cases. While its original application was restricted to the Federal Government, the Establishment Clause also applies to State government under

the Fourteenth Amendment.³ Let's consider a few cases that helped to clarify the Establishment Clause. First, *Reynolds v. United States* in 1879 clarified that while the government could not outlaw the religious belief in polygamy, it could regulate its practice, noting that it could be argued that the government should not interfere in other 'religious practices' like human sacrifice.

In *Everson v. Board of Education* (1947), the Supreme Court found that taxpayer money could be used to reimburse transportation costs to and from schools, including private, religious schools. The case is significant in that it clarified the scope of the Establishment clause, noting that it applied not merely to the Federal Government but also State Governments. The ruling clarified that there should be a "wall of separation between Church and State" and that there should not be any public spending "to support any religious activities, or institutions whatever they may be called, or whatever form they may adopt to teach or practice religion." Nevertheless, it was argued that the State was not in violation of the Establishment Clause since the money was for a "public purpose" rather than: the purpose of the reimbursement was not to establish or support a religion but to assist parents of all religions to get to and from school.

In *Illinois ex rel. McCollum v. Board of Education* (1948), the Champaign County Board of Education created a program where religious teachers (funded by a private party) provided weekly religious education instruction. The instruction was optional but McCollum argued that his son was disparaged for not attending these classes. The Supreme Court found that this was a violation of the Establishment Clause as it involved (1) the utilization of the tax-supported public school system to promote a religion instruction and (2) utilized the "its machinery for compulsory public school attendance" (which involves punishment for not attending) for religious instruction.

Whether the Establishment Clause prohibits prayer is particularly complex. In *Marsh v. Chambers*, the Supreme Court noted that there was a long history extending back to the Continental Congress in 1774 of opening sessions with a prayer from a paid chaplain. The Court argued that while historical pattern alone does not justify the constitutionality of the practice, this pattern gives insight into what the Establishment Clause was intended to mean. Relying on this historical pattern, the Court argued that the mere activity or praying does not reach the threshold of a government "establishment" of a religion nor was it a "step toward establishment". Instead, prayer before a governmental session was taken to simply be "a tolerable acknowledgment of beliefs widely held among the people of this country" (*Marsh v. Chambers*). The decision in *Marsh v. Chambers* was further affirmed in *Town of*

Greece vs. Galloway.

While prayer is permitted in legislative sessions, courts have declared school policies that involve school-led prayer to be unconstitutional. In the controversial case of *Engel v. Vitale* 370 U.S. 421 (1962), the New York Board of Regents recommended that all New York schools began each school day with the following prayer (following the Pledge of Allegiance):

“Almighty God, we acknowledge our dependence upon Thee,
and we beg Thy blessings upon us, our parents, our teachers and
our Country.”

The Regents thought it was the role of school to intensify a child's love of God and it was essential to teach that ‘Almighty God is their Creator’(see *Engel v. Vitale*, *Oral Argument - April 03, 1962, Part 1*). In 1958 the Board of Education in Long Island, run by William Vitale, directed schools to begin with the ‘Regents’ prayer’. Teachers were not allowed to comment on participation or non-participation in the prayer and, while initially the saying of the prayer was mandatory, the school district permitted, upon written request, to abstain from saying the prayer (although the teacher was required to utter the prayer). In contrast to *McCullum v. Board of Education*, while there were no explicit punishments for non-participation, it was argued by William J. Butler that both the content and the manner in which the prayer was performed were (1) forms of state-sponsored religious practice and (2) an effort to *teach* a specific kind of theistic religion and practice⁴. In addition, Butler argued that the implementation of the school prayer was coercive since requiring students to leave the classroom or parents to write to the school to ask for the child to be excused would identify children as *nonconforming* and isolate them from their peers. As Butler put it “[l]ittle children want to be with other little children” and to excuse kids from this dominant practice would “cast upon this child's mind some indelible mark”(see *Engel v. Vitale*, *Oral Argument - April 03, 1962, Part 1*).

Ultimately, the Court remarked that the issue was whether “New York oversteps the bounds when it finances a religious exercise” (*Engel v. Vitale*). In consideration of the issue, the Court noted that the amount of time required to say the prayer was ‘minuscule’ and what the New York school was doing was similar to what is done before court and legislative sessions. Nevertheless, the Court found that New York did violate the Establishment Clause since the “prayer was composed by governmental officials as a part of a governmental program to further religious beliefs” (*Engel v. Vitale*). In short, since the saying was used to inculcate religion into students, it violated the Establishment Clause.⁵

One of the reasons that prayer in school and *Engel v. Vitale* is so controversial is not only because many Americans pray, but many Americans believe prayer should be a part of school. As of 2025 roughly half (52%) of Americans believe that public school teachers should be allowed to lead their classes in prayer that specifically refers to Jesus (Christian prayer), while 57% are in favor of teacher-led prayer that refers to God (generic prayer).^[109] Opponents of school-led prayer contend that students are free to pray if they want to and that the government ought to be neutral when it comes to constitutionally protected religious matters. By mandating teacher-led prayer, the government is interfering in the religious lives of young people and potentially coercing certain religious beliefs.

Debates concerning the placement of prayer in schools is still ongoing. When the Florida High School Athletic Association (FHSAA) blocked a Cambridge Christian School from offering a prayer over the loudspeaker in a 2015 high-school football championship game, the school took them to court saying that the FHSAA violated their First Amendment rights. After the courts found that the free speech of the school was not violated, Florida lawmakers passed a bill allowing schools to make brief announcements (including prayers) before various events.^[49]

In 2025 while not mandating teacher-led prayer, Texas lawmakers passed two bills: (1) requiring the Ten Commandments in all K-12 classrooms and (2) requiring public schools "to provide students and employees with an opportunity to participate in a period of prayer and reading of the Bible or other religious text on each school day" (TX SB11).

2.1.3.3 *The Argument from Prayer*

In the previous section, we saw that many people believe in God and that prayer should be woven into the very fabric of social life. Presumably, one reason people believe in the importance of prayer is because they believe that their prayers (at least at one time in their lives) have been answered by God. Their belief is so strong that they believe it should be mandated by schools. To not foster prayer would be to prevent others from reaping some of life's greatest goods (e.g., blessings from God).

In sum, God's answering their prayers is an *indirect proof* of God's existence.

- P1: I prayed for *X*.
- P2: I received *X*.
- P3: The best explanation for why I received *X* is that God gave me *X*.
- C: Therefore, God exists.

The main problem with this argument is P2. Support for P2 is found both in personal testimony and religious texts (e.g., Genesis 20:17, 18; Numbers 12:13; Acts 28:8). However, since there are plenty of cases where a person's prayers were not answered, several studies have aimed to investigate this topic experimentally.

The effects of personal prayer (prayer done alone rather than in a group) can be divided into two types: *effects on the person* and the *effects on a person being prayed for*. Several studies suggest that praying has a variety of different effects on the person praying. It is correlated with decreased anger and aggression, increased marital faithfulness, better coping, and increased self-control and regulation.[50] To illustrate, one study investigated whether there was a correlation between prayer and alcohol use. It found that greater prayer frequency was negatively correlated with total alcohol consumption as well as problems relating to alcohol consumption.[67]

Studies involving the effects of personal prayer on the person praying are problematic due to third-variable explanations. Tek may begin praying to God for confidence and later find himself more confident, but it is possible the mere act of visualizing himself as a confident person or verbally clarifying his desire leads Tek to take steps to increase his confidence *or* lead Tek to believe himself to be more confident. Lambert et al.[67] point out that while there is a negative correlation between being religious and depression, it is possible that what accounts for this fact is the larger social networks of religious people. In other words, increased socialization is protective against depression, not a relationship with God.

Efforts have been made to try to control for third variables. For example, while prayer and alcohol consumption are negatively correlated perhaps some type of pseudo-prayer activity that is a part of prayer is responsible for decreased alcohol usage. In a follow-up study[67] involving prayer and alcohol, religious individuals (who already comfortable with the act of praying and who were currently in a relationship) were divided into three groups:

1. directed prayer group (told to pray about topics involving their relationships),
2. an undirected prayer group (instructed to pray but not given a topic),
3. a reflection group (asked to write about the same topics as those prayed upon by the directed prayer group).

The study found that those in the prayer groups only drank slightly more than half those in the reflection group.

While several studies suggest positive benefits associated with prayer, these

studies do little to show that these effects are due to the power of God. Exercise is correlated with positive health effects but these effects are due to the activity of exercise on the body, not God. It is thus quite plausible to assume that the activity of prayer works in an analogous way.

There are two other problems with research on the effects of personal prayer on the person praying. First, the vast majority of studies focus primarily on the positive benefits of prayer. It is possible that the activity of prayer also has several negative outcomes, e.g., increased gullibility, overly optimistic behavior, etc. If God is responsible for the effects of prayer, then presumably he would be responsible for both the positive and negative effects. Second, the vast majority of studies focus on the use of prayer for positive purposes, e.g., praying for loved ones, praying for peace, etc. But many religious people pray for evil outcomes, e.g., death to an enemy, the annihilation of a group of people, misfortune to befall one's boss. Further research is needed to see what kind of effects on the prayer these prayers have.

The second type of effect associated with prayer are its effects on others or the world. For example, I might pray for a new car to magically appear in my driveway or for my neighbor to recover from cancer. Unlike the former effect, if prayer elicited this second type of effect, it would be extremely convincing to non-believers.

On its face, this prayer does not seem to work. Set a pencil on a table and pray to God to rotate the pencil. The pencil will not rotate.

A common response from theists is that this profoundly misunderstands how prayer works: prayers do not *cause* or *force* God to respond since prayers asking God for something are *requests* to God. As C.S. Lewis puts it "The essence of request, as distinct from compulsion, is that it may or may not be granted. And if an infinitely wise Being listens to the request of finite and foolish creatures, of course he will sometimes grant and sometimes refuse them."

In addition, Lewis contends that our prayer must match our purpose. In praying for a pencil to appear out of thin air, my purpose is not to materialize a pencil. Instead, my purpose is to test the efficacy of prayer. So, since my prayer does not match my purpose, God (according to Lewis) won't grant my request.

Let's change the experiment then to something less foolish than making a pencil appear out of thin air. In addition, we can let the prayer match the purpose. Take two starving parents. Put them in a country undergoing a civil war. Now suppose they pray for food, not for themselves, but for their

starving children. Given the many starving people everywhere, this example is certainly not hypothetical. This sort of request (or plea for help) is routinely denied. There are countless examples of such cruel denial. People in war pray for their children and comrades to be saved. They are captured and tortured instead. A child may pray for God to save their sick parent. They die the next day. While good outcomes do occur after prayers, so do bad ones. So, if God is granting prayers, it seems to be without rhyme, reason, or sympathy.

While some argue that prayers cannot be tested[24], let's consider the experimental evidence on prayer.

From an experimental point of view, a study from Byrd[16] tested whether intercessory prayer (praying to God on behalf of others) to the Judeo-Christian God had any effect on a patient's medical condition while they were hospitalized. Each person who prayed was an active Christian who engaged in devotional prayer and they were asked to pray for a "rapid recovery and for prevention of complication and death". This study found that prayer was effective as patients who were prayed for had a lesser need for antibiotics, diuretics, and intubation / ventilation.

This study, as well as several others, some showing that prayer was effective, others showing it to be ineffective have been criticized for their design.[85] To cite just a few concerns. First, in a typical experiment there is an experimental (treatment) group and a control group. In testing for the efficacy of intercessory prayer, the experimental group would be patients receiving prayers from experiment participants and the control group would not receive prayers. Unfortunately, there is no practical way to prevent people from praying for the control group. There are scores of people who pray for the sick, the poor, the needy, and so on.[24, p. 41]

More recent studies aimed at correcting these errors, have not shown prayer to be effective in the medical context. For example, Benson et al.[9] investigated the effect of prayer on people who were recovering from a coronary artery bypass graft (CABG). That study showed had strangers pray for 14 days starting the night before the CABG prayer had no effect on their recovery. However, medical complications were higher in when patients were certain that they would receive prayers as opposed to (1) those who were uncertain they would receive prayers and did receive prayers and (2) those who were uncertain they would receive prayers and did not receive prayers.

2.1.4 From Miracles

One reason to believe in God is because of various reports of miracles, e.g., the dead coming back to life, turning water into wine, etc. Some people use the word 'miracle' to refer to any unusual and fortuitous event. Suppose I am walking in a field and my contact lens falls into the grass. It would be extremely difficult for me to find it. If I find it, I might say "it is a miracle that I found my contact." When used in this way, it is unclear whether my usage is meant to imply the acting of a supernatural agent on the natural events in the world.

In addition, some people use 'miracle' in an extremely broad sense. For example, I might say that 'childbirth is a miracle' or 'nature is miraculous'. This usage signals that there is something amazing or wonderful, but it is unclear whether it indicates any action by a supernatural being.

Finally, there is a usage of 'miracle' that does signal action by God. This is when God acts in a way that conflicts with the laws of nature.

Definition 2.4: miracle

A miracle is a violation of the laws of nature by the action of God or by one of God's intermediaries (e.g., angel, spirit, etc.). As Hume defines it "a transgression of a law of nature by a particular volition of the Deity, or by the interposition of some invisible agent" or "a miracle is a violation of the laws of nature." [58, sec.10]

Let's consider some examples. Jesus is said to have performed several miracles. Some of these include:

1. Turning water into wine (John 2: 1-11, 4:43-54)
2. Heals a paralyzed person (Matthew 9:1-8, Mark 2:1-12, Luke 5:17-26)
3. Healing the blind (Mark 8:22-26, 10:46-52)

While it is not physically impossible to make wine, it is physically impossible for water to instantly turn to wine by waiving one's hands over it or saying a few words. While it is not physically impossible to heal paralyzed people, it is physically impossible to cure a paralyzed person by saying 'Rise and walk'. Finally, while some forms of vision loss are curable (e.g., cataracts, refractive errors), blindness cannot be cured by spitting on a person's eyes and laying your hands on them.

Here is a more commonplace example. In a Reddit forum on Christianity, a person (let's call them 'Tek') tells a story about how they were volunteering at a food delivery organization. While delivering food to one family, Tek

noticed that they were one bag of food short. As Tek delivered the food to the family, he apologized for the delivery being one bag short. However, when Tek returned to his car, he checked the trunk of the car one more time and found an extra bag. Tek's interpretation of the event was that the bag was not there originally but God decided to create a bag out of thin air so the family would not go hungry.

To summarize, here is an argument from miracles to the existence of God:

- P1: There are miracles.
- P2: The best explanation for the occurrence of a miracle involves God.
- C: Therefore, God exists.

Next, let's consider an argument against the use of miracles to support the existence of God.

- P1: Assume it is reported that some event e is an apparent miracle.
- P2: One interpretation of the report of e is that the report is a genuine law of nature.
- P3: Another interpretation of the report of e is that the report was a hoax, misinterpretation, or hallucination.
- P4: It is rational to adopt the more likely (probable) interpretation.⁶
- P5: It is more likely that the report was a hoax, misinterpretation, or hallucination than a violation of the laws of nature.
- IC: Therefore, it is not rational to believe in miracles.
- C: Therefore, if miracles form the primary basis for a belief in God, then it is not rational to believe in God.

P5 needs some brief elaboration. There are many precise treatments of the above argument. Our approach will be to make the premise as intuitive as possible.

Approximately, 150,000 people die every day worldwide. When they die, they are buried, cremated, or treated in some other fashion. Of those 150,000 people, none of them come back to life. We do not see the ashes of the cremated body combine back together to form a human being that resembles the human before their death and cremation. We do not see cases where the skeletal remains of a deceased person grow flesh and then begin moving.

Next, suppose someone were to claim to have witnessed someone die, be buried, and then dig their way up from the grave. What could explain this report? Here are some possibilities:

1. Someone really did die and come back to life
2. The person is lying

3. The person was mistaken (they dreamed the event) or tricked (they were the victim of a prank)
4. The body that was buried was mistakenly buried alive and so never died
5. It is a hoax or prank.

What is the probability of the first option (a real miracle) compared to the other options? The first option seems the *least likely* given the fact that it goes against what we have observed about the thousands of people who die every day and the millions of people who die every year. It goes against what we know about the functioning and maintenance of the human body. It goes against what we know about other non-human life forms. And, it goes against

In contrast, we know that people routinely lie or are mistaken about what they see. We know that people do sometimes mistaken what they dreamed with what they experience. We also know that people do play pranks on people. So, P5 seems supported by the fact that a miracle (e.g., coming back from the dead) goes against what we know with a high degree of certainty. And, while it is possible that miracles do occur, believing in them is more an act of faith than an act of reason.⁷

2.1.5 *Argument from Common Consent*

One of the most interesting arguments for God's existence is the 'argument from common consent'. Let's consider a simple and crude version of the argument from common consent for the existence of God.

- P1: Everyone believes in God.
- C: Therefore, God exists

P1 is certainly false, so let's revise it a little more:

- P1: Every culture / nation, in near enough every historical era, has believed in a god.
- C: Therefore, a god exists

One of the merits of this argument is that it is easy to grasp. It doesn't require knowledge of science, to contemplate whether the universe is infinite or finite, or to carefully consider steps in an abstract proof. However, this is certainly not a very persuasive argument.

First, P1 is false. Polytheists don't believe in 'a god' but many gods. In addition, there are small atheist communities.

Second, the argument is invalid. It is possible for the premises to be true and the conclusion false. Just because everyone believes in God does not entail God exists. There are countless cases where many believe in a proposition *P* but *P* is (or turns out to be) false.

So, let's refine the argument a little more. In our newest version of this argument, we want to drop any claim that everyone believes in God or every culture believes in God, and we want to say that *most* people believe in God. In addition, we also don't want to directly say *just because most people believe in God*, it follows (with necessity) that God exists. Finally, rather than saying cultures believe in a single God, let's be clear and say they believe in *at least one God*. Here is our new version.

- P1: *Most* people believe in God (*nearly* every culture in every historical period has believed in some (at least one) God).
- P2: Either most people are wrong about one of the most important beliefs in their lives or they are not.
- P3: For any proposition *P*, if most people believe *P* is true, then it is *more likely* that *P* is true than not true.
- C: Therefore, it is more likely that God exists than does not exist.

This version of the argument has more promise. First, there is some support for P1.

- By some estimates, if we examine cultures throughout time, anywhere between 93-99% of the population believes in some higher power deity.
- In one poll, approximately 80% of Americans say that they believe in God, and of 10% of the people who say they didn't believe in God, they nevertheless claimed to believe in a higher power / spiritual force

P3 is the crucial premise. Here is the general thought process behind why P3 is true. If there is a *near consensus* that *P* is true, we can ask the question "why is there a consensus"? What caused the consensus? Many explanations can be given. Perhaps everyone believes the proposition due to chance, perhaps there is a global conspiracy, or perhaps there is some tricky aspect of the proposition that makes people believe think *P* is true. However, in support of P3, one might claim the following:

In the absence of a better explanation, the best (or most likely) explanation for the consensus is that *P* is true.

In other words, the reason that there is a consensus about *P* is because something in the world is driving people to believe *P*. Let's illustrate this idea with some examples and then apply it to God.

- Suppose 99.9% of people believe the sky is blue. What is the best explanation for why this is the case? Is it an optical illusion making us think it is blue, global conspiracy run by Jeff Bezos who wants people to believe the sky is blue, or does everyone think this because the sky is blue. The last explanation is the best (most) likely explanation and so we can conclude that the sky is blue.
- 97% of climate scientists believe in climate change. What would best explain this fact is that their beliefs are driven by the world (phenomena) and not due to chance, or a widespread conspiracy, or they are all hallucinating the facts, or they are all being bribed. Therefore, it is more likely that their belief is true.
- Approximately 95% of people believe in a higher-power. The best explanation for this phenomena is that God exists (not some global conspiracy, or chance).

The general point then is that in many cases (but not all), we can use the fact that there is a consensus on a topic as a sign that the world drives that consensus.

Now that we've clarified the argument and provided some initial support for some of the premises, it is worth noting that people tend to take this argument to be *laughably bad*.

2.1.5.1 Objections

There are at least two objections to the above argument.

The first objection is that P1 is too vague to be true since different cultures mean different things by "God" so there is no single proposition they agree upon. A quick response to this is that this objection commits the *inflation of conflict fallacy*. Just because people cannot agree about the precise details does not undermine the agreed upon propositions accepted by both parties. A group of people might agree that pumpkin bread is the best dessert but disagree about which how much sugar makes the best pumpkin bread. You might say that a 1-cup of sugar pumpkin bread is the best while I say it is 3/4 cup of sugar pumpkin bread. We disagree about the details but this disagreement doesn't undermine the fact that pumpkin bread ranks as the best dessert. Scientists may disagree about minute details of how a physical phenomena is to be explained but this disagreement takes place against a broad background of agreement.

The classic example of this involves the parable of three blind men each touching an elephant. The first is touching the trunk and thinks he is touch-

ing a massive snake. The second is touching the leg and thinks he is touching a tree. The third is touching the tail and thinks he is touching a rope. While all three men disagree about what they touching, this disagreement is contextualized against a broader agreement that they are touching something. The many believers may disagree about whether God is male, female, genderless, all-good, all-knowing, but this disagreement is set against a broad background of agreement that there is a God.

The second objection is more powerful. Recall that P3 says the following:

For any proposition P, if most people believe P is true, then it is *more likely* that P is true than not true.

And recall that P3 is supported by an inference to the best explanation argument:

In the absence of a better explanation for why a consensus that P is true occurs, it is likely that P is true.

P3 certainly seems true in some cases. Let's illustrate this with an example. Suppose I have a box with something in it. Suppose there are 100 people *independently* open up a box and record what they saw. All 100 people solemnly testify that there is a cake in the box. What best explains the curious phenomena that everyone separately came to the same belief that there is cake in the box:

1. Each person decided to say some random thing was in the box and it is only by *pure chance* that they all said "cake"
2. There is a global conspiracy run by the illuminati and they said that if they are ever asked what is in a box, they should say "cake"
3. It is a magic box capable of producing optical illusions in the minds of box openers.
4. There is a cake in the box. When each person opens the box, they see the cake and they say "there is a cake in the box"

All of these are possibilities, but if we have to pick *the best, most plausible, most likely explanation for the consensus*, then the best explanation is simply that there is a cake in the box. When we apply this same reasoning to explaining why there is a near consensus that God exists, the claim is simply that God exists.

The objection from critics is that *P3 is false* and so they contend that there is a better (more likely) explanation for why the majority of people believe in God. To illustrate, let's use our box example. Suppose I show one person what's in the box and they tell 9 *chosen people* what's in the box. Those

9 chosen people are given instructions to go out and try to increase the percentage of people that believe people that there is a cake in the box. They use a variety of practices:

1. A small number of them murder non-believers
2. They grant favors or social status to cake-believers and diminish the social or financial status of non-cake-believers
3. They educate young people to become cake-believers⁸
4. They exclude non-cake-believers from certain aspects of society
5. They create fun holidays and traditions for cake-believers
6. They present beautiful and uplifting stories that lead people to believe or stories that elicit fear and discourage nonbelief.

After some amount of time, a near consensus forms around their being a cake in the box. Now notice that in contrast to our first example involving the cake-in-the-box, the consensus is produced in a different way. In the first case, the consensus was produced by each person coming to their belief *independently*. Let's call this an *independent consensus*. In the second case, the consensus was produced by pressures and incentives. Let's call this a *manufactured consensus*.

In an independent consensus, it seems more plausible (in the absence of additional evidence) to posit some feature of the world to explain why the consensus occurs. We posit the existence of a cake to explain why everyone believes the cake is in the box. But, in the case of the manufactured consensus, it no longer seems necessary to posit the object of the belief. We don't need to posit a cake in the box to explain why everyone believes there is a cake in the box. Instead, we can posit the trickery of the leader and the proselytizing, tradition, propaganda, political trickery, force, murder, and/or coercion by the followers of the leader. The critic will argue that this explanation is more conservative since it doesn't involve positing a supernatural entity. In sum, P4 is false in our revised argument because the consensus is not best explained by God's existence.

One quick clarification is that this objection is not saying that all belief in God is the result of coercion, indoctrination, propaganda, etc. Some people may come to believe in God on their own. The objection says that *the reason there is a consensus* is best explained by these factors. Support for this claim is based on the fact that belief in a high Gods in isolated cultures (e.g., the monotheistic God) is low (24%).

2.1.5.2 Common consent for Nature Spirits

We have revised the argument from common consent a few times. Let's give it one more try. The final refinement we'll consider has been called the best version of the argument from common consent. In putting forward this version of the argument, I'm drawing on a 2020 article by Tiddy Smith (see the description below for details).

The problem we saw with our earlier version of the argument from common consent is the consensus around God's existence is likely manufactured. Let's try to create an argument from common consent where consent around God's existence is independent. In other words, what do people believe about higher powers if they all separately look into the box (if they were to form their beliefs free from pressure)? Since most of us live in a highly-connected, global culture, the best place to look for what people would believe about higher powers are isolated cultures or from very early cultures before various forms of proselytizing could stamp beliefs across the globe. Now depending on how far back we look, it is recognized by anthropologists that beliefs in high Gods is low to nonexistent. For example, in one report, beliefs in high Gods (e.g., the monotheistic god) was around 24%. However, it has long been noted that 100% of isolated cultures believe in the existence of some type of *nature spirit* (e.g., mountain spirit). A belief in nature spirits is recognized as the "oldest trait of religion".^[99] The next two most common beliefs are (1) belief in afterlife and (2) belief in religious specialists (visionaries/shamans). This suggests that the best version of the common consent argument is for *animism*. Animism is complex religious tradition (some say it is not even a religion at all but a way of living or sentiment). For our purposes, however, we only need one small part of it. The part we need is the belief that nature spirits exist, they have a viewpoint, i.e., they can, in some sense, "see" the activities of humans, and they have some amount of agency in (they can influence) the world. With this in mind, I'm going to define animism as follows:

Definition 2.5: animism

Animism is a religious tradition or temperament (associated with hunter-gatherers) that attributes sentience (or life force) to a wide range of beings or places in nature (e.g., animals, plants, minerals, stones, oceans, rivers, mountains, and in modern types of animism to technologies like cars or computers).

An example would be the Yup'ik people, an indigenous group who live in

the western and southern parts of Alaska.^[43] The Yup'ik people believe that the ocean has a “sense of awareness” and knows when people have fallen through the ice, when children have died, or when women have had miscarriages. Various practices must be taken to “close the eyes” of the ocean to prevent invoking its anger.

With all of these terminological matters out of the way, let's try to revise our argument, not for God but for the existence of nature spirits:

- P1: Most people living in isolated cultures believe that there are natural spirits.
- P2: The consensus that there are natural spirits is formed independently (not coerced).
- P3: Either most people are wrong about one of the most important beliefs in their lives or they are not.
- P4: For any proposition P, if most people believe P is true (and do so independently of each other), then P is more likely than not-P.
- C: Therefore, it is more likely that natural spirits exist than not exist.

While this new version of the argument from common consent does seem attractive since it address the problem of manufactured consent that applies to the argument for a monotheistic God, it is still not without serious problems.

First, a critic can argue that P4 is false. If we are uncertain about whether nature spirits exists, it may be reasonable to assume that if many people believe a proposition they exist, it is likely that they exist. However, if we are not uncertain about this proposition because we know (or have strong evidence) that they do not exist, then what people believe about the topic should play no significant role in deciding whether nature spirits exist or not. For example, suppose nearly everyone began believing that some scientifically-proven belief was false. This surely wouldn't make that belief false.

Second, and the more interesting criticism is that P4 is false since *there is a better explanation* (one that does not imply the existence of nature spirits) for why isolated cultures all came to independently believe that nature spirits exists. The usual explanation is that there is an *evolutionary benefit* to believing in nature spirits. The general line is that since human decision-making isn't perfect, there is an evolutionary benefit for making the “least costly error”. Take a smoke alarm. Since your smoke alarm isn't perfect, it is better for you if the smoke alarm is too sensitive than not sensitive enough. It is a minor trouble to deal with the false alarm when you put the wrong cooking oil on a hot pan than to deal with the house burning while you slumber. Here is a simpler example, it is better to think that sticks are snakes than to think that snakes are sticks.^[63]

All of this sounds nice for smoke alarms, sticks, and snakes, but what is the evolutionary benefit of believing in the supernatural? What do we get from thinking that there are nature spirits like mountain spirits, river spirits, forest spirits and so on? What benefit is there for hyper-inflating agency in the world?

The claim made by some evolutionary psychologists is that it is simply *less costly* to believe in god, gods, or spirits than it is to believe they don't exist. They are posited for all sorts of reasons, but let's look specifically at thinking there are nature spirits. Here is one possible explanation.

Now, the greatest predator of man is man. Between warring people then, consider which mistake is more costly.

- Thinking an enemy is present when they are not
- Thinking an enemy is not present when they are.

The less costly false belief is clearly thinking that (1). It thus pays to believe that there are agents, beings, spirits watching even when no such being exists.

But, we need not simply consider external threats. Beliefs in nature spirits is also benefit within one's group. The usual claim is the following: *belief in supernatural punishment for selfish actions encourages individuals to cooperate*, and this is beneficial for, not only the group, but *the individual* as it decreases the likelihood their selfish activity will be detected and punished. To illustrate, suppose you are going to do some selfish action that is taboo within your culture (let's say its theft). You want to avoid detection and punishment but you are not a perfect being and so are bound to make mistakes. Just like it is better for you to believe that a stick is a snake rather than a snake is a stick, it is better for you to believe that someone is watching rather than someone is not watching. If you incorrectly believe someone is watching (and they aren't), the cost is that you don't get to have the thing you wanted to steal (a minor cost). In contrast, if you incorrectly believe someone is not watching (and they are), the cost is major: you are likely to be punished by the group.

In short, our brains are "trip-wired to spot lurkers" [4, p. 719]. Since humans aren't perfect, we are bound to make mistakes, but when we make mistakes, we want to make the "least costly error". Since we can't always know when someone is watching us, we are simply better off positing a being or beings that are watching us (e.g., spirits or a God) even when no such beings exist.

And there we have it, while it is striking that there is an independent consensus that nature spirits exist, there is no need to posit the existence of spirits

to explain why this consensus occurs. We can give a more conservative explanation: the consensus exists because groups that believe in a world with spirits are better off (from an evolutionary point of view) than those that don't. It is an evolutionary benefit to have a cognitive bias to thinking that nature spirits exist.⁹

2.1.6 *Kalam Cosmological Argument*

Every object seems to have a cause. Your birth has a cause, your continued presence here has a cause, the clothes you wear have a cause, the food you eat has a cause, etc. If every object has a cause and the universe itself is an object, then surely it has a cause. But what could that cause be? It cannot be another object in the universe because then it would need a cause as well. So, it must be something outside of space and time (the typical causal order of the world) and an eternal / timeless being. And, since it is the cause of the universe, it must also be extremely powerful. The best explanation then for the cause of the universe is God.

- P1: Everything that *begins* to exist has a cause of its existence (evidence: experience)
- P2: The universe began to exist (evidence: science)
- IC: Therefore, the universe has a cause to its existence (a first cause).
- P3: The first cause is God.
- C: Therefore, God exists.

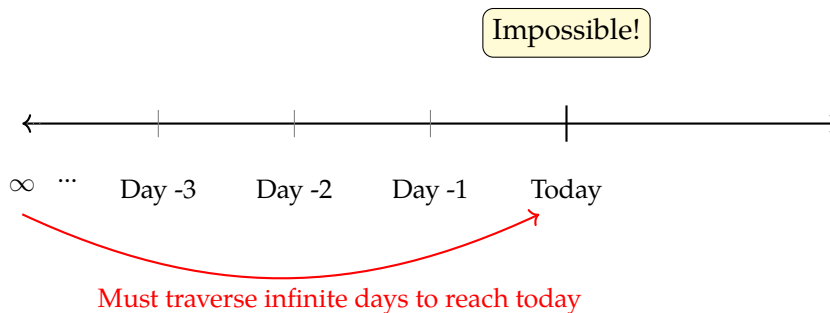
P1 is supported by our experience with objects in the universe, e.g., people, cell phones, even mountains.

P2 is supported by two arguments. First, it is supported by empirical evidence / science. The Big Bang theory is a theory about the expansion of the universe 13.7 billion years ago. It extrapolates from the current expansion of the universe to a point where the universe was infinitely dense (compact) and infinitely hot. The theory is supported by its prediction (and discovery in 1965) that there would be microwave radiation leftover from the Big Bang.

Second, it is further supported by arguments against an infinite sequence of causes / past events. Consider that the history of the world (the past) is formed by adding one event to another event. For example, the past is formed by adding a new day to the prior day, or a new year to the prior year. More simply, we traverse through past events step by step.

Now consider the events occurring today and suppose the past is infinite. If the past is infinite, then an infinite number of prior events would need to occur. That is, an infinite number of days, years, centuries would need to

elapse. But an infinite number of moments could not be traversed step by step. Therefore, the past is not infinite.



There are several objections to the above argument.

First, one might accept the argument to IC but not to C. That is, one might accept that everything *in the universe* begins to exist has a cause, that the universe began to exist, but reject that God is the cause of the universe. In putting forward this argument, one might say that the *singularity* is (1) not a part of the universe and (2) is the cause of everything in the universe. As Park^[97], pp. 325, 326] puts this point,

“the physical universe does have a cause. It is the singularity. The singularity is not in three-dimensional space. It exists outside of and before the existence of space time just as God is thought to as well. Therefore, it not a part of the early universe. [...] The singularity is not a part of the universe but rather, is the cause of the universe.”

A supporter of the kalam argument might respond by trying to revise the above argument as follows:

- P1: Everything that *begins* to exist has a cause of its existence.
- P2: The universe began to exist (evidence: science)
- IC: Therefore, the universe and the singularity has a cause to its existence.
- P3: The cause is God.
- C: Therefore, God exists.

The problem with the above argument now is that it commits the ‘hasty generalization fallacy’. To see this more clearly, consider that P1 refers to (1) every physical object (every object in the universe) and (2) every object that *begins to exist*. The problem then with P1 is that it does not apply to the singularity because (1) it is not a physical object (it is not spatiotemporal). and (2) since it is outside of time, it never begins to exist.

Suppose you randomly survey people with blue hats and you find that all of them are voting for Tek for President. From this, you conclude that everyone wearing a hat will vote for Tek, including Jon, who is wearing a *red* hat. The problem with this conclusion is that you are generalizing that anyone wearing a hat will vote for Tek, when the sample on which your generalization depends only involves *blue-hat* people.

A similar problem is found with the above argument for the existence of God. Recall that the evidence for P1 is that every *physical object* (in spacetime) we have ever encountered has a cause. However, note that the singularity is not a physical object (it is a theoretical starting point that is not spatiotemporal). Therefore, just as it would be unjustified to conclude that red-hat people are voting for Tek on the basis that all blue-hat people are voting for Tek, it would also be unjustified to conclude that non-physical things (e.g., numbers, propositions, perfect circles, the singularity) have a cause on the basis that all physical things have a cause.¹⁰

In addition, our evidence that objects have causes applies to objects that *began* to exist. We have no evidence that objects that never began to exist or always existed have causes. In the case of P1 however, since the singularity is outside of time, it *never begins to exist*. Therefore, we cannot conclude from our evidence of things that began to exist having causes that something that never began to exist has a cause.

Let's consider a few final responses from the theist.

First, the theist might insist that the singularity (even if it is not spatiotemporal) has to have a cause and that cause is God. The atheistic response is to argue that there is no justification for this view. Just as God's existence is brute (does not have a cause), the singularity's existence is brute. And, all other things being equally, the atheist's position is actually simpler since it only supposes one 'strange' thing exists (just the singularity) rather than two 'strange' things (the singularity and God).

Second, the theist might insist that it is not merely the case that everything that *begins* to exist has a cause, but everything that begins to exist *and* everything that exists has a cause. The obvious problem with this revision is that it would also apply to God. That is, since God exists, God would need a cause, and that cause would need a cause, and so on, ad infinitum.

2.1.7 Cosmic Watchmaker Argument

Suppose you are walking along and you find a working watch on the ground. Where did the watch come from? Here are two hypotheses about its origin:

1. The parts of the watch formed the watch through entirely random (arbitrary) collisions. The particles collided together to form the watch parts and the parts randomly collided together to combine into the watch.
2. There was a designer (or team of designers) who intelligently used various materials to form watch parts and then assembled these parts to form a working watch. Or, the designers crafted a machine that was capable of crafting the watch.

Hypothesis (2) is much stronger as it aligns with every watch we know. Generalizing from (2), we might say that every *complex* object whose parts must be coordinated in a specific way to serve some purpose entail a designer. In addition, the more complex and organized the object, the more intelligent the designer must be. For example, a modern cell phone or requires more intelligence than a watch or a clock.

The next step in this argument is to point out that some natural objects are even more complex (e.g., the human eye) than artificial objects. So, just as the highly complex artificial objects imply a extremely intelligent designer, complex natural objects (eye, ecosystems, the universe) imply an almost supremely intelligent designer (God).

- P1: There are complex, artificial objects whose parts must be precisely coordinated to serve some purpose (e.g., watches).
- P2: The only reasonable explanation for the existence of such objects is the existence of an intelligent designer (watchmaker).
- P3: Many complex, natural objects in the universe have parts whose parts must be precisely coordinated to serve some purpose (e.g., the human being, the biological eye, the universe).
- IC: Therefore, the only reasonable explanation for the existence of natural objects is the existence of an intelligent designer.
- P4: Many natural objects are more complex and sophisticated than the artificial objects created by human beings.
- C: Therefore, the designer of these objects must be vastly more powerful and intelligent (God).

In support of P2, Newton argued that men and animals could not be "fortuitous jumbings of attomes". If they were, their bodies would be composed of many useless parts. As Newton[87, 128r] put it, "here a lump of flesh there a member too much some kinds of beasts might have had but one eye some more than two & the two eyes."

Some objections:

1. Darwin. Critics of the argument contend that the complexity, order, and functionality of natural objects can be explained without appealing to a designer. Complexity of human organisms (species) can be explained by evolution, complex rock patterns (strata) can be explained by heat build up from pressurization, etc., etc.
2. Does not prove God is omnibenevolent, omnipotent, or omniscient. Only proves the existence of a powerful and intelligent being.
3. Polytheism. Consider that complex artificial objects do not imply a single designer. A cellphone is designed by a team of individuals who create the hardware, software, aesthetics, etc. Applying this to complex natural objects would imply polytheism (the existence of multiple gods) rather than monotheism (the existence of a single god).

2.1.8 *Pascal's Wager (Short Version)*

In this section, Pascal's wager for the existence of God is introduced. The format for presenting this argument is first to introduce some key concepts in decision theory. Once these ideas are presented, we'll then turn to a clear presentation and evaluation of the argument.

2.1.8.1 *Key ideas in decision theory*

Before considering this argument, let's explain how to calculate the expected value of an action using a ``payoff matrix''. A *payoff matrix* is a table that displays the outcomes (or ``payoffs'') of different choices under different possible states of the world. The rows represent the choices (actions) we can take, and the columns represent the different possible states of the world (which may occur with certain probabilities). The intersections of the rows and columns (the cells) represent the payoffs we receive if we take that action and that state of the world occurs.

Let's consider an example. Suppose we are considering whether to move to a new city or stay in our current location. As these are two actions we might take, let's add these to the rows of a table:

Move		
Stay		

To keep things simple, let's consider two outcomes (states of the world): moving would be beneficial (make our life go better) and moving would not be beneficial (make our life go worse). Let's add these to the columns of the table:

	Beneficial	Not beneficial
Move		
Stay		

Next, let's add the payoffs in the cells. Note that in saying that moving would make our life go better, this does not imply that staying would make our life go bad.

	Beneficial	Not beneficial
Move	+10	-2
Stay	+1	+1

So, if we move to a new city and this move is beneficial, then our life improves by 10 units. On the other hand, if we move to a new city and this move is not beneficial, then our life gets worse by 2 units.

Given the above payoff matrix, what should we do? The answer to this question largely depends upon the type of person we are.

If we are adventurous (focusing on the highest possible upside), we might use the *maximax rule*: choose the option with the best possible outcome.

1. Move: best case = 10
2. Stay: best case = 1

According to maximax, we ought to move.

If we are cautious (wishing to avoid the worst possible outcome), we might use the *maximin rule*: choose the option with the best worst-case outcome.

1. Move: worst case = -2
2. Stay: worst case = 1

According to minimax, we ought to stay.

Finally, if we are the type of person who wishes to minimize the maximum possible regret, then we might use the *minimax regret rule*. We'll use a three-step process to apply the minimax regret rule. First identify the best payoff in each state of the world. In our example, we get the following results

1. Beneficial (column 1): 10
2. Not beneficial (column 2): 1

When we think about regret, we think about what we could have had if we chose optimally. So, to calculate "regret", calculate the amount by which each choice falls short of the best option in that state.

In the Beneficial state, here is our regret:

1. Move: $10-10=0$
2. Stay: $10-1=9$

Regret is then calculated by subtracting the payoff of each act from the best payoff in its column. This yields regret values of 0 and 3 for Move, and 9 and 0 for Stay. Since the maximum regret attached to Move is 3, while the maximum regret attached to Stay is 9, the minimax regret rule recommends choosing Move.

Consider the state where moving is beneficial, here is our regret:

1. Regret of moving: $10-10=0$
2. Regret of staying: $10-1=9$

Notice that we have no regrets in moving since it is the optimal choice. In contrast, we have considerable regret in staying if moving is beneficial. Now consider the state where moving is not beneficial. Here is our regret:

1. Regret of moving: $1-(-2)=3$
2. Regret of staying: $1-1=0$

If moving is not beneficial, then we would have some regret moving and no regret staying.

In each outcome, we can calculate the regret we have for each choice we make in that outcome. Let's take the *maximum* regret we would have for each choice:

1. Stay: $\max(9, 0) = 9$
2. Move: $\max(0, 3) = 3$

The minimax regret rule says to choose the option that minimizes your maximum regret. In other word, make the choice that would have you regret the least. Since the choice that would minimize our maximum regret is to move, we ought to move.

Let's introduce two more ideas by way of an example. Suppose that your current living situation is terrible. If you continue to stay where you are, your life will get worse. However, it isn't a guarantee that your life will get better if you move. With that in mind, consider the following payoff matrix:

	Beneficial	Not beneficial
Move	+10	0
Stay	-5	-5

Notice that the decision to move is better in both states. When an action *A* gives a better payoff than *B* in every state, then *A* *strictly dominates* *B*. So, in the above example, *moving* strictly dominates *staying*.

Consider the following payoff matrix:

	Beneficial	Not beneficial
Move	+10	0
Stay	0	0

Notice that the decision to move is better in at least one state and equally as good in the other. When an action A (1) is better than B in at least one state and (2) is at least as good as B in every other, then A *weakly dominates* B . So, in the above example, *moving* weakly dominates *staying*.

The following table summarizes the key decision-making rules discussed above:

Decision Rule	Description	Attitude
<i>Maximax</i>	Choose the option with the best possible outcome	Optimistic/Adventurous
<i>Minimax (Maximin)</i>	Choose the option with the best worst-case outcome	Cautious/Risk-averse
<i>Minimax Regret</i>	Choose the option that minimizes your maximum regret	Regret-minimizing
<i>Strictly Dominates</i>	Action A gives a better payoff than B in every state	Dominant strategy
<i>Weakly Dominates</i>	Action A is better than B in at least one state and equally as good in all others	Weakly dominant strategy

2.1.8.2 Decision theory applied to God

Suppose that God is infinitely incomprehensible: we can neither know what God is or that God is. But consider the fact that God either exists or does not exist: "God is, or He is not." There are thus two different possible states of affairs:

- God exists
- God does not exist.

According to Pascal, there are two different choices (attitudes) we can have with respect to whether God exists.

1. Believe that God exists.
2. Not believe that God exists.

Unlike the choice between heads and tails in a coin flip, it seems that we must choose one or the other. This is because trying to not choose (ignoring the question) is equivalent to not believing.

Given the two states of affairs and the two choices, we might try to set up a payoff matrix like the following:

	God exists	God does not exist.
Believe		
Not believe		

The question then becomes: which attitude will you choose? Rather than choosing blindly, Pascal says we ought to “weigh the gain and the loss in wagering that God is.”[98, p. 233] According to Pascal, “If you gain, you gain all; if you lose, you lose nothing.” In other words, if you believe in God and God exists, then you gain infinite life and happiness. But, if you believe in God and God does not exist, then you neither gain nor lose.

	God Exists	God Does Not Exist
Believe in God	$+\infty$	0
Do Not Believe		

What about if we do not believe in God? If God exists, then we miss out on eternal life and happiness. If God does not exist, then there is no real benefit.

	God Exists	God Does Not Exist
Believe in God	$+\infty$ (infinitely happy)	0 (lose nothing)
Do Not Believe	$-\infty$ (lose everything)	0 (gain nothing)

With the payoff matrix complete, notice that the decision to *Believe* weakly dominates the decision to *Not Believe*. Belief in God offers at least the chance at infinite life and happiness and costs nothing. We thus have the ingredients for the following argument for *believing that God exists*:

- P1: If you believe in God and God exists, then you will receive a great reward but if God does not exist, then you lose nothing.
- P2: If you don't believe in God and God exists, then you lose out on a great reward but if God does not exist, then you gain a minor reward.
- C: Therefore, in the absence of knowledge, it is better to believe that God exists.

2.1.8.3 Objections to Pascal's Wager

Let's consider three objections to Pascal's Wager.

First, in our construction of the payoff matrix, we assumed that believing in God costs nothing and not believing in God gains you nothing. One objection is that this is not correct. In many religions, belief in God also involves a variety of different personal sacrifices, e.g., fasting, abstinence, other forms of asceticism. If belief in God requires abstaining from certain prudential goods, then there is a cost to believing in God. And, if the non-belief in God allows for the partaking of these goods, then there are certain benefits in not believing in God.

	God Exists	God Does Not Exist
Believe in God	$+\infty$ (infinitely happy)	-10 (lose goods)
Do Not Believe	$-\infty$ (lose everything)	10 (gain goods)

In the above payoff matrix, no choice strictly (or weakly) dominates the other.

Exercise 2.6

In the revised payoff matrix, no choice strictly (or weakly) dominates the other. However, consider the other decision rules (maximax, minimax, minimax regret). What should one believe given these rules?

Second, while Pascal suggests that we know nothing about God's nature, Pascal seems to assume that God rewards belief and does not reward non-belief. But suppose that it was the other way around. God might say "I gave you no sign that I exist and you foolishly believed anyway. Therefore, you deserve to be punished for all eternity." That is, if we suppose that God rewards those who believe based on evidence and punishes those who believe in the absence of evidence, then assuming then that God has given us no sign whether God exists or does not exist, the payoff matrix would be the following:

	God Exists	God Does Not Exist
Believe in God	$-\infty$ (infinite punishment)	0 (nothing)
Do Not Believe	$+\infty$ (infinite happiness)	0 (nothing)

On this revised form of Pascal's Wager, it is better to not believe than to believe.

A third problem with Pascal's Wager is it seems to assume that we can *will* ourselves to believe that God exists. For example, suppose an old professor

promises his students an A in his philosophy course if they believe he is under 34 years of age.^[147] While the students can *say* they believe, saying one believes something is not sufficient for actually believing. A common test to determine whether one believes a proposition *P* is true is whether they would be willing to act on the basis of that belief. For example, suppose the lives of the students were on the line. If the student's belief is true, then they survive. If it is false, then they perish. Surely, the students would take one look at the old professor and say "if I had to stake my life on whether the professor is under or over 34 years old, he is older."

The same then is true concerning a belief in God. While Pascal's Wager may make a belief in God look attractive, the argument alone cannot lead someone to believe in God.³

To put it bluntly, Pascal's suggestion is to try to "fake it until you make it". He suggests that you will come to a belief in God not by considering more rational arguments for the existence of God, but by pretending to believe.⁴ Pascal^[98, p. 233] writes,

"You would like to attain faith, and do not know the way; you would like to cure yourself of unbelief, and ask the remedy for it. Learn of those who have been bound like you, and who now stake all their possessions. These are people who know the way which you would follow, and who are cured of an ill of which you would be cured. Follow the way by which they began; by acting as if they believed, taking the holy water, having masses said, etc. Even this will naturally make you believe, and deaden your acuteness."

2.1.9 *The pragmatic-moralistic argument*

This argument is attributed to William James. Similar to Pascal's argument, it does not argue that God exists but instead argues that we are better off believing in God. The rationale is that if we believe in God, we will behave better (more ethically).

Bertrand Russell^[110, p. 545] directs several criticism against this argument.

First, he points out that it does not prove that God exists. He write "at its

³Pascal considers someone in this situation when he writes, "I have my hands tied and my mouth closed; I am forced to wager, and am not free. I am not released, and am so made that I cannot believe. What, then, would you have me do?"^[98, p. 233]

⁴"Endeavour then to convince yourself, not by increase of proofs of God, but by the abatement of your passions."^[98, p. 233]

best, it cannot prove that there is a God but only that politicians and educators ought to try to make people think there is one. [...] The arguments are of the same sort as those which urge that children should be taught respect for the flag."

Second, Russell claims that the premise upon which the argument relies is doubtful since there are good people who do not believe in God and bad people who believe in God. Russell writes,

"Many of the best men known to history have been unbelievers. John Stuart Mill may serve as an instance. And many of the worst men known to history have been believers. Of this there are innumerable instances. Perhaps Henry VIII may serve as typical."

2.1.10 *The pragmatic-prudential argument*

Similar to Pascal's argument and the pragmatic-moralistic argument, the pragmatic-prudential argument does not argue that God exists. It instead argues that people are "happier" if they believe in God. The argument then is as follows:

- P1: If I don't believe in God, I will be less happy than if I did.
- P2: You are better off believing in what makes you happier.
- C: Therefore, you are better off believing in God.

Russell levels several criticisms of the above argument. First, he calls the above argument a "coward's argument". He writes:

"Nobody but a coward would consciously choose to live in a fool's paradise. When a man suspects his wife of infidelity, he is not thought the better of for shutting his eyes to the evidence." [110, p. 546]

In other words, P2 is false since there are cases where you are "better off" knowing certain things even if they will make you "unhappier".

Second, he takes issue with P1. [110, p. 546] He notes that there are several things that make a person better off, e.g., good health, food security, shelter, friends, good mental health. He contends that if you have these things, you are likely to be happy independent of whether you believe or not believe. And, if you lack these things, you are likely to be unhappy regardless of whether you believe or not believe. Finally, he suggests that there is no evidence to suggest that, all other things being equal, you will be happier if you believe than not believe.

2.1.11 Ontological Arguments

There are many versions of this ontological argument. Since our goal is to provide an overview of arguments for the existence of God, let's consider a relatively simple version of the ontological argument.

2.1.11.1 Definitional ontological arguments

Let's start with the simplest and shortest version of the ontological argument: the *definitional modal argument*:

- P1: God is as a being that is perfect in every way (a perfectly supreme being).
- P2: Existence is a perfection.
- C: Therefore, God exists.

At first glance, an atheist might deny P1. P1 assumes there is a being that has all of the perfections, but if no such being exists, then P1 is false. The theist might then revise P1 and restate their argument as follows:

- P1: By *definition*, God is as a being that is perfect in every way.
- P2: Existence is a perfection.
- C: Therefore, God exists.

The above argument no longer *assumes* God exists since P1 is merely defining "God" as the being that has all of the perfections. The problem however with this revision is that conclusion no longer follows from the premises (the argument is invalid). The atheist will respond that the correct conclusion is that *By definition*, God exists, rather than *God exists*.

- P1: By *definition*, God is as a being that is perfect in every way.
- P2: Existence is a perfection.
- C: Therefore, by *definition*, God exists.

Some might wonder what is wrong with the above conclusion. The problem is that there is a distinction between something *existing* and *existing by definition*. To illustrate, consider the famous "Perfect Island" parody of the ontological argument from Gaunilo

- P1*: By *definition*, a "perfect island" is an island that is perfect in every way.
- P2: Existence is a perfection.
- C: Therefore, by *definition*, the perfect island exists.

Since the arguments have the same structure, they are either both valid or both invalid. If they are both invalid, then the ontological argument fails. If

they are both valid, then if the ontological argument is legitimate, the parody version of the argument must have a false premise. The false premise cannot be P2 since this premise is shared by both arguments.

So, it must be P1*. It seems the theist must say that the property of *being perfect in every way* or *being the greatest* can apply to God but not to islands. The question is why that is the case. Let's consider two possible explanations and replies from the atheist.

First, the theist might argue that greatness is sometimes *multidimensional* rather than univocal: some objects admit of greatness only in a non-unique way. In such cases, there is no single greatest object, but several objects that qualify as greatest relative to different perfections. Basketball players, for example, may count as the "greatest" in different ways depending on whether one emphasizes scoring, defense, playmaking, or how they contribute to their team's overall success. In the case of islands, there is no single perfect island but there is a single perfect being (God).

The atheist might grant that greatness is sometimes multidimensional, while insisting that the theist has not shown why it should be univocal in the case of God. Consider two kinds of athletes: a sprinter in the 100-meter dash and a basketball player. In sprinting, one might plausibly argue that *greatness* is univocal: it belongs to whoever runs the fastest (the individual who clocks the fastest time). This is because success in sprinting is measured along a single dimension, even if the abilities required for sprinting are themselves diverse. Basketball, by contrast, offers a much stronger case for thinking that *greatness* is multidimensional, since its standards of excellence are more varied and sometimes non-overlapping. A player might be considered great because of elite scoring, even if their team loses; another might be considered great because of exceptional passing and playmaking, even if they score far fewer points.

If multiple basketball players can each qualify as "greatest" by exemplifying distinct and non-overlapping excellences, then there is reason to think the same is true of persons--or beings--more generally. If so, the theist cannot simply assume that maximal greatness must converge in a single, uniquely greatest being. The ontological argument instead points to a more polytheistic conclusion.

Second, the theist might say that some objects admit of perfection (an upper limit) while others do not. Persons (or beings) admit of an upper limit and so the property of *being the greatest* applies to them. However, islands do not admit of maximal perfection and so there can be no perfect island since there is always a better possible island, e.g., more palm trees, softer sand, etc.

Exercise 2.7

Consider whether there can be an absolutely perfect (or greatest) version of the the following kinds of entities or whether there can always be a ``better" version of that entity.

1. A sprinter competing in the 100-meter dash
2. A student
3. A song
4. A book or video game
5. A stuffed animal

Let's consider two responses from the atheist. First, the atheist can again question what reason there is for accepting that perfection applies to persons (beings) but not to the perfect island. If there is no perfect island because for any possible island, there is always a more perfect version of that island (e.g., sandier, more palm trees), then why isn't the same true of persons (beings) in general?

Second, the atheist may grant that their specific example of a perfect island was poorly chosen. But consider a variation on the perfect-island parody argument, one involving a perfect multiple-choice exam:⁵

- P1*: By *definition*, a "perfect score" on an exam is an exam that has all correct answers (100%).
- P2: Existence is a perfection.
- C: Therefore, by *definition*, a perfect score on an exam exists.

Again, the conclusion does not follow from the premises since while a perfect score on an exam is possible, it does not follow that for a certain exam that someone has achieved a perfect score.

2.2 GOD DOES NOT EXIST

Suppose that arguments for God's existence are unconvincing. The theist might accept this fact, claiming that there can be no rational argument for God's existence. This alone does not entail that the theist should give up their belief in God for they might argue that we have countless beliefs for which we have no rational proof. A theist might say "while reason does not support God's existence, it does not act against it."

Arguments against the existence of God aim to provide rational support against God's existence or the belief in God's existence.

⁵...The parody does not need to be formulated in terms of islands. It can, for example, be formulated in terms of scores on a test for which there is a maximum possible score".[92]

2.2.1 *Argument from Disappointment*

Theists sometimes point out that they believe in God because God has revealed himself to them. There are two primary ways that God reveals himself to people: via general revelation and special revelation.

General revelation refers to God revealing himself to people through the created universe. For example, someone might see the beauty or regularity in nature and interpret this as a mark or indicator of God. While people often will say they “experience” or “see” God in nature, they are discovering God's existence through observation and reason.

Special revelation refers to direct communication between God and inhabitants of the world. Examples include seeing God, hearing God speak, or sensing that God has commanded them to do something. Other examples include experiencing a sudden feeling of peace in a state of grief, feeling strength in a state of crisis, finding meaning in tragedy, or feeling powerful support in a state of weakness.

Let's focus our attention on special revelation. While general revelation is available to all (evidence via general revelation is universal), God does not provide special revelation to everyone. Let's assume that general revelation is not sufficient to show that God exists (that is, while the natural world is perfect, theists often claim that we can easily misinterpret the marks of God in nature). If that is the case, then the only other way one would know that God exists is through special revelation. But, if a person does not receive special revelation, then they could not know God exists. This is problematic since many claim that knowing that God exists is a good thing (some even say it is the greatest good). Therefore, God is engaging in a particular cruelty to people by denying them a good (or the greatest good) and there is no reason for certain people to believe that God exists.

- P1: If an individual has been given no visible sign of God's existence, then they have no reason to believe in God since the primary reason to believe in God is because of personal revelation / experience of God.
- P2: Certain individuals have been given no visible sign of God's existence.
- IC: Therefore, they have no reason to believe in God's existence.
- P3: It is claimed that knowledge of God is a good thing (God provides practical guidance (purpose), moral guidance, and inner peace).
- C: God chooses not to reveal himself to certain people (therefore depriving them not merely of a good but the greatest good) and there is no reason (for some people) to think that God exists.

2.2.2 *Argument from Inconsistent Revelation*

The following argument against God's existence is sometimes called the “avoiding the wrong hell problem”. It essentially argues that revelations that God exists point to the existence of different Gods, e.g., the Roman Catholic God, the Methodist God, YHWH, Allah, Ahura Mazda (the Zoroastrian God).

Since these gods are not the same God, either God does not exist, or there is no rational way to choose, or believing in God based on revelation runs the risk of believing in the wrong God.

- P1: The existence of God and God's laws are revealed through general and special revelation (e.g. beauty in nature, direct communication, miracles).
- P2: Revelations are *inconsistent* with each other and there is no *rational* ground to resolve these revelations, e.g., Christians believe that Jesus is the Messiah, while Jews and Muslims don't. Even the acceptance of our own “religious” experience involves rejecting the “religious” experiences of others.
- C: Therefore, we cannot rationally choose to believe in one God (or set of God's laws) over another.

One objection to the above is that the argument commits the *inflation of conflict fallacy*. Just because individuals in a field disagree about *P* does not mean that (1) there is no fact concerning *P* or (2) there is not some underlying agreement about *P*. In the case of the above argument, while *specific* revelations might conflict, the underlying agreement in all of the revelations is that God exists.

2.2.3 *Dysteleological Argument*

Argument has several names, including “argument from poor design”, “argument from incompetent design”, “argument from unintelligent design”. This argument is often attributed to Haeckel *History of Creation*.

- P1: If God exists, is omnipotent, omniscient, omnibenevolent, and created the world, then organisms in the world would be designed perfectly (optimally).
- P2: But, organisms in this world are designed suboptimally.
- C: Therefore, God does not exist.

Some support for P2:

1. Ectopic pregnancies (fertilized egg implanted in Fallopian tube, cervix, or ovary rather than uterus; leads to death of mother / fetus without

- surgical intervention)
- 2. common mis-formation of spinal column (scoliosis)
- 3. Nearly all genetic and congenital disease
- 4. Crowded teeth / wisdom death
- 5. Chlorophyll absorbs green light poorly, making plants green. But if plants were black, they would be able to absorb more light

Objections

- O1: The world and everything in it is designed optimally! We are constantly finding the purpose of things we thought were useless (e.g., appendix turns out to play an important role in the immune system and perhaps fetal development).
- O2: God can exist but God is either not omnipotent, omniscient, or omnibenevolent.

2.2.4 *Argument from Entropy*

Imagine you are living in an apartment in a city. You have a window that overlooks the city. As you sit in your air-conditioned apartment, you turn away from your computer and stare out the city at night. You see the city lights, cars moving around, a hospital, and a baseball stadium. These events require the use of concentrated energy, e.g., the power plant burning fuel, engines combusting gasoline, and even human bodies ingesting food and expelling waste (e.g., heat, disordered waste). The great order, complexity, and functioning of the city and you require the consumption of energy.

The second law of thermodynamics says that in a closed system, entropy tends to *increase* over time. Roughly speaking, "entropy" is a measure of the degree to which energy is spread out and unavailable for useful work. A system with low entropy would have concentrated energy that is available to do work. A system with high entropy would have unconcentrated energy that is not available to do work. The second law of thermodynamics tells us that entropy increases, which means that energy becomes more evenly distributed, and this more even distribution means that there is less energy available to do work.

The second law of thermodynamics does not imply that there never can be an increase in a *local* pocket of increasing order. Organisms, ecosystems, and civilizations become more structured but only by drawing on concentrated energy sources and releasing more dispersed energy back into their surroundings. Thus, these pockets of order exist only within a larger process of increasing entropy.

Theists tend to point out that the universe is either perfect as it is or trending toward greater and greater perfection. However, this contrasts with the second law of thermodynamics: physicists tell us that energy will increasingly become more and more distributed. As a result, eventually complex and organized beings like plants, animals, and humans will ultimately cease to exist. As Russell[110, p. 547] puts it, "The cosmos is like a theatre in which just once a play is performed, but, after the curtain falls, the theatre is left cold and empty until it sinks in ruins."

What does this imply about God? Russell[110, p. 547] points out that the fact that the universe is trending toward the extinction of the human race implies that God is evil. He writes, "A man who commits a murder is considered to be a bad man. An Omnipotent Deity, if there be one, murders everybody." The argument then is as follows:

- P1: The second law of thermodynamics entails that eventually all life will cease to exist.
- P2: If God created the universe, then it is created in a way that will extinguish human life.
- P3: Anyone who would extinguish all life in this way is evil.
- C: Therefore, either God does not exist or God is evil.

2.2.5 *Argument from conjunctive probabilities*

Conjunctive probability, also called "joint probabilities", is the probability that two events both occur. If A and B are events, then the conjunctive probability of A and B refers to the probability of both A and B occurring.

Care must be taken in determining the conjunctive probability of two events by determining whether the events are independent events or dependent events. But, for our purposes, we will ignore this consideration and treat all events as independent from each other.

Events are *independent* of each other if the probability of one event does not change the probability of the other event. Suppose you are in one room and I am in another. We roll a die at the same time. The probability that I will roll a 6 and the probability you will roll a 6 are independent since your roll has no influence on my roll, and vice versa.

If two events are independent, you can calculate the conjunctive probability by multiplying their individual probabilities. Suppose consecutive flips of a coin are independent. Since the probability of the coin landing on heads for each flip is $1/2$, we can calculate the probability of coin landing on heads twice in a row as follows:

$$P(A) \times P(B) = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$$

Or, to use our dice example, we can calculate the probability that we both roll a 6 by multiplying the probability of me rolling a 6 by the probability of you rolling a 6.

$$P(A) \times P(B) = \frac{1}{6} \times \frac{1}{6} = \frac{1}{36}$$

What can we take away from the above? What we see is that we evaluate a conjunctive probability (the probability A and B will both occur) by multiplying the probability each will independently occur.

In contrast to the formal methods for determining probability like those detailed above, people (in their daily lives) rely upon their "intuitions" or availability heuristics.^[136] These intuitions are sometimes misguided. A quick example. Tversky & Kahneman^[137], asked individuals about how many seven-letter words of the form "`_ _ _ _ i n g`" are in four pages (2,000 words) of a novel. They then asked individuals about how many seven-letter words of the form "`_ _ _ _ i n _`" are in four pages (2,000 words) of a novel. Note that the latter form includes the former form so the number of words of the form "`_ _ _ _ i n _`" should be greater than or equal to "`_ _ _ _ i n g`". Tversky & Kahneman reported that median estimates were 13.4 for *ing* words and 4.7 for *_ n _* words.

With that example in place, let's consider a case involving conjunctive probabilities. Consider the following:^[137]

"Linda is 31 years old, single, outspoken, and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations."

Next, rank the probability of each in order of most likely to least likely:

1. Linda is a bank teller (T)
2. Linda is active in the feminist movement (F)
3. Linda is a bank teller and is active in the feminist movement (T&F)

Tversky & Kahneman^[137]^[137] found that 88% of respondents (N=88) ranked the probability of the conjunction to be greater (more likely) than at least one of the constituents of the conjunction. For example, respondents might claim that F is more likely than $T \& F$. In sum, individuals are susceptible to the "conjunction fallacy".

Definition 2.6: conjunction fallacy

The conjunction fallacy is a type of fallacy where an individual reasons that $A \& B$ is more likely than A or B independently.

What does any of this have to do with God? First, according to a 2022 Gallup poll, 81% of Americans respond “Yes” to the question “Do you believe in God?”. This result was down from 87% in 2017. In the same 2017 poll, Gallup found that 64% of adults were *certain* God exists, 21% said that God *probably* exists, 6% said God probably does not exist, and 7% were certain God does not exist. For the sake of argument, let's assign some relatively high probability to the proposition *God exists*.

Next, let's call the set of beliefs *about* God an *auxiliary religious belief about God*. Some examples might include:

1. God is all-good.
2. God is all-powerful.
3. God is all-knowing.
4. God created the universe.
5. God has a plan for human beings.
6. God prayers of human beings.
7. God directed such-and-such person to write these words in this religious text.

As far as I'm aware, there is minimal data about the confidence levels of people concerning their an auxiliary religious belief about God. For example, while few Americans believe the Bible is the literal word of God (24%), there does not appear to be data that indicates how certain people are about this view.^[111] Nevertheless, it seems reasonable to assume that a person's confidence about their auxiliary religious beliefs about God is *lower than* their confidence that God exists ($P(\text{Auxiliary}) < P(\text{God})$). For example, a person may be absolutely certain that God exists but only think it is highly probable that God is all-powerful or hears their prayers.

Let's suppose someone is almost certain that God exists. They are willing to assign a probability of $\frac{99}{100}$ to God's. Next, suppose that they have 30 auxiliary beliefs about God. For the sake of simplicity, let's assume that the probability of each of these beliefs is a coin-flip: $\frac{1}{2}$. Treating each of these probabilities as independent, the probability of the theist's view being true can be expressed by the following equation: $\frac{99}{100} \times (\frac{1}{2})^{30}$. The probability that God exists *and* that all of their auxiliary beliefs are true is 9.22×10^{-10} or 0.000000092175%. Alternatively put, the odds of a person's belief in God and their beliefs about

being true is 1.08 billion to 1.

- P1: The probability that a person's beliefs in and about God are true are extremely small.
- P2: It is irrational believe things that are extremely unlikely.
- C: Therefore, the theist ought to either abandon *a large portion of their auxiliary beliefs about God* or *be agnostic about them*/

A few criticism. First, the main problem with the above argument is that it does not prove God does not exist. It only shows that it is improbable that *all* of our beliefs about God are true.

A second problem concerns a person who is absolutely certain about *all* of their religious beliefs. That person calculates their beliefs as follows: $(\frac{1}{1})^{31}$.

2.2.6 Bayesian argument against God

TBA

2.2.7 Argument from Russell's Teapot

The following argument comes from a paper titled "Is There God?" by philosopher and logician Bertrand Russell.^[110] The argument appears in different forms and by different names. It is sometimes called the "Argument from Invisible Pink Unicorns" and the "Argument from Flying Sphaghetti Monsters (Occam's Razor)".⁶

Russell points out that many people put the burden of proof on the skeptic: it is the skeptic's job to prove that God does not exist rather than the theist's job to prove that God exists. Russell claims that this is mistaken and supports this claim with the following analogy:

"If I were to suggest that between the Earth and Mars there is a china teapot revolving about the sun in an elliptical orbit, nobody would be able to disprove my assertion provided I were careful to add that the teapot is too small to be revealed even by our most powerful telescopes. But if I were to go on to say that, since my assertion cannot be disproved, it is intolerable presumption on

⁶van Inwagen^[140, pp. 15–16] points out that it while "many atheists regard the "teapot argument" as a knockdown argument for atheism, it is not easy to discover *what* argument this knockdown argument is." I agree with van Inwagen, but I take a different approach in the presentation of Russell's argument in that I ignore the many different arguments mixes together (e.g., the argument from common consent) and extract what I take to be the core of the teapot argument.

the part of human reason to doubt it, I should rightly be thought to be talking nonsense." [110, pp. 547–548]

Let's clean up Russell's argument here. First, Russell suggests that the following two claims are analogous:

1. Between the Earth and Mars there is a very small china teapot (not visible by our most powerful telescopes) revolving about the sun in an elliptical orbit.
2. God exists.

Second, he contends that the following two claims are analogous:

1. Since you cannot disprove the existence of the teapot, you should not doubt its existence.
2. Since you cannot disprove the existence of God, you should not doubt God's existence.

The choice of the teapot is somewhat arbitrary. We could swap out a variety of other objects: an invisible pink unicorn or a flying spaghetti monster. Putting these two analogies together, we get the following two arguments:

- P1: If you cannot disprove the existence of the teapot (Spaghetti monster, invisible pink unicorn), you should not doubt its existence.
- P2: You cannot disprove the existence of the teapot.
- C: Therefore, you should not doubt that the teapot exists.

Applied to God, we get the following:

- P1: If you cannot disprove the existence of the God, you should not doubt God's existence.
- P2: You cannot disprove the existence of God.
- C: Therefore, you should not doubt that God exists.

With this in place, we can now state how this applies to arguing that God does not exist.

- P1: Even if you cannot disprove the teapot exists or God exists, it is rational to doubt that they exist.
- P2: You cannot disprove the existence of the teapot or God.
- C: Therefore, it is rational to doubt that the teapot and God exists.

There are several problems with the above argument. First, the argument does not disprove that God exists. Second, suppose God exists and reveals himself to Tek (and a few others). While it may be rational for others to doubt God's existence, it is at least plausible to claim that it is rational for Tek (and the chosen few) to believe in God.

2.2.8 Arguments from Evil

There are several different arguments from evil against the existence of God. At its most basic, it raises the question of how God could allow for evil in the world when God could stop it.⁷

- P1: If God is all-knowing, all-powerful, and all-good, then there should not be evil in the world since for any evil there could be, God would know it, would have the power to prevent it, and would have the goodness to want to stop it.
- P2: There is evil / suffering in the world.
- C: Therefore, either God does not exist or God is evil.

A common response is that P1 is false: God's existence does not entail a complete absence of evil. It only implies a complete absence of *gratuitous evil* (evil whose presence is not necessary for some greater good). For example, the goodness of human free will (which results in evil) outweighs the evil it causes.

Here is a revised argument.

- P1: If the God exists, then there should be no preventable events in this world whose moral wrongness is not outweighed by its moral rightness (there should be no unnecessary suffering, senseless evil).
- P2: There are preventable events in this world whose moral wrongness is not outweighed by its moral rightness (there is preventable suffering that God could have prevented but did not to prevent).
- C: Therefore, the God does not exist.

Here are some common objections.

First, one might argue that the above argument does not disprove God's existence. God can exist but is not-omnipotent (lacks power to prevent evil), or is not-omniscient (lacks knowledge to prevent evil), or is not-omnibenevolent (doesn't want to prevent evil).

Second, one might respond that P2 is false. The usual response along these lines is to assert that suffering or evil only appears unnecessary and preventable but it is actually *necessary* all things considered. But this is a tough

⁷For example, Elie Wiesel[150, pp. 33–35] writes “A truck drew close and unloaded its hold: small children. Babies! Yes, I did see this, with my own eyes ... children thrown into the flames. [...] How was it possible that men, women, and children were being burned and the world kept silent? [...] Why should I sanctify His name? The Almighty, the eternal and terrible Master of the Universe, chose to be silent. [...] Never shall I forget those moments that murdered my God and my soul and turned my dreams to ashes.”

pill to swallow for consider a natural disaster or genocide. Given that these are great evils, there must be some even greater good that is not merely made possible but that will be actualized by the existence of such an evil? But what actualized good could ever outweigh the existence of such an evil?

Let's consider the following reasons for this view:⁸

1. Good cannot exist without evil as it is a necessary counterpart to evil.
2. Evil is necessary as a means to obtain good.
3. A universe with some evil is better than a universe with no evil as it allows for the achievement of greater goods.
4. A universe with free will is greater than a universe without free will, but free will is the cause of evil in the world.

First, if good cannot exist without evil, then God is not omnipotent since this would imply that God cannot create a world that is wholly good. Second, let's accept the view that something cannot exist without its negated counterpart. For example, *bigness* cannot exist without *smallness*. This would imply that God's goodness could not exist without the existence of evil, and so God's nature would depend upon the actual existence of evil things. Third, it is not clear that good cannot exist without actual evil in the world. Suppose a world where every object is small. Surely, such a world could exist.

Second, some argue that evil is *needed* for good. This view again implies that God is not omnipotent since God could not create a good world without adding evil into it as a means of achieving that good. Some reply that while God could create a perfectly good world, evil is necessary as a means for human beings to do good. The idea is that human beings, in their stupidity and weakness, need evil events to perform good acts. But this again implies God is not omnipotent since it implies God could not create humans without this shortcoming.

Third, some note that the universe is better (a greater good) than a universe without evil. Mackie notes that there are two versions of this position:

1. A literary or musical work that contrasts evil, discord, suffering against goodness, harmony, peace is often greater than one that does not. The discord adds to the greatness of the whole.
2. That the best possible universe is one of dynamic progress toward the good (a gradual overcoming of evil)

The underlying idea is that God injects pain, suffering, and other forms of evil into the world so that humans can obtain "greater goods". For exam-

⁸The following list and much of the discussion of the list is from [75]

ple, suffering, death, and various calamities and evils allow for higher goods like sympathy, charity, heroism, and the development of character traits that minimize future evils in the world.

A common example of this is the use of evil by God to “test” or “develop” the strength of an individual's belief, one might argue that the suffering is there to test true believers. For example, one inmate at the Auschwitz concentration camp was quoted as saying the following: “God is testing us. He wants to see whether we are capable of overcoming our base instincts, of killing the Satan within ourselves. We have no right to despair. And if He punishes us mercilessly, it is a sign that He loves us that much more...”. [150, p. 45]

The claim then is that pain and suffering are the means to produce sympathy, charity, heroism, and other virtues. This implies that at least some humans lack these virtues or embody the corresponding vices: antipathy (indifference, callousness, or apathy), selfishness (or even malevolence), and cowardice. But if God exists, is omnibenevolent, and omnipotent, then one would not expect to see these vices in the world. But they are. Therefore, either God does not exist or he is not omnipotent or he is not omniscient.

Finally, consider the claim that a universe with free will is greater than a universe without free will and it is free will that is the cause of evil in the world. The central problem with this view is that even if it were true, it only accounts for a small fraction of the evil and suffering in the world. Namely, it fails to explain why preventable *natural* evils exist, e.g., droughts, famines, earthquakes, tornadoes, pestilences, etc.

2.2.9 Paradox of Omnipotence

God is said to be omnipotent (all-powerful): there isn't anything that God can't do! One popular objection to this view is called the “Paradox of the Stone”. According to this argument, if God does exist and is omnipotent, then God (1) could do certain actions that are impossible or (2) God could perform an action that are inconsistent with God being omnipotent. Let's consider some of these:

1. God could draw a square circle.
2. God create a Euclidean triangle whose internal angles do not add up to 180.
3. God could create an inescapable prison
4. God could break the laws of logic

5. God could create a stone so heavy that not even God could lift it.⁹
6. God (who is all good and cannot sin / do evil) could sin.¹⁰

The argument then is the following:

- P1: Assume God exists and is “absolutely” omnipotent.
- P2: If God is omnipotent, then God could create a stone so heavy that not even God could lift it.
- IC1: Therefore, God could create a stone so heavy not even God could lift it.
- C: Therefore, either God does not exist or is not omnipotent.

Next, let's consider some responses to the above argument.

First, the usual response is to (1) accept the argument by saying God is not *absolutely* omnipotent and (2) limit what it means to be “omnipotent” to *possible states of affairs*. That is, God cannot do *anything* but can do *anything that is possible*.¹¹

Another way of putting this response[31, see] is to consider this sentence “God created a round square”. This sentence entails that there exists a possible object *O* that has properties that entail a contradiction (being round and being a square). The theist can assert that God's inability to create *O* does not count against God's omnipotence because an omnipotent being can only create what can be created (what is logically possible to create).

One problem with the above response is that it only seems to deal with *some* of the cases above. For example, limiting God's omnipotence in the above way entails that God cannot draw a square circle (which is impossible) but it does not imply that God cannot create a stone so heavy that God cannot lift the stone. The former is an impossible object, but there is nothing logically impossible about the latter provided we do not already assume God is omnipotent.

As Mavrodes[76, p. 221] puts it, the stone example creates a dilemma:

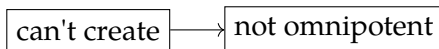
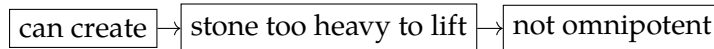
⁹In the *Summa Theologica*, Part I, Q. 25, Art.3, Aquinas[2] writes “It seems that God is not omnipotent. For movement and passiveness belong to everything. But this is impossible with God, for He is immovable, as was said above (Q. 2, A. 3). Therefore He is not omnipotent.”

¹⁰In the *Summa Theologica*, Part I, Q. 25, Art.3, Aquinas[2] considers the example of the stone but also that God cannot sin. He writes “Further, sin is an act of some kind. But God cannot sin, nor deny Himself as it is said in 2 Tim. 2:13. Therefore He is not omnipotent.”

¹¹As Aquinas[2] puts it in the *Summa Theologica* Part I, Q. 25, Art.3, “If, however, we consider the matter aright, since power is said in reference to possible things, this phrase, God can do all things, is rightly understood to mean that God can do all things that are possible; and for this reason He is said to be omnipotent.”

“If we say that God can create such a stone, then it seems that there might be such a stone. And if there might be a stone too heavy for Him to lift, then He is evidently not omnipotent. But if we deny that God can create such a stone, we seem to have given up His omnipotence already. Both answers lead us to the same conclusion.”

Expressing this in a diagram:



In response, the theist can say that the stone example is just like the square circle example in that it is asking God to do something that is impossible. The impossible thing is to create a stone so heavy that not even God can lift it. This poses no problem to God's omnipotence since it is already admitted that God can only do what is logically possible.¹²

Let's consider a second response.^[47] In the previous response, we limited God's omnipotence by saying that God is not capable of performing a task that is contradictory. God can neither draw a square circle nor create a stone so heavy He cannot lift it. Focusing on the stone, let's suppose that God can create such a stone. If God is omnipotent and God can create such a stone, then God can perform a self-contradictory task. God can do the impossible. And, if God can do what is impossible, then surely God can lift the stone as well since this is just a second impossible task. As Frankfurt^[47] points out, “[i]f an omnipotent being can do what is logically impossible, then he can not only create situations which he cannot handle but also, since he is not bound by the limits of consistency, he can handle situations which he cannot handle.” In short, if God can do the impossible, then he can create such a stone and then lift it.

2.2.10 From free will against God

- P1: If God knows everything (omniscience), then he knows everything I will do in the future.

¹²Quite interestingly, Mavrodes^[76, p. 222] notes that “it is the very omnipotence of God which makes the existence of such a stone absolutely impossible, while it is the fact that I am finite in power that makes it possible for me to make a boat too heavy for me to lift”

- P2: If God knows what I will do in the future, then I am not free since if God knows I will do P, then it is fact that I will do P, and I cannot do not-P (the ability to do otherwise is a necessary condition for freedom).
 - IC: Therefore, God is either not omniscient or I am not free.
 - P4: There is more of a reason to believe I am free than there is that God is omniscient.
 - C: Therefore, God is not omniscient.
1. Similar to the argument from omnipotence, one can argue that God is not absolutely omniscient. Instead, God knows all there is to know but the future is unknown. The universe is indeterministic and the future is open.

2.2.11 *Hybrid-Omni Argument*

In earlier subsections, we considered the (1) argument from evil, (2) the paradox of omnipotence, and (3) the argument from free will. Each of these arguments concludes that either “God lacks property *X* or God does not exist.” For example, the paradox from omnipotence concluded that “either God is not omnipotent or God does not exist.” Let’s combine arguments (1)-(3) together in what we will call the “hybrid-omni argument against God’s existence”:

- P1: If God exists, then God must have one of the following properties: be all-good, be all-powerful, or be all-knowing.
- P2: Per the earlier arguments, God either does not exist or is not all-good, not all-powerful, and not all-knowing.
- C: Therefore, God does not exist.

2.2.12 *The No Reason Argument*

Sometimes called the “praxeological argument”.¹³ The argument claims that since God is perfect (complete), God has no reason to do anything (including create the world). So, either there is no God or God is not perfect. Sometimes this argument is used to further argue that God destroys himself after creating the universe (turning himself into fundamental particles).

- P1: For all intelligent beings *x* and acts *y*, if *x* does *y*, then *x* has a reason, need, want, or desire to do *y*.
- P2: If God created the universe, then God must have had a need, want, or desire to create the universe.

¹³Attributed to Scott Adams *God’s Debris*.

- P3: But, by hypothesis, God is omnipotent, omniscient, and perfect (not lacking in anything), and thus has no need, want, or desire to do anything.
- C: Therefore, either God does not exist or God is not perfect.

2.2.13 *Buridan's Ass Argument*

Buridan's ass refers to a paradox concerning rational free choice. Let's suppose that a rational agent is an entity whose choices are always guided by reasons. Since people often have conflicting reasons, their choice is always guided by the strongest reason. For example, I might desire a ham sandwich and a turkey sandwich. I have reason to eat either one, but since I prefer turkey over ham, I have a stronger reason to choose turkey over ham.

Consider a perfectly rational donkey that is dying of hunger. Now suppose the donkey is equidistant between two bales of hay. The donkey has no rational reason to prefer one bale of hay over the other since each bale is identical in every relevant way and they are equidistant to the donkey. In the end, since the donkey is a fully rational agent (the donkey's choice is guided by the strongest reason) and the donkey has no reason to prefer one bale over the other, the donkey is unable to act. The donkey thus dies of hunger.

A similar argument (credited to Clarke against Leibniz) can be applied to God since before God created the world, there were many different possible, but equally good, worlds before God's mind. To simplify things, suppose there are two equally good possible worlds: w_1 and w_2 . If they are equally good and God is perfectly rational, then there is no reason to actualize w_1 over w_2 (or w_2 over w_1). Therefore, either God is not perfectly rational or God does not exist.

- P1: Suppose God exists and is perfectly rational.
- P2: Suppose there are two possible, and equally good, worlds w_1 and w_2 .
- P3: If God is perfectly rational, then there is no reason to create one world rather than the other.
- P4: Our world does exist.
- C: Therefore, either God is not perfectly rational or God does not exist.

Here are two possible responses. First, one could simply accept that God exists but is not perfectly rational. Some actions performed by God are arbitrary. Our world exists not because God had a reason for preferring it over another but because of God's personal whim. There is no logical basis, rational principle for why God chose this world (and why you exist) rather than some equally good alternative (one where you don't exist).

A second response is to deny P3. One could reject the claim that a perfectly rational is always guided by their strongest reason. Instead, one could contend that being perfectly rational involves utilizing other methods for handling those edge cases where reasons compete. In the case of the donkey, while there is no compelling reason to pick one bale over the other, there is a compelling reason to randomly select one bale rather than be deadlocked in a state of indecision and starve to death. In other words, while the default procedure of the rational donkey is to choose based on the strongest reason, the donkey makes use of a rule that handles edge cases where there are equally good reasons for opposing actions, e.g., the use of a random number generator to decide between equally good outcomes. Similarly, the reason God created this world rather than some other equally good alternative is due to chance, but such chance does not mean God isn't perfectly rational. Instead, when faced with two equally good worlds, God takes the rational approach to resolving the deadlock: flipping a coin.

2.3 THE ARGUMENT FROM ABSURDITY

This is my argument for God's existence. It is hated by atheists and theists alike. In fact, I would venture to say it is the most disliked argument for or against the existence of God. Let me first give a short, but incomplete version of the argument before developing it into a more robust version.

- P1: The world and its inhabitants are absurd.
- C: Therefore, God exists.

I'll return to a more detailed defense of P1, but for now, what do I mean by P1? I have two things in mind. First, the world itself is far from an unchanging perfection. It is instead a dynamic mess and largely inhospitable to the continued existence of human beings or their flourishing. Second, not only is the world absurd but so are the people in it. By this I mean that they are extremely stupid and violent.

The above argument is *invalid* (the conclusion does not follow from the premises). To address this problem, let me add a premise.

- P1: The world and its inhabitants are absurd.
- P2: Given this absurdity, it is amazing that we continue to exist.
- P3: The best explanation for our continued survival (non-extinction) or at least there not being more destruction in the world is that God *sometimes* prevents the world or humans from engaging in further destruction that would ensure our extinction.
- C: Therefore, the best explanation for our continued existence is God's existence.

A defense of P1 and P2 involves two parts. First, it involves pointing out the immense amount of violence and stupidity that exists in the world. As a species, humans are amazingly violent, dumb, weak, and irrational. Billions have *unnecessarily* died due to war, genocide, or starvation. We have weapons capable of completely destroying every human being on the planet in a matter of minutes. We not only devise all sorts of ingenious ways to torture our kin but we institutionalize it in the legal system.^[154]^[114] Despite the vast stores of information at our disposal, educational and social policy still keeps the poorest of us in a state of stupor and forces the brightest to the narrowest of technical tasks. Economic and political policy is such that there is no realistic escape from poverty. Rape, murder, assault, robbery, fighting, starvation: all of it is so commonplace and unsurprising that most of us are numbe to it.

The second part of the defense of P1 involves eliciting a sense of amazement. As of 2025, there are 12,241 nuclear weapons in the world. While only two nuclear weapons have been used, isn't it amazing that more of these weapons haven't been used either intentionally or accidentally? Millions of people have been murdered by political regimes, isn't it amazing that the children of parents who were murdered don't exact their vengeance to a greater degree? Healthcare companies wrongfully deny coverage to people suffering with painful and life-threatening diseases. Giving their disregard for human life and suffering (and the general disregard from politicians to correct these issues), isn't it amazing that they don't deny coverage to more people? And, while some people do enact vengeance, isn't it amazing that more people (either the sick or the families of the sick) don't rise up and violently destroy those responsible for these injustices? And on and on and on. The key part of P1 involves identifying some evil part of the world and then considering why that evil or some knock-on evil isn't even worse.

P3 is relatively simple. After one has considered how evil and stupid people are and has added up not only how much worse things could be given *every* evil and absurdity, the only force that would be strong enough to prevent that worse state of affairs is some kind of God.

2.3.1 *Further supporting considerations*

Before considering some objections to this theory, let me point out a few attractive features of this argument.

First, theists do not need to abandon some of the evidence they have traditionally used to support their belief in God. The only catch is that it supports God's existence in a radically new way.

Consider that some theists argue that miracles provide evidence that God exists. On my view, the fact that so many people these so-called miracles is evidence that God works miracles. Consider a relatively mundane miracle reported by a believer: a person delivering bags of food to a hungry family. The person notices that there are only four bags of food for a family that is supposed to receive five bags. They remove four bags from the trunk of their car and deliver the food to the family. Before returning to the car, they pray for a fifth bag. When they return to their car, they check the trunk and, amazingly, a fifth bag of food is in the car. In this scenario, God is believed to have materialized a bag of food out of thin air.

If one accepts the argument from absurdity, the miracle is not that God has materialized a bag of food. Instead, the real miracle is that someone could believe that (1) a bag of food is created out of nothing, (2) be so unobservant to not notice the bag of food that they thought was in their trunk, and (3) not be seriously injured or die in a car crash.

Here is another example. In Mark 8:22-26, Jesus is said to spit on the eyes of a blind man and then lay his hands on his eyes. The blind man's eyesight is then restored. In John 9:6-7, Jesus spits on the ground, makes mud, and the puts mud in a blind man's eyes. After the man washes off, he can now see. Interpreted literally, these passages would have us believe that Jesus was able to cure blind people.

If one accepts the argument from absurdity, the miracle is not that Jesus has cured blind people. Instead, the real miracle is that millions of people could believe (1) a report that a person thousands of years ago cured blindness with spit and mud and (2) have other irrational beliefs that cause them serious injury or death, e.g., not go blind themselves by putting mud in their eyes or by ignoring medical advice from an optometrist.

In addition to miracles, consider that theists often believe that God hears or answers our prayers, and this response to our prayers gives us evidence that God exists. If one accepts the argument from absurdity, the absurdity of prayers is evidence that God does act on the basis of your prayers. However, this does not mean God answers *your prayers*. Instead, God only intervenes in a general way to prevent humanity from completely destroying itself.

A second attractive feature of the argument is that it does not conflict with much of the evidence that atheists point out in support of the claim that God does *not exist*. Atheists commonly point to evil in the world to support the conclusion that no supernatural being could be all-loving, all-knowing, and all-powerful since such a being would prevent the evil in the world. I accept their conclusion whole-heartedly. There is no being that has all of these

properties. However, in accepting their premises, my argument says that the evil in the world (especially the evil in humans) points to some supernatural being that prevents us from destroying ourselves.

2.3.2 *Common objections*

Let's consider a few objections.

First, there is what I will call the "clock-work universe objection". Suppose the universe is a mechanical cosmos. Since such a universe does not leave room for God, no God could interfere and prevent us from destroying ourselves.

To this objection, I have two replies. First, the objection does not preclude God setting up the universe in a way so that we don't kill ourselves. Second, and more importantly, science offers no convincing explanation for the limits to our violence in the technological age.

Second, traditionally, while God might not be omnipotent, God is still extremely powerful. However, if the argument from absurdity were sound, it would indicate that God is rather measly since he is not preventing the immense amount of violence and stupidity in this world.

Third, your argument posits God as the primary force that prevents *more* evil from existing in the world, e.g., human extinction. However, there is another, more plausible hypothesis that explains why humans haven't annihilated themselves: human stupidity. In other words, the reason that humans haven't destroyed themselves is because they are too foolish to enact the right kind of violence on themselves or others for this to happen.

Chapter skipped.

In an earlier chapter, the great chain (or ladder) of being (GCB) was introduced. In GCB, God exists at the top of the chain as the highest (most perfect) being while less complex beings exist below. Toward the top of the GCB exist human persons. This chapter considers what properties an object are necessary (required) and sufficient (enough) to count as a person.

4.0.1 The identification question

It is a widespread belief that persons exist but it unclear what it means to be a person. That is, while children and adults both believe themselves to be persons, it is not clear what properties are necessary and sufficient to be a person. This question is sometimes known as the *identification question*.

Definition 4.1: identification question

What properties must a being have to count as a person?

Answers to this question give a theory of *synchronic personhood*.

Definition 4.2: synchronic personhood

Object O at a time t is a metaphysical person if and only if O has properties F and having F is both necessary and sufficient to be metaphysical person.

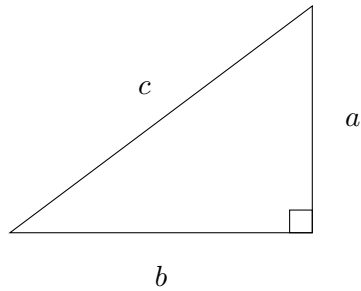
Theories of synchronic personhood will typically face three problems.

First, a theory might be too strict or narrow. When a theory is too strict an object O is a person, but the theory says they are *not* a person. For example, the theory might say that for an O to be a person, O must have such-and-such property. But, O is a person and does not have that property.

Definition 4.3: strictness problem

A theory T is too strict iff (1) O is of kind K , but (2) T says X is necessary for O to count as K but O lacks K .

For example, suppose a definition of a triangle says that in order for a shape to be a triangle, all of the interior angles must equal 60 degrees. Now consider the following triangle:



The above triangle is a triangle even though it is not a triangle *according to the definition*. This is because the definition is too strict (too narrow): it excludes triangles from counting as triangles.

The second problem a theory of synchronic personhood might face is that it is not strict enough (it is too loose). When the theory is too loose, it counts objects that are not persons as persons. It does this because it makes the barrier to entry to be a person too low.

For example, consider the following definition of a triangle: to be a triangle, you only need to be a polygon. Now consider the following figure:



Since the above figure is a polygon, it is (according to the definition) a triangle. But the above figure is not a triangle, so the definition of a triangle is incorrect.

Definition 4.4: looseness problem

A theory T is too loose iff (1) O is not of kind K , but (2) T says X is sufficient for O to count as K but O lacks K .

The third problem a theory of synchronic personhood will face is that it is unclear. If a theory is too unclear, we will not be able to say whether evidence counts for or against the theory because it is unclear what the theory is saying in the first place.

Consider the following definition of a triangle: A triangle is a polygon that has the right number of sides whose interior angles add up to the right number.

On this definition, it is unclear whether a square or a right triangle would count as a triangle since it is unclear what it means for a polygon to have the "right number of sides" or to have "angles that sum to the right number".

Let's summarize. The *identification question of personhood* asks what are the necessary and sufficient conditions for an object to be a person. In other words, what is both required and enough to be a person? Answers to this question give us a *synchronic theory of personhood*. Typical problems with a synchronic theories of personhood are (1) the theory is too strict, (2) too loose, or (3) too unclear to evaluate.

4.1 METAPHYSICAL PERSONS

In this section, we will propose several theories for what are the necessary and sufficient properties an object must have in order to count as a metaphysical persons (hereafter simply *persons*). It is worth noting that not all of the theories we will consider are taken seriously by philosophers, or, if they are, they are only taken seriously in a more developed form. The purpose of considering them is partly to consider some views that people have when they first begin to think about what it means to be a person.

4.1.1 Looking Like a Human

The first theory we will consider is the *humanlike theory of personhood*.

Definition 4.5: humanlike theory of personhood

Object S is a person if and only if S looks like a human being.

This theory has some serious problems.

First, having the properties of looking like a human is *not necessary* for being a person. The theory is thus too strict since it says that certain objects that are persons are not persons.

To illustrate, consider a rational, intelligent, and conscious human being that is disfigured in a horrible accident to the point that they no longer resemble a human being. On the above theory, this individual would lose their personhood on account of a change to their outward appearance. Similarly, people often contend that if super intelligent aliens exist, then they are persons. However, if the super intelligent aliens did not look humanlike, then they would (contrary to our intuitions) not count as persons. Finally, while it is controversial, some people believe that highly intelligent (or rational) animals like dogs, dolphins, cats, and some apes are non-human persons. But, the above theory entails that these are not persons.

Second, another problem the humanlike theory faces is that having the properties of looking like a human is *not sufficient* for being a person. The theory

is thus too loose since it says that certain objects that are not persons are persons.

To illustrate, take a super realistic mannequin or a human-looking robot. On the humanlike theory, the mannequin and the robot are both persons. But, intuitively they are not persons. The mannequin is just like many other inanimate objects in the world: nothing more than fiberglass or plastic in the shape of a human.

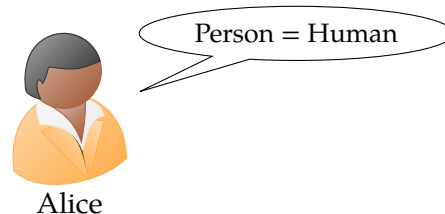
4.1.2 *Being a Human*

The next theory of personhood we will examine will be called the *human theory of personhood*.

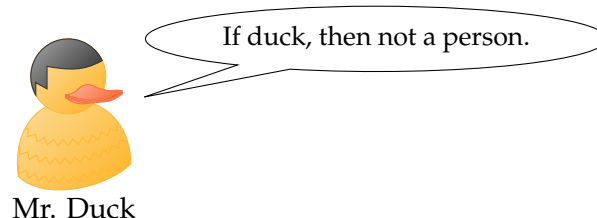
Definition 4.6: human theory of personhood

Object O is a person if and only if O is a human being.

On this theory, you and I and anyone that is a member of *homo sapiens* is a person. In contrast, rocks, tables, chairs, and every other animal is not a human being. This theory improves upon the humanlike theory in that mannequins and robots are not persons. It also improves upon the humanlike theory in that disfigured humans count as persons.



However, the theory faces several problems. First, it is not clear that *being a human* is a necessary condition of personhood. The theory can thus be criticized for being too strict. Generally, it is thought that super intelligent, rational, and self-conscious aliens are persons. In addition, angels, God, and other intelligent and self-conscious beings are taken to be persons. The human theory would not count them as such.



Second, it is not clear that merely being a human being is a sufficient condition of personhood. The theory can thus be criticized for being too loose. A non-functioning human being (e.g. dead body, permanently brain-dead individual, etc.) is a human being but they are often taken not to be persons any longer. Finally, in the context of the abortion debate, it is argued that the human fetus is not a being with a right to life whereas it is universally accepted that newborns (0-3 months), infants (0-12 months), and toddlers (1-3 years) do have a right to life. Some (not all) argue that the latter as opposed to the former have this right to life precisely because they are the type of being that has a right to life.

Exercise 4.8

1. The human theory of personhood is widely held. The most common objection to it is that while being a human is sufficient (enough) for personhood, it is not necessary for personhood. That is, there may be non-human persons. Besides intelligent aliens what other types of beings (if they did exist) would be persons?
2. If being a human is sufficient but not necessary for personhood, is there some trait that humans have that guarantees their personhood? If so, what is it?

4.1.3 Having a soul

It is generally accepted that there are material things that exist in the world. There are tables, chairs, and computers. However, some argue that not only are there material things but also immaterial things. For example, abstract objects (e.g., numbers or geometrical objects) are sometimes said to be outside of space and time and so can be considered immaterial. Another kind of immaterial object are souls. Souls are said to exist for a variety of different reasons. First, there are religious reasons. Second, they may be posited to explain our ability to exist after death. Third, it is claimed that while we can doubt the existence of our own body (which is known through perception), it is not possible to doubt that we exist. This suggests that we exist as a non-bodily object (immaterial thing). In other words, it is claimed that we simply cannot conceive various acts of thinking, willing, doubting and so on without some *being / agent* that is doing these actions. Quinton[105, p. 395] writes:

“A more usual argument for the principle [that there is a soul] starts from the premise that mental states are acts that cannot be conceived without an agent in the same way as there cannot be a

blow without a striker or a journey without a traveler.”

Let's define the soulful theory of personhood.

Definition 4.7: soulful theory of personhood

Object O is a person if and only if O has a soul.

While we will consider the claim that persons are souls in more detail in later chapters, let's point out an initial problem.

The soul theory is unclear because the status of what constitutes *having a soul* is unclear. This results in making it difficult to determine whether an entity is or is not a person.

It is generally believed that human beings have souls. Some religions contend that only humans have souls. Within these traditions, it is unclear whether intelligent aliens have souls. Some eastern traditions contend that animals (and even insects) have souls. More animistic traditions contend that not only do animals have souls, but other natural objects (e.g., plants, bodies of water, mountains) have souls. Finally, more recently, questions have been raised whether artificial intelligent life (some form of embodied artificial intelligence) could have a soul.



Do penguins have souls?

In sum, the soul theory suffers from a lack of clarity about what is a soul. Without a clear description of what a soul is (and how to determine whether an entity has a soul), the theory is too unclear to say whether it is true or false.

Exercise 4.9

1. Create a table with three columns labeled: (1) Has a Soul, (2) No Soul, and (3) Uncertain. Under each column, list the entities that belong in that category. Use some of the entities mentioned in this section as a guide.
2. Try to give a definition of a soul by completing the following: Object O has a soul if and only if ...

4.1.4 Self-consciousness

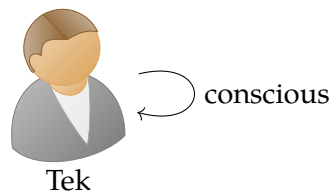
Locke writes that one of the necessary characteristics of a person is that a person “can consider itself as itself”.^[72, p. II.xxvii.11] Let’s call the view that to be a person is to be “self-conscious”, the “self-conscious theory of personhood”.

Definition 4.8: self-consciousness theory

Object O is a person if and only if O has the ability to consider themselves as themselves. That is, they are consciously aware that they are an object that is consciously aware.

Let’s clarify this theory further. First, let’s suppose that a being is “conscious” provided they are cognitively aware of some object O , whether that object be a physical object or an abstract object like a circle. Being conscious does not imply being self-conscious since an object can be aware of something without being aware of themselves.

Second, let’s suppose that a being is “self-conscious” if and only if the object O that they are cognitively aware of is themselves. So, for example, Tek would be *conscious* if he believed he was aware of a lion in front of him. And, Tek would be *self-conscious* if he was aware of himself (the being that is conscious of the lion). Sometimes, it is said that self-consciousness requires an experience of a *private self*.

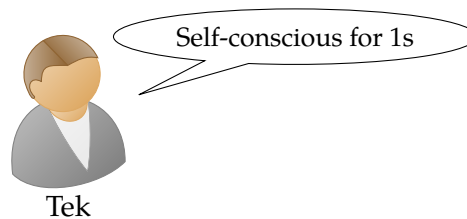


Third, the self-conscious theory does not say that in order for an object O to be a person, O must be self-conscious at every moment of their existence. For example, if Tek is sleeping, it is not the case that he becomes a *non-person* and then, when he wakes up, he transforms back into a person. Rather, the theory only asserts that an object O is a person if and only if they are *capable* of self-conscious.

One strength of the theory is that it seems to identify a property that is *necessary* of personhood: any entity that is not self-conscious would not qualify as a person. Amy Kind^[66, p. 15], for example, points out that we would *not* regard highly-functional (intelligent) robots as persons: even though they are capable of answering questions better than humans (or playing Jeopardy).

We would regard these robots as non-persons because they lack an awareness of themselves. That is, even though they mirror the rational capacity of human beings and display a great deal of intellect, the fact that they lack self-consciousness entails their non-personhood.

However, one problem with the theory is that it is unclear whether *being self-conscious* is *sufficient* for personhood. The theory is thus criticized for being too loose: it says objects are persons that are not persons. Kind suggests that an entity whose sole property is being self-conscious is not a person.



That is, if Tek is a being that simply reflects upon Tek's existence for a single second (so lacking the powers of memory, communication, etc.) we would be reluctant to call Tek a person. So if self-consciousness is not sufficient for personhood, personhood must require a more expansive set of properties. The question then becomes: what other properties are needed for personhood?

Exercise 4.10

1. Is being capable of having emotions necessary for personhood?
2. To be a person, one must be aware of one's self. But, we generally think that people can have degrees of self-awareness. How much self-awareness do you think is sufficient for personhood?

4.1.5 Being Rational

The next theory of personhood we will examine will be called the *rational theory of personhood*.

Definition 4.9: rational theory of personhood

Object S is a person if and only if S is rational.

This theory immediately raises the following question: what does it mean to be rational?

Definition 4.10: rationality as abstract thought

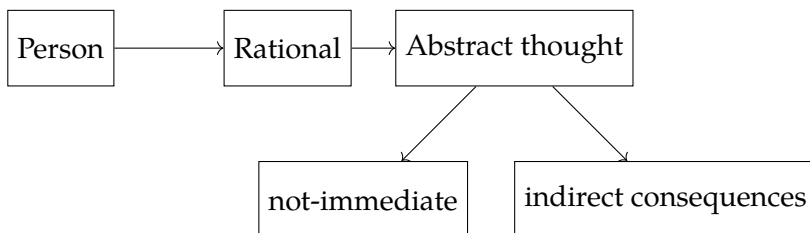
Object O is rational if and only if O is capable of *abstract thought*.

Two important points are worth making. First, in saying that O is rational, this does not mean that O is *always* rational. Take a person who sometimes makes the less-than-optimal choice. While the person's choice may be irrational, this does not mean that the person is irrational. In other words, to be "rational" is to be capable of acting rationally not to always be acting rationally.

Second, the above definition clarifies rationality in terms of the ability to produce abstract thought. This raises the question of what it means to be capable of abstract thought? Let's say that O is a being capable of abstract thought if and only if O is capable of having ideas of things (or events) that are:

1. beyond those ideas which are formed immediately from sense perception
2. beyond those ideas which are about the direct consequences of ideas formed immediately from sense perception

As an example, of (1), abstract ideas would include mathematical ideas (e.g., numbers, circles), the idea of the future, and the idea of possibilities. In terms of (2), some examples would include long-term consequences of certain actions as well as the indirect consequences of certain actions.



All of this is somewhat wordy so let's illustrate the above with a rather wicked example.¹ Suppose a dog is being beaten by its master. Let's also suppose the dog is not a person. As such, the dog is not rational and not capable of abstract thought. This does not mean that the dog does not have ideas for the dog is aware of being beaten while the beating occurs. In addition, the dog has the idea that it will feel pain when it sees its master's fist moving toward it. In this example, the dog has ideas about what it senses along with ideas about the direct consequences of what it senses. However, the dog does not have any ideas of any greater abstractness. For example, the

dog does not have ideas about *why* it is being beaten, or whether it is a good or bad dog (or that its master is evil), or what tactics it should take to prevent future beatings.

On the rational theory of personhood, the following beings are not persons: rocks, tables, chairs, vehicles, plants, and perhaps a good portion of organic life. In contrast, the majority of human beings, certain aliens, angels, and God are persons. The theory thus appears to select a property that (1) excludes beings that intuitively are not persons from being characterized as persons and (2) it also characterizes those beings that are intuitively persons as persons.

One problem with the above approach concerns the claim that *animals* are not rational since they are not capable of abstract thinking. Animals are said not to be capable of abstract thinking since (as our definition indicates) they do not have ideas beyond what they immediately sense (or their immediate consequences).



Mr. Quacker

But this is not correct since as several authors have pointed out, animals do appear to have abstract ideas. For example, van Inwagen[142, p. 136] points out that a dog has the idea of being *beaten by one's master* and this is an abstract idea since it is an idea of something (a universal) that can happen again but not precisely in the same way. To illustrate, suppose a dog has been beaten with a newspaper. Now the dog fears being beaten again in exactly the same way but also in different ways, e.g., with a stick or the master's open hand rather than the newspaper. The dog's fear that the master is capable of future beatings or a variety of different kinds of cruelty suggests that the dog is capable of abstract ideas since it is able to conceptualize a variety of different possible future affairs.

In response to this objection, van Inwagen[142, p. 136] suggests that rationality "does not consist simply in the capacity for abstract thought" but instead "consists in the capacity for a certain *kind* of abstract thought." Let's define this alternative approach to defining rationality as the *capacity view*

Definition 4.11: rationality in terms of capacity

Object O is rational if and only if O has certain cognitive capacities C .

The immediate question then is what cognitive capacities does a being have to have in order to be rational. Let's consider one answer from van Inwagen[142, pp. 136–137]. For van Inwagen, a rational being is a being that has the following capacities:

1. representing to itself a complex state of affairs (including non-actual states)
2. believing that some states are actual and others are non-actual
3. desiring that certain states are actual while others are non-actual
4. contemplating that states without considering whether they are actual or non-actual
5. recognizing logical and causal relations between states of affairs
6. sorting states of affairs into “probable” and “improbable”
7. assigning relative values to states of affairs, e.g., eating a salad is better than eating these chips.
8. devising plans that draw on its beliefs about (1) what is actual, non-actual, probable, and improbable and (2) logical and causal relations between these states to attempt to cause a state of affairs it values to become actual.
9. recognizing other beings as having the above traits
10. communicating to other beings (with the above traits) various facts, orders, questions relating to the states of affairs it is capable of representing to itself
11. communicating to other beings (with the above traits) an assortment of cognitive attitudes (beliefs, values, desires) one has to various states of affairs it is capable of representing to itself

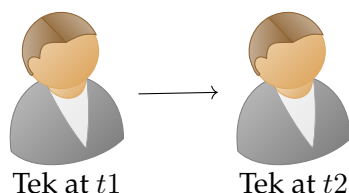
In short, on the above account, an object O is a person if and only if they are rational. An object O is rational if and only it has the above cognitive capacities.

Given the above definition of personhood, what beings are persons? On van Inwagen's[142, p. 139] account, the only beings who are rational and uncontroversially exist are *humans*. However, there are a variety of other beings that would be persons if they exist. Some of them include: “God, angels, gods, elves and fairies and trolls, rational extra-terrestrial beings and rational computers or robots.”[142, p. 139] In addition, on his account, chimpanzees, dolphins, and other animals are not rational beings.²

Exercise 4.11

1. First, select one of the traits in van Inwagen's theory of rationality. Next, clarify this trait with an example. Finally, specify what types of beings have this trait.
2. According to van Inwagen[142, p. 137], "rationality marks a great divide, a discontinuity between humanity and the beasts." Do you agree with this statement, why or why not?
3. van Inwagen notes that since human beings are likely the only rational beings that exist, we are metaphysically unique and cosmically alone since there is no other being like us in the universe. van Inwagen[142, p. 141] writes "it may well be that we are absolutely alone in the cosmos". Similarly, Wilson[153, p. 1110] writes "[a]s far as we know, humanity is alone in the Universe: there is no definite evidence for the existence of extraterrestrial life". Wilson[153, p. 1114] goes on to say that "[o]ne can only speculate on the effect that contact with an ETC [extraterrestrial civilization] would have on humankind. [...]But the simple knowledge of the existence of ETCs would doubtless have far-reaching effects on our relationship with the Universe---as, no doubt, would be a persistent failure to detect ETCs, as this would confirm our uniqueness." Even though we are likely alone, we have engaged in a search for extraterrestrial intelligence (SETI) since the 1960s. The US government funded SETI efforts (it is now funded by private sources and non-profit organizations) until the 1960s. Putting aside the possibility that intelligent extraterrestrial life might have advanced technology that would improve human welfare, what other benefits would there be to discovering extraterrestrial intelligence?

Let's assume that there are objects in the world, that these objects have properties, and that these objects exist at multiple times. For example, the chair I am sitting on this second is the same chair I was sitting on 1 second ago (I didn't stand up). In addition, let's assume that among the many objects that exist in the world (tables, chairs, animals, and computers), there are metaphysical persons. It is generally thought that persons exist through time. The Tek who loaned me ten dollars yesterday is the same Tek who I am paying back today.



But what makes Tek the same Tek? The Tek at t_1 is not perfectly identical to the Tek at t_2 . Tek may have lost a few hairs due to the stress of me not paying him back. Tek at t_2 is slightly older, has more experiences than Tek at t_1 , and he has undergone several other biological changes. What are the conditions under which a person S_1 at time t_1 is one and the same as person S_2 at time t_2 ? How do we (as persons) exist through time?

5.1 THE PROBLEM OF DIACHRONIC PERSONAL IDENTITY

The question of how we exist through time is the question of diachronic personal identity. In this section, this question is clarified.

5.1.1 *Qualitative and numerical identity*

Before diving into various theories of what allows a person to stay the same person over time, let's clarify two different senses in which two objects are "the same" or "identical". The first sense is *qualitative* sameness / identity. The second is *numerical* sameness / identity.

5.1.1.1 *Qualitative identity*

Let's first consider qualitative similarity.

Definition 5.1: qualitatively similar

Two objects A and B are qualitatively similar relative to a comparison class F iff they share enough properties in F to satisfy being "relevantly similar".

Let's make this definition clearer. First, when we say that two objects are *relevantly similar*, we invoke a *comparison class*. A comparison class F is a set of properties that we take to be relevant for judging whether things are similar. In other words, a comparison class tells you what properties count for determining whether things are similar.

To illustrate, take three different dogs

- Fido: a German Shepherd
- Spike: a German Shepherd
- Rex: a Labrador Retriever

In answering the question "which dogs are qualitatively similar", we need to know the comparison class. That is, we need to know "what properties are relevant for *being similar*?" Here are some examples of comparison classes that might be used to judge whether two dogs are qualitatively similar:

- F_1 : the visible properties of the dog (e.g., how the dog looks)
- F_2 : the abilities of the dog (e.g., whether the dog is trained as a police dog)
- F_3 : the health of the dog (e.g., whether the dog is healthy or sick)

It is important to recognize that two objects that are qualitatively similar relative to one comparison class F_1 may not be qualitatively similar relative to another comparison class F_2 . Suppose that Fido is a household pet while Spike and Rex are trained as police dogs. While Fido and Spike are qualitatively similar with respect to their appearance, Spike and Rex are qualitatively similar relative to their abilities.

Second, for two objects A and B to be qualitatively similar, they must have enough of the same properties in the comparison class to count as being relevantly similar. In other words, there is some threshold of shared properties from that comparison class that must be met for the two objects to be qualitatively similar.

To illustrate, suppose we are comparing dogs and our comparison class is *the visible properties of the dog*. In addition, suppose that the threshold to be qualitatively similar is high. With this in mind, two German Shepherds might be judged as qualitatively similar since they share enough of the same

properties found in the comparison class. While a German Shepherd and a Labrador Retriever might share some of the same properties, they do not share enough to meet the threshold.

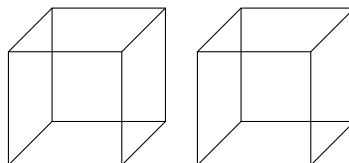
When two objects share many of the same properties, we often say the objects are "the same". For example, imagine children playing with blocks. Two children might say to each other "my block is the *the same* as your block." Or, imagine that there is a blue shirt sold at a store. Now suppose Jon and Tek are friends and Jon buys one on Friday and Tek buys one on Saturday. On Sunday both men wear that shirt and see each other at an event. Jon might say to Tek "we have the same shirt".

Two objects A and B are "qualitatively the same" provided, for every property P in the comparison class F , if A is P , then B is P , and vice versa. When two objects are qualitatively identical for every property P in the comparison class, they are said to be "qualitatively identical" or "indiscernible".

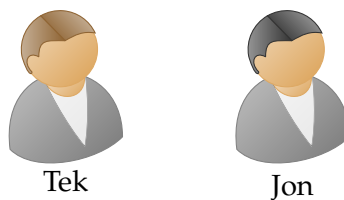
Definition 5.2: qualitatively identical

Two objects A and B are qualitatively identical (the same) provided, for every property P in the comparison class F , if A is P , then B is P , and vice versa.

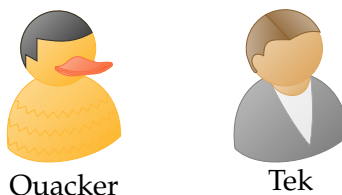
To illustrate, take two blocks A and B that are sitting next to each other. Suppose these two blocks share many of the same properties, e.g. both blue, both cubes, both hard, etc. Since for any property P , if A is P , then B has P , and vice versa, the blocks are said to be qualitatively identical (indiscernible).



Let's consider two key points concerning qualitative similarity. First, qualitative similarity is a *matter of degree*. Suppose Jon and Tek are brothers. They share a lot of properties. Intuitively, we would say that Jon and Tek are similar but not identical.



Contrast this with Jon and his pet duck named Quacker. They share a few properties, e.g., they both like to walk, to look at bodies of water, and to eat crackers. Therefore, they are qualitatively similar. But, Jon and Quacker don't share several important properties. To name just one, Quacker has a bill but Jon does not. So, Jon and Quacker are not qualitatively identical.



Given that Jon and Quacker do *not* share many properties, it seems intuitive to say that not only are they *not* identical but they are also *not* qualitatively similar: they are different in so many ways. One way to resolve this issue is to modify what we mean when we say that two objects are (qualitatively) similar. Rather than saying that one object is or is not qualitatively similar to another object, we can instead specify the degree to which two objects are qualitatively similar to each other. Namely, the degree to which object O_1 is qualitatively similar to object O_2 depends upon the number and significance of the intrinsic qualities that O_1 and O_2 have in common. On this account of qualitative similarity, Jon and Tek are more qualitatively similar than Jon and Quacker.

Second, it should be obvious from our earlier examples that just because two objects are qualitatively similar, it does not follow that they are *one and the same object*. Jon and Tek are qualitatively similar but they are different objects: they are brothers.

Exercise 5.12

1. What object in the world is the most qualitatively identical to you?
2. Suppose there are two objects that are so perfectly (qualitatively) identical to each other that it is impossible for even God to visually tell them apart. If they have all of the same intrinsic properties, are they the same object? What other method could you use to determine that they are two objects instead of one?

5.1.1.2 Numerical identity

In the previous section, we pointed out that two objects can be qualitatively similar but be different objects. That is A and B may be qualitatively similar but $A \neq B$. Let's consider another question. If A and B are *qualitatively*

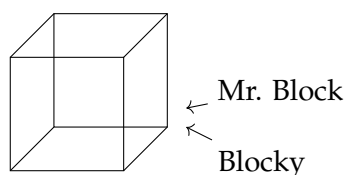
identical, is it necessarily the case that $A = B$ or is it possible that $A \neq B$? To get clearer on this question, let's introduce another concept: numerical identity.

There are several definitions of numerical identity.

Definition 5.3: numerical identity

A is numerically identical to B if and only if $A = B$, viz., they are one and the same object.

Let's try to make numerical identity more understandable. Consider a cube that has two names "Mr. Block" and "Blocky".



While there are two names, both names pick out *one and the same object*. When we say "Mr. Block is Blocky" or "Mr. Block is identical to Blocky", what we mean to say is that the two names pick out *one and the same object*.

Let's consider a few points of clarification regarding numerical identity.

First, the concept of numerical identity is related to counting. Consider the block named 'Blocky' and 'Mr. Block'. We have two names, but there is only *one* object. Similarly, suppose Tek enters a room where he sees a pencil on the floor. He takes there to be exactly one object in the room (ignoring the walls, floor, etc.). He may come up with several names for the pencil, but there is just one object wholly occupying the space of the pencil.

Second, if we fix our attention to objects at an instant of time t , numerical identity is taken to entail qualitative identity. That is, if there is an object A and an object B , and A is numerically identical to B , then for every property F , A is F if and only if B is F . More simply, if $A = B$ (there is just one object), then any property A has B has, and vice versa.

Third, still fixing our attention to an object's identity at an instant of time t , it is unclear whether qualitative identity entails numerical identity. That is, suppose there are two spheres A and B that are qualitatively identical. If they are qualitatively identical, then for every property P , A is P if and only if B is P . Does this entail that there is just one object (so $A = B$) or is it possible for there to be two objects ($A \neq B$)?

It is unclear how to exactly answer this question. To see why, try to picture what two spheres that are qualitatively identical might look like? If spatial location is a property, then A and B would be in the same location. So, our question would then be whether there can be two distinct objects A and B entirely overlapping in space: is it possible for $A \neq B$ even though A and B wholly occupy the same location?

If spatial location is not a property of an object, then A and B could be qualitatively identical while A and B could be in different locations. Our question then would be whether it is possible for A and B to be qualitatively identical, A and B to be in different locations, and $A \neq B$.

Fourth, it is not always immediately obvious *to us* whether two objects are numerically identical. A famous example involves naive astronomical observations. For example, suppose I look up in the sky in the morning and see a shining star. I call it the "morning star". In the evening, I look up at the starry night and see a shining star. I call it the "evening star". At this point, it is not clear to me whether these two objects are numerically the same object. However, later I find out that what I thought was a star wasn't a star at all, it was a planet, it was Venus! The morning star and the evening star are numerically identical. That is, they were one and the same object.

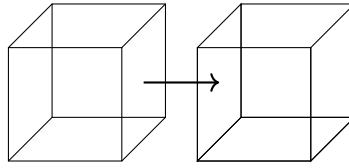
Exercise 5.13

The game of Go is played with black and white stones. The stones are bi-lens-shaped (biconvex) and it is extremely difficult to tell any two similarly-colored stones apart. For any two stones, are the stones qualitatively similar, qualitatively identical, or numerically identical? Explain your answers.

5.1.2 Diachronic identity

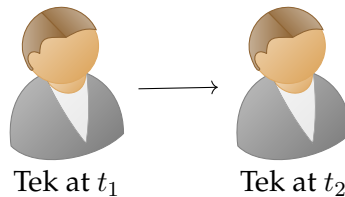
Now that we have introduced the concepts of qualitative and numerical identity, let's consider the concept of *diachronic identity* before turning to the concept of *diachronic personal identity*.

When discussing numerical identity, it was asserted that if two objects A and B are numerically identical at a time t , then they are qualitatively identical. However, consider an object's identity not at a time t but at two different times: t_1 and t_2 . Can an object A that exists at t_1 be numerically identical to an object B at t_2 ? For example, suppose you are looking a block A sitting on a table at time t_1 . You close your eyes for a split second and see a block B at time t_2 sitting on the table.



Is block A numerically identical to B ? Is it one and the same block? At least in the above example, it seems that $A = B$ because nothing about A has changed. B just is A but at a later time.

However, it is also claimed that an object A at one time t_1 can be numerically identical to an object B at another time t_2 even though A and B are not qualitatively identical. That is, an object can survive change. Suppose Tek is sitting on a chair at time t_1 and, one second later (at t_2) he sneezes.



Even though Tek at t_1 and Tek at t_2 are not exactly the same in every way (some qualities in Tek t_2 are different, e.g. at the cellular level), there is often said to be just one object Tek that at one point occupied time t_1 and another time occupied t_2 .

The question of an object's numerical identity (being one and the same object) through time and through change is the question of *diachronic identity*.

Definition 5.4: diachronic numerical identity

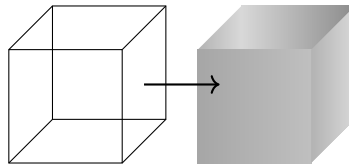
Object A at t_1 is numerically identical to object B at t_2 (where $t_1 \neq t_2$) if and only if $A = B$, viz., they are one and the same object.

One helpful way to remember this is that "dia" comes from the Greek meaning "through" while "chrono-" comes from the Greek "chronos" meaning time, so "diachronic identity" concerns identity through time.

We can ask about an object's diachronic identity in several different ways:

1. how does an object stay one and the same object over (through) time?
2. how do objects persist through time?
3. given an object O_1 at time t_1 and an object O_2 at time t_2 (where $t_1 \neq t_2$), what are the necessary and sufficient conditions such that O_1 is numerically identical to O_2 ?

Let's consider a few additional illustrations of diachronic identity. Consider a cube. Now suppose we shade that cube with a pencil.



Our shading of the cube occurs over time. Let's specify different stages of the shading process as follows:

1. The cube is unshaded at time t_1
2. The cube is partially shaded at time t_2
3. The cube is fully shaded at time t_3

At times t_1, t_2, t_3 , we believe that there is a *single* cube that undergoes change. Its properties change from unshaded, to partially shaded, to fully shaded. Intuitively, we believe that an object can gain and lose properties (e.g., changing color) and stay the same one and the same object. But surely it cannot lose all of its properties and stay the same thing. For example, if we manipulated the lines of the cube (transforming it into a cylinder), it would no longer be the same thing.

The question then of diachronic numerical identity here concerns what sorts of changes can the cube undergo and still be the same cube?

To further illustrate, let's a cellphone under a pair of scenarios.

1. Scenario 1: I drop the phone and the screen breaks. I then take my phone (with the cracked screen) to be repaired. My phone is repaired and no longer has a cracked screen.
2. Scenario 2: I drop the phone and the screen breaks. The phone is under warranty and I am given a new phone.

In Scenario 1, it seems natural to say that the phone I now own is my original phone. In Scenario 2, it seems natural to say that the phone I now own is *not* the original phone. The question then of diachronic numerical identity here concerns what sorts of replacements to the parts of a thing can a phone undergo and still be the same phone?

Exercise 5.14

1. Create a pair of examples using the same object that illustrates diachronic identity. In one example, show how certain changes would result in the object no longer being the same object. In the other example, show how an object could survive changes to its properties or parts but still be one and the same object.

2. Consider the following two stories. Story 1: Suppose there is a pencil on your desk. You pick it up and put it down. How many pencils are there in this story? Story 2: You reach into a bag and pull out a pencil. It breaks. You reach into your bag again and pull out a pencil and continue writing. How many pencils are there in this story?
3. Suppose you have a five dollar bill on the table. Now suppose I pick up your five dollar bill, take a red pen, and mark it with my initials. You complain that I took your five dollars. In response, I say “this is not your five dollar bill. Two objects are numerically identical if and only if they have exactly the same properties, but your five dollar bill did not have my initials on it, while this five dollar bill does have my initials on it. Therefore, your five dollars and my five dollars are different objects.” What metaphysical assumption (what fact about how objects exist in our world) am I denying? Or, how would you respond to what I am saying?

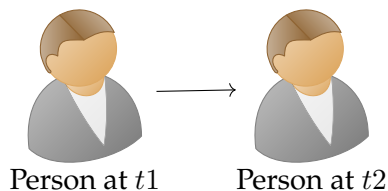
5.1.2.1 Diachronic personal identity

Now that we have illustrated the question of diachronic identity, let's formulate the question of *diachronic personal identity*. The question of diachronic personal identity takes two different forms (the narrow form and the broad form):

Definition 5.5: Narrow question of diachronic personal identity

If A is a person at time t and B is a person at time t_i (where t and t_i are different times), what are the necessary and sufficient conditions such that A is numerically identical to B ?

Let's consider a more concrete version of this question. Suppose there is a person at time t_1 and another person at time t_2 (where $t_1 \neq t_2$):



We can ask “what makes a person the same person over time?” [66, p. 27] Or, if we bump into some woman at an airport, we can ask “Is this tired-looking woman I’ve just bumped into in the Dallas airport the very same person as

the energetic young girl who was my bunkmate at summer camp”?[66, p. 27]

A broader version of this question is the following:[90, p. 25]

Definition 5.6: Broad question of diachronic personal identity

If A is a person at time t and B is an object at time t_i (where t and t_i are different times), what are the necessary and sufficient conditions such that A is numerically identical to B ?

The difference between the narrow and broad question is the following: The broad question leaves open the possibility that a person A at a given time t could survive as a *non-person* at some other time t_i . For example, the broad question leaves it open whether you could survive as something that is not a person, e.g., a rock, tree, some non-human animal that is not a person.

Now that we have formulated the question of diachronic personal identity, let's offer a few clarifying points before trying to answer it.

First, the question is a metaphysical one rather than an epistemological one. Epistemology is a branch of philosophy that concerns the nature of knowledge and how we know something is the case. Categorizing a question as “epistemological” is to categorize it as a question about *how we know something is the case*. While there are epistemological issues concerning the question of diachronic personal identity, the question is a metaphysical one. It asks *what makes it the case* that person A at t_1 and person B at t_2 are one and the same (numerically identical) objects. It concerns what aspects of *reality* determine that these two persons are the same person. The question does not ask *how do we know* that A and B are the same object.

Second, the question of diachronic personal identity is an instance of the more *general* question of the diachronic identity of objects over time. In the more general question, the question concerns what allows *any kind of object* to maintain its identity over time. In the case of the diachronic personal identity, the question focuses specifically on how what allows *persons* to maintain their identity over time.¹

Exercise 5.15

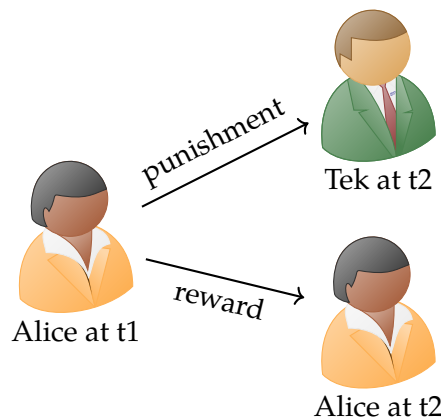
1. Articulate the problem of diachronic personal identity. Supplement your articulation with an example.
2. Could you survive as anything other than a person? Or, do you need to stay a person (or always be a person) to be identical with who you are right now?

3. Suppose that Tek commits a crime. A year passes and he is arrested, tried, and found guilty. When he is being sentenced, Tek pleads with the court, claiming that it would be unjust to punish him as he is *not the same person* as the person who committed the crime. Under what conditions (if any) might it make sense not to punish Tek for the crime? That is, what sort of changes would it make sense not to punish Tek?

5.1.3 Why care? The egoistic concern

Now that we have formulated the question of diachronic personal identity, let's pause for a moment and consider a more practical question: why care about this question at all?

First, people have *egoistic concerns*: a motivation or desire to promote your own self interest over that of at least one other person. To illustrate, suppose that you (Alice) are informed that, in one year, a great reward and a great punishment will be given to either you or your enemy (Tek). You are then asked whether you prefer the reward or the punishment.



If you prefer the reward, then you are concerned about what happens to a future person that you take to be the same person as yourself. In other words, you care about diachronic personal identity because you care about being able to assign goods (and avoid evils) to a future person that you take to be the same person as you.¹

¹Here is how Marya Schechtman describes the "egoistic concern": "we all know the difference between fearing for our own pain and fearing for the pain of someone else. The difference here consists not in *degree*---I may care more about the pain of my beloved than about my own---but in kind."

Let's consider a more realistic example. Today you might forego certain pleasures (food, drink, play) and engage in activities that are somewhat undesirable (e.g., putting money into a retirement account, studying, being faithful in a relationship). Why? You likely forego these pleasures because you believe that your actions today (done by you) will impact a future person. But not *any* future person! You believe that your actions today will have an effect on a future person that you take to be numerically identical to you. So, again, the question of diachronic personal identity matters because you have an egoistic concern and this egoistic concern involves being able to identify whether some future person is the same person as you.

A second reason why the question of diachronic personal identity has practical value concerns *reward and punishment*. Intuitively, if people who commit crimes or do good deeds at one time are to be punished or rewarded at another time (on the basis of these deeds), then we have a concern that we are punishing or rewarding the same person who performed the punishable or rewardable act. Similarly, if a child makes a mistake and we wish to correct (not punish) the child so they don't make the same mistake in the future, it is important to the child (and hopefully to us) that the child who made the mistake is the same child that we are correcting. Correcting the same child should matter if the development of the child matters. And surely it does!

One final example. Suppose Tek does something terrible. It is so bad that Tek deserves serious punishment. Now imagine we apprehend someone who believe to be the same person as Tek. Should we punish the person? The answer to this question surely depends upon whether or not the person is, in fact, Tek. If the person is you, then you will surely yell, plead, and do everything in your power to avoid the punishment that Tek deserves. At that moment in time, nothing will be important to you than to prove that you are not the same person as Tek.

Exercise 5.16

Create an illustration of the egoistic concern.

5.2 OVERVIEW OF SOLUTIONS

In this very brief section, a broad overview of the main solutions to the problem of diachronic personal identity are provided. The goal of this section is not to define any of the solutions in detail but, instead, to supply the reader a general overview of what is to come. There are five main answers to the problem of diachronic personal identity:

1. soul-based solutions
2. physical solutions
3. psychological solutions
4. hybrid solutions
5. skeptical solutions

In what follows, let's briefly characterize each of these solutions. Each of these approaches will be discussed in more detail in subsequent sections.

First, on a soul-based solution, a person is (or has) a soul. The soul is an immaterial and indivisible entity that persists through time. On a soul-based approach, a person A at t_1 is the same person B at t_2 if and only if A and B have the same soul. Soul accounts differ based on the nature of the soul (e.g., does the soul think?) and when it begins (e.g., does it begin at conception?).

Second, on a physical (animalist) approach, a person is their body (or some part of their body). On a physical-based approach, a person A at t_1 is the same person B at t_2 if and only if there exists a specific kind of physical (biological) relation between A and B . A simplistic version of the physical approach is the whole body theory: you stay the same person over time provided you have the same body over time.

Third, on a psychological approach, a person is a thinking thing. On a psychological based approach, a person A at t_1 is the same person B at t_2 if and only if there exists a specific kind of psychological relation between A and B . A simplistic version of the psychological approach is the memory theory: you stay the same person over time provided you can remember earlier experiences you had.

Fourth, on a hybrid approach, we are some combination of a soul, a physical being, or a thinking thing *or* what preserves our identity through time is some combination of soul-soul, physical, psychological relations. For example, a simplistic version of a hybrid approach is what might be called "the conjunctive theory": you stay the same person over time provided you can remember earlier experiences you had (memory) and you have the same body.

Fifth, on a skeptical approach, there is no answer to what preserves our identity (or, that answer cannot be known). On this approach, there is either no fact to the matter as to what whether A at t_1 is the same person B at t_2 , or, perhaps there is a fact to the matter but it cannot be known.

Exercise 5.17

Before considering the various theories of diachronic personal identity, try to situate yourself within one of the theories above. Once you have selected one of the above theories, try to articulate why you think this theory is correct.

5.3 SOUL APPROACHES

Let's introduce the soul approach by articulating several assumptions. Note that as assumptions, we may wish to revise or reject them after we've considered them in more detail. Here are our assumptions:

1. we are souls / spirits
2. the soul is immaterial and distinct from the body (which is material)
3. the soul *always thinks*.
4. bodies are ensouled (acquire souls) at conception
5. the soul is capable of surviving death
6. the soul is the bearer of moral qualities, e.g. wickedness, goodness.
7. the soul is the bearer of personal traits, e.g., virtues, vices, emotional responses (e.g., feeling pity for someone)

Next, let's examine a few of these assumptions.

Proposition (1) asserts that we are souls and (2) says something about the nature of the soul: it is immaterial. In claiming that the soul is immaterial, the soul is thus distinct from the body. The soul is not your brain.

In accepting we are souls, we can also answer the question of diachronic personal identity by saying that to stay the same person over time is to have the same soul at different times. So, if there is a person S_1 at time t_1 and a person S_2 at a later time t_2 , then S_1 is the same person as S_2 if and only if S_1 and S_2 have the same soul.

Definition 5.7: soulful theory of diachronic personhood

If A is a person at t_1 and B is a person at t_2 , then A is the same person as B iff A and B have the same soul.

Proposition (3) says that the essential property of a soul is *to think*. That is, for any object O , O is a soul if and only if O thinks. In other words, if an object O thinks, then it is a soul. If O is a soul, then O thinks.

Recall that (2) says that the soul is immaterial and the body is material. As such, it is not technically correct to say that the brain thinks. Rather, the

brain is capable of electrical activity correlated with different thoughts. In contrast, the soul is said to have the power of thought, and other forms of mental activation: judging, willing, believing, etc.

To clarify this view a bit more, let's say that the soul *always thinks* and if it ever stopped thinking, it would cease to exist. This is a view that some philosophers (e.g., Descartes) can be seen as accepting. For example, Descartes writes, "the human soul [...] is always thinking [...]. [It] cannot ever be without thought; it can of course be without this or that thought, but it cannot be without *some* thought."

Let's restate the soulful theory of diachronic personhood:

Definition 5.8: thinking soulful theory of diachronic personhood

If A is a person at t_1 and B is a person at t_2 , then A is the same person as B iff A and B have the same soul, which is that which has the power to think and is always exercising that power.

With this definition in place, let's consider some reasons in favor of the soulful theory of diachronic personal identity.

Proposition (4) claims your soul is created at conception. Conception (also known as fertilization) is the process of the sperm penetrating the egg. It is recognized as the start of pregnancy. The soulful theory contends that when conception occurs, ensoulment occurs.

Definition 5.9: ensoulment

Ensoulment is the moment at which an object gains a soul.

The object that is ensouled is the zygote (fertilized egg).

5.3.1 Arguments for the soulful theory

There are several reasons why this view is attractive.

First, there is an epistemological argument for the soulful theory. Take the classic argument given by Descartes that while we may doubt whether we have a body, it is impossible to doubt that we exist since the very activity of doubting (thinking) presupposes the existence of a thinking thing. In other words, we know that we exist with more certainty than we know that we have a body. From this Descartes concludes that we are essentially thinking things.

“Just because I know certainly that I exist, and that meanwhile I do not remark that any other thing necessarily pertains to my nature or essence, excepting that I am a thinking thing, I rightly conclude that my essence consists solely in the fact that I am a thinking thing. And although possibly [...] I possess a body with which I am very intimately conjoined, yet because, on the one side, I have a clear and distinct idea of myself inasmuch as I am only a thinking and unextended thing, and as, on the other, I possess a distinct idea of body, inasmuch as it is only an extended and unthinking thing, it is certain that this I [that is to say, my soul by which I am what I am], is entirely and absolutely distinct from my body, and can exist without it.”

Now the thought then is that if we are essentially thinking things (souls), then what is necessary for the preservation of our identity as persons is not the preservation of our bodies but the preservation of our souls.

Second, the theory can account for *strict numerical identity*. Let's suppose we are just our bodies. We know that our bodies are constantly changing. So, from one moment to the next, we are not *strictly identical* since at time t_1 we might have one more skin cell than the t_2 . So, if we are our bodies, while I may think I am *identical* (one and the same) to the person I was two seconds ago, this is not (strictly speaking) true. But, if I am just my immaterial soul (a kind of power source for my thoughts), then, insofar as the soul does not change, I can say (strictly speaking) one and the same person throughout my life.

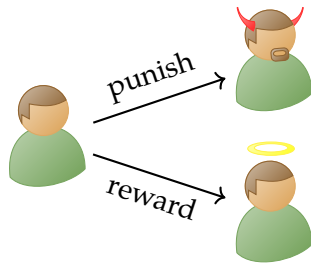
The inability to account for strict (numerical) identity creates a problem for theories of personal identity over time. Leibniz's law states that two objects x, y are identical to each other $x = y$ if and only if x and y have the same properties. But, if Tek at t_2 has undergone some small psychological or physical change between t_1 and t_2 , then Tek at t_2 is not identical to Tek at t_1 . Thus, all talk by personal identity theorists of persons being one and the same person over time is not (strictly speaking) accurate. However, if there is some non-physical, non-psychological and unchanging fact about an individual (e.g. an immaterial soul) then such a theory could accommodate the view that Tek at t_1 is one and the same person (in the strict sense) to Tek at t_2 .

only the further fact
theory can account for
the egoistic concern

Third, the soul theory can explain why our egoistic concerns about post-mortem (after death) are rational.

Suppose you believe in the afterlife. With respect to the afterlife, you likely have concerns about what happens to you in the afterlife. If there are cer-

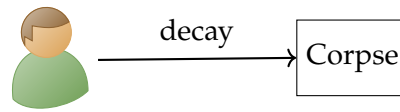
tain rewards (e.g., happiness), you want to experience them. If there are punishments (e.g., torture), then you want to avoid them.



But now suppose we are just our bodies. If we are just our bodies, then these egoistic concerns seem irrational. To see why this is the case, consider what happens to your body after you die (note that times are approximate as they depend upon several factors, especially the environment):

- Immediate post-mortem (first 24 hours): most of the changes occurs in the eyes (segmentation of retinal blood vessels, loss of intraocular pressure, corneal clouding), loss of skin luster and elasticity
- Early post-mortem (3-72 hours): algor mortis (body loses ability to regulate temperature), rigor mortis (muscle stiffening starting with the smallest muscles, e.g., those in the face, moving to the larger muscles, before disappearing), purple-blue discoloration starting as spots before turning into patches due to blood cell disintegration
- Early decomposition (days 1-5): Skin slippage and hair loss, appearance of maggots, brownish extremities, bloating that starts in the abdomen and is later found in other parts
- Advanced decomposition (beginning days 4-10): extensive maggot infestation, loose and sagging skin, skin is leathery, loss of skin, collapse of abdominal (stomach) cavity, body appears brownish-blackish, tissue is slimey and paste-like, development of adipocere (also called "corpse wax"; it is a wax-like (grayish or tannish) substance that is formed by anaerobic bacterial hydrolysis).
- Skeletonization phase (can occur over years): half the skeleton is exposed, adipocere is located on high-fat areas, limited tissue remaining
- Extreme decomposition phase (as early as 2 months to as late as two and half yeas): erosion of skeleton

Egoistic concerns relating to rewards and punishments seem absurd if we continue to survive as decaying corpses because corpses are not the time of thing that can be rewarded with happiness or punished with torture.



But now consider that we are souls. If we can continue to exist as souls, then concerns we may have after death are rational. Intuitively, if I am an immaterial thinking thing then I can be subject to pleasing thoughts or painful ones.

Fourth, the previous argument assumed that the afterlife exists and then argued that rational concern for ourselves in that afterlife requires positing the existence of an immaterial soul. Swinburne puts forward a similar argument that if existence in the afterlife is a *logical possibility*, then it follows that we have an immaterial soul.

To see this argument somewhat clearly, first let's accept that it is logically possible that the soul can be separated from the body.

“a person living on Earth consists of two parts---a material part, the body; and an immaterial part, the soul. The soul is the essential part of a person, and it is its continuing which constitutes the continuing of the person. While on Earth, the soul is linked to a body (by the body being the vehicle of the person's knowledge of and action upon the physical world). But, it is logically possible, the soul can be separated from the body and exist in a disembodied state [...] or linked to a new body.” [116, p. 27]

On the view above, a person can survive the loss of their body and continue on in the form of a soul. While we exist on earth, we are also connected to a body, but when our body dies, we are capable of continued existing as our souls. Next, here is Swinburne's statement that we could only exist as a soul:²

“Given that for any present person who is currently conscious, there is no logical impossibility, whatever else may be true now of that person, that that person continue to exist without his body, it follows that that person must now actually have a part other than a bodily part which can continue, and which we may call his soul---and so that his possession of it is entailed by his being a conscious thing. For there is not even a logical possibility that if I now consist of nothing but matter and the matter is destroyed, that I should nevertheless continue to exist. From the mere logical possibility of my continued existence there follows the actual fact that there is now more to me than my body; and that more

is the essential part of myself. A person's being conscious is thus to be analysed as an immaterial core of himself, his soul being conscious." [116, pp. 29–30]

Here is the argument in condensed standard form:

- P1: If I am now consist only of matter (I'm just my body) and my body is destroyed, then it is not logically possible that I should continue to exist.
- P2: It is logically possible that a person can exist after death.
- C: Therefore, I am (right now) not just my body (I'm a soul).

In this section, we considered an attempt to answer the question of diachronic personal identity by positing the existence of an immaterial soul that emerges at conception. This soul is not a thinking thing nor is it our bodies. The attractiveness of such a theory is that it can account for our continued existence as persons in the strict (numerical) sense since there exists something that stays the same (the soul) throughout our existence. In addition to being able to explain how we are numerically the same, it also is capable of accounting for the concern we have for ourselves (rather than someone that is merely psychologically or physically related to us). In the next section, we consider several criticisms of the soul theory of diachronic personal identity.

Exercise 5.18

1. In your own words, articulate the soul theory of diachronic personal identity.
2. What are some advantages of the soul theory (consider any advantages not mentioned in this section)?
3. While some egoistic concerns are rational, is the egoistic concern about life after death rational? That is, is it rational to be worried about what will happen to you after you die? If so, what does this imply about your continued existence after death?
4. The soul theory is a non-reductionist (further fact) theory in that our personal identity over time does not depend primarily upon psychological and / or physical facts. Construct your own version of a non-reductionist theory.

5.3.2 Objections to the soul theory

Recall the first four assumptions concerning the soul:

1. we are souls / spirits

2. the soul is immaterial and distinct from the body (which is material)
3. the soul *always thinks*.
4. bodies are ensouled (acquire souls) at conception

There are three problems with our assumptions (1)-(4). First, if the soul exists if and only if it is thinking, then what about scenarios where we are not thinking? Intuitively, we are (at least temporarily) not thinking while we are sleeping, unconscious, or in a coma. If the soul is always thinking and we are not thinking in these scenarios, then it appears that whenever we are in these scenarios, we lose our souls. In addition, since our continued existence depends upon our having the same soul, the loss of the soul means that we temporarily go out of existence. This result is highly counterintuitive since I may be deep in thought (and therefore existing) but as soon as I start daydreaming (spacing out), then I no longer exist, but then when I am back in thought, I begin to exist again.

A second problem with this view is that even if there are periods of time when we go in and out of existence because we are temporarily not thinking, the view that the soul always thinks also conflicts with the idea that we begin at conception. As a zygote, it isn't that we have temporarily stopped thinking but it appears that we do not think at all. That is, we do not remember anything about our life inside the womb and current research suggests that consciousness only begins after birth (or very late in a pregnancy: 35 weeks).

A third problem concerns the assumption that the soul is entirely separate from the body. We tend to think that some of the actions of our body are caused by thoughts that we have. For example, suppose I am in my office and want to get some water. Further suppose that there are two different locations for me to fill up my water bottle. I could walk upstairs to a water-bottle refilling station or I could walk downstairs to a water-bottle refilling station. I start to think about which station is better based on which location is more likely to be congested or more sanitary. Suppose I choose to go to the station upstairs and I begin walking in that direction. In this scenario it appears that my thoughts have influenced my body by telling it where to go. But if the soul is wholly separate from the body, how did this occur? If my soul is immaterial, how did this immaterial thing interact with my material body? While there may be good answers to these questions, the soulful theory has some explaining to do.

5.3.3 *The soul has the capacity to think*

In the previous section, we made several assumptions about who we are and used these to give an account of how we persist through time. Our theory faced several problems. Rather than completely abandon the idea that we persist through time by having the same soul, let's modify the assumption that *the soul always thinks*. Instead of saying that *the soul always thinks*, let's assume the following:

- 3* The soul has the capacity to think (it is not the mind but the "basis for the mind")

On this view, while the soul does not always think, it has the power to think. This modification allows for us to say that we (right now as we think) are the same person as the unthinking person that is asleep or temporarily unconscious. Furthermore, it also allows for us to preserve the intuition that we begin at conception. Namely, while the zygote does not think, it has the capacity to think. On this view then, our existence begins at conception. At that point in time, we are unthinking souls with the power (capacity) to think. And, there is a time over the course of our biological development, where we exercise this power to think by actually thinking.

Our modified view of the soulful theory runs into several problems.

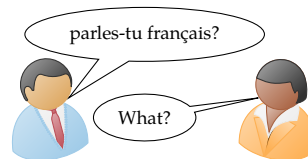
First, in the previous section, we noted that the soulful theory faces the problem of how the soul can influence (interact with) the body if the two are entirely separate (one immaterial and the other material). But this type of interaction problem also works the other way. If the soul is entirely separate from the body, then events that happen to the body should not impact the soul's ability to think. But, as we know, our thoughts and our ability to think are dramatically impacted by events in our body and brain.

Let's consider two cases. First, if the soul (not the brain) is capable of thought, then it is unclear why the soul only begins to think only when the brain is sufficiently developed. Given the regular connection between brain development and the capacity for thought, it is more plausible to say that the brain (not the soul) makes thought possible. Second, if the soul is the basis for thought, then it is unclear why our thoughts are so heavily influenced by our physical condition. If Tek is heavily intoxicated, his use of alcohol not only influences his motor skills and his perception (bodily skills), but his ability to think (cognitive skills). Or, if Tek suffers from a serious brain injury, while we might expect various perceptual or motor skills to be impaired, it is not clear why (if the soul is separate from his body) his cognitive skills would be impaired.

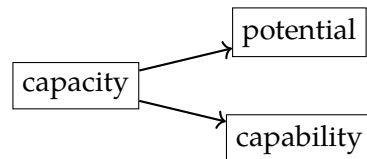
A second problem concerns what is meant by *being capable of thought*. Consider a sleeping person and a human fetus. A sleeping person is capable of thinking in that if we woke up the thinking person, they would have the power to think. A human fetus is capable of thinking in that if the fetus develops for a longer period of time, the fetus would have the power to think. But, some individuals may wish to say that there is an important difference between the sleeping person and the fetus. They might point out that the sleeping person has *the capability to think* while the fetus only has *the potential to think*.

When something has the capacity to do something, nothing about the intrinsic nature of that thing as an object must change in order for it to be able to exercise that power. In contrast, when something has only the mere potential to do something, something about the intrinsic nature of that thing must change in order for that object to exercise that power.

To illustrate the difference between a *capability* and *potential* to do something, consider Tek and Liz. We might say that they both have the capacity to speak French. But this claim overlooks an important difference. Suppose Tek learned French as a child and speaks French daily (it is his native language). In contrast, Liz never learned French but she would be able to learn it if she studied regularly for two years.



Rather than merely saying that they both have the "capacity to think", it is more informative to say that Tek has the capability to speak French while Liz has the potential to speak French. Nothing about Tek needs to change in order for Tek to speak French, while Liz must undertake some grueling study.



Here is another example. Suppose Sam is a professional weight-lifter and Jon is a weakling. To say that they both have the "capacity" to lift 200lbs is ambiguous. Sam has the capability of lifting 200lbs while Jon only has the

potential. Nothing about Sam needs to change in order for them to be able to lift 200lbs while Jon must undertake some physical training.

Exercise 5.19

Create your own example that illustrates difference between someone having the *capability* of doing something as opposed to the mere *potentiality* to do that thing.

What is problematic about saying that in order for an object to be a soul, it only need to have the *capacity* to think? Why is it important to tease out the differences between capability and potentiality when talking about capacity?

First, if *being a soul* only requires having the *potential* to think, then lots of objects have souls. A human fetus would have a soul because it would be capable of developing into an entity that is able to think. In addition, various animals would have souls because even if they do not meet the requirements of being able to think now (or in their mature state), they could undergo various forms of genetic engineering to make them capable of high / more complex forms of thought.

Second, if *being a soul* instead requires having the *capability* to think, then lots of objects that are intuitively regarded as persons are not persons. Take an individual who can no longer think because of brain damage. These individuals do not have souls and are not persons.

Finally, a third problem facing the soul theory relates to ethical consequences. Let's consider two consequences.

First, if being a soul only requires having the potential to think, then this potentially clashes with other ethical views we have about persons. From the ethical side, McMahan[78, p. 17] points out, those who hold the view that the soul thinks (or is always capable of thought) must come to terms with how animals are treated. Since animals have the potential for thought, they have souls. According to the soul theory, if an object has a soul, then it is a person. And, if morally wrong to kill persons, then it is morally wrong to kill animals.

Second, if being a soul requires the capability of thought, then seriously brain damaged individuals have no souls and are therefore not persons. For some, while persons have a right to life and are worthy of respect, non-persons (mere objects) are capable of being killed or used for the benefit of human persons. Therefore, it would follow that individuals no longer capable of

thought may be killed and have their organs harvested since they are mere objects.²

In this section, we consider the soulful theory of diachronic personal identity. We assumed that the soul begins at conception and the soul either thinks or has the capacity to think.

5.3.4 *The soul does not think*

In the previous section, we attempted to modify the soulful theory of diachronic personal identity by replacing the assumption that (1) the soul always thinks with the weaker claim that (2) the soul has the capacity to think. This revision still faced serious problems. Again, rather than abandon the theory, let's try to save the theory by modifying this assumption once again. Rather than saying that the soul always thinks or has the capacity to think, let's assume that the soul does not think.

Always thinks → Capacity to think → Never thinks

The soul is thus neither the mind (since it does not think) nor the basis for the mind (since it is not the capacity for thinking).

Definition 5.10: unthinking soulful theory

If A is a person at $t1$ and B is a person at $t2$, then A is the same person as B iff A and B have the same soul and the soul does not think.

Revising the theory in this way also takes us into problems.

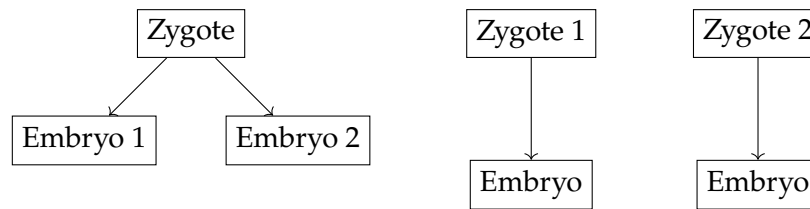
First, if the soul is not the mind (our consciousness) nor the body, then it appears to be nothing at all. It is an empty shell that is responsible for our continued survival but not our thoughts, feelings, memories, and so forth. As McMahan[78, p. 9] puts it “[w]hen defined by negation, as neither body nor mind, it [the soul] emerges as entirely featureless.”

- P1: If X exists then it must have properties, features, characteristics.
- P2: The soul has no properties, features, or characteristics.
- C: Therefore, the soul does not exist.

Second, if the soul begins at conception, then a problem is raised by the existence of monozygotic twins (identical twins). In the case of monozygotic twins, a single zygote is formed during conception. This contrasts with dizy-

²Of course, there are other moral reasons beyond personhood that may morally require us to respect non-persons.

gotic twins (fraternal twins) which develop from separate eggs, separate sperm, and so separate events of conception.



Give our current assumptions, it would follow that if there is only one event of conception, only one soul is created (or is present) and therefore there is only one person occupying (or related to) that zygote. In the case of monozygotic twins, the single zygote splits and forms two embryos, which may ultimately lead to the birth of two distinct human beings (the “identical twins”). Although relatively rare (3 in 1000 births), there appears to be no good explanation of how a single event of conception can account for the formation of two persons. Here are some possibilities:

1. There are two souls at conception but they both occupy the same body. Problem: This allows for two people to occupy a single body.
2. There is only one soul at conception. When the zygote divides into two embryos, there is one soul that occupies two bodies. Problem: This means that the identical twins are not distinct persons since they don't have their own souls. They are just parts of a single person.
3. There is only one soul at conception. When the zygote divides into two embryos, the soul only goes with one of the bodies. Problem: This implies one twin is a person while the other is a “soulless automaton”[78, p. 19]

Third, intuitively we might take *we who are* to have something to do with our *psychology* (e.g., what we believe, our perspective, our beliefs, our memories, our consciousness). But if the soul is not conscious, then we could exist (for millions of years) and not be aware of it (or be aware of anything). For example, suppose Tek is born, becomes conscious, but immediately loses consciousness. His body is then kept (artificially) alive for a million years. According to the soul account, Tek is a million years old. But, this is a highly counterintuitive result.

Fourth, if the soul is not our mind (our psychological states), then this allows for several of what we will call *wild possibilities*. Let's consider a few.

The first possibility is that it is possible for a single soul to occupy the multiple different consciousnesses (minds, psychologies) at different times with-

out any one consciousness having knowledge of any of the others. Alternatively put, if the soul does not think, then it is possible for a single soul to have a series of distinct minds (a consciousness with memories, a personality, etc.) to occupy *what we take*. Here is how Locke[72, Bk.II, xxvii, 21] puts it:

“it must be allowed possible that a man born of different women, and in distant times, may be the same man. A way of speaking, which whoever admits, must allow it possible, for the same man to be two distinct persons, as any two that have lived in different ages without knowledge of another's thoughts.”

Let's consider some different illustrations of how a distinct soul might have psychologically distinct lives. The first is that we can imagine a single soul inhabiting a human being. Once the human being dies, the soul then moves to a new human being. On this account, the second human would be the same *person* as the first person (since they have the same soul) but they would not have any knowledge of the first human. Locke writes,

“if the *identity* of soul alone makes the same man, and there be nothing in the nature of matter, why the same individual spirit may not be united to different bodies. [...] [I]t will be possible, that those men, living in distant ages, and of different tempers, may have been the same man: which way of speaking must be from a very strange use of of the word *man* [...]. And that way of speaking would agree yet worse with the notions of those philosophers, who allow of transmigration, and are of opinion that the souls of men may, for their miscarriages, be detruded into the bodies of beasts [...]. But yet I think no body, could he be sure that the soul of *Heliogabalus* were in one of his hogs, would yet say that hog were a *man* or *Heliogabalus*.”[72, Bk. 2, xxvii, 6]

Let's illustrate Locke's argument more concretely. Suppose that the person is the soul. Now let's take the following individuals:

1. Heliogabalus (a Roman emperor), who lived from c. 204 - 222 and was the subject of various religious controversies (e.g., replacing Jupiter as the primary god, marrying a Vestal Virgin) and ultimately assassinated.
2. Ockam (a Franciscan friar and scholastic philosopher) who lived from c.1287- 1347, who is probably best known for his version of the ontological argument for the existence of God.
3. Georg Cantor (German-Russian mathematician) who lived 1845-1918

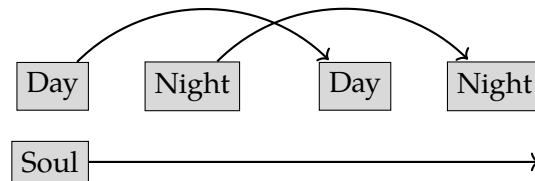
and is widely known for his contributions to set theory and that there are there are different sizes of infinity.

4. Rover: a collie (dog) living today.

According to Locke, if the soul is not tied to any individual matter (and if the soul does not think), then it is possible that the same soul occupied Helioabalus from 204 - 222, Ockam from 1287-1347, Cantor from 1845-1918, and now occupies Rover today.

Here is another variation of this argument. Let's call this *Lockes's Nightman Argument*. Suppose that a single body (call it "Tek") has two distinct consciousnesses. The first consciousness occupies Tek during the day (the Dayman) and the second consciousness occupies Tek during the night (the Nightman). Let's further suppose that neither has knowledge of the other's thoughts or remembers the other's experiences. According to Locke, we would take there two to be two distinct people occupying a single body.

Let's further suppose that the soul does not think (it is not conscious) and can survive without being consciousness. Locke contends that if our personal identity were accounted for in terms of the soul, then it would be possible to say that there is one person instead of two. That is, Tek has one soul: it is conscious of the Dayman's life during the day, then forgets everything about this in the night, and is now conscious of the Nightman's life during the night, then forgets everything about the Nightman's life, and recalls everything about the Dayman's life the next day, and so on.



So, intuitively, we might take there to be two persons in one body. However, on the soul account, there is only one person (since there is only one soul) with two distinct consciousnesses. Since this gives us the wrong result, there is further reason to doubt the soul theory of personal identity.

Exercise 5.20

1. Create your own "wild possibility".
2. Even if the soul exists, is it responsible for personal identity over time?
3. Let's assume that you are a soul and the soul theory is correct. Now suppose your body dies but you need a body to be conscious. What

would this imply about the afterlife?

5.3.5 *The soul does not begin at conception*

Skipped.

5.3.6 *Hylomorphic views of the soul*

Skipped.

5.4 PHYSICAL APPROACHES

Our bodies distinguish us from a great many other objects in this universe. In contrast to tables and chairs (non-living things), our bodies are *living things*. We are different from other living things in that we are complex (multicellular) organisms. Our bodies also possess parts and some of these parts possess a distinct nucleus (eukaryotes). We are also heterotrophs in that we cannot produce our own food but instead depend for our food on other sources of organic carbon (e.g., plants, other animals). Now while our bodies distinguish us from many other beings, it is unclear whether (1) we *are* nothing but our bodies and (2) whether having the same body over time is what preserves our identity. Nevertheless, let's explore the possibility that what preserves our identity over time is having the same body.

5.4.1 *The body theory*

The body theory asserts that what makes us the same person over time is having the same body.

Definition 5.11: Same body theory

If A is a person at t_1 and B is a person at t_2 , then A is the same person as B iff A and B have the same body.

To illustrate the same body theory, suppose Tek is hungry. He buys a sandwich, eats it, and is no longer hungry. In this scenario, Tek is the same person as he was before he ate the sandwich. This is because he has the same body. Or, suppose Liz commits a series of crimes. She is arrested, tried, and convicted. She serves her time in prison and is released. In this scenario, Liz is the same person as she was before she committed the crimes. This is because she has the same body.

The main problem with the body theory is that it does not specify what it

means to *have the same body*. First, having the same body cannot mean that the body A at t_1 and the body B at t_2 are qualitatively identical since our bodies are always changing. Tek's body before eating a sandwich is different than Tek's body after the sandwich is digested. Intuitively, the same body theory must have an account of what it means to *have the same body* but that body is capable of change.

If the same body theory allows for bodies to change while staying the same body, it needs to specify the *kinds* of changes that a body can undergo and be the same body. After all, if our bodies perserve our identity, we want to know what types of changes we should avoid so that we do not perish.

To illustrate, consider the following changes a body might undergo:

1. replacement of an old skin cell with a new skin cell.
2. loss of an arm
3. replacement of an arm with a robotic arm
4. Gradual replacement of 49% of your cells through a natural process
5. Gradual replacement of 49% of your cells through an unnatural process
6. Gradual replacement of 100% of your cells through a natural process³
7. Rapid replacement of 100% of your cells through an unnatural process
8. Gradual replacement of 100% of your cells with non-human cells through an unnatural process

Do we survive some or all of the above? The body theory does not tell us.

Let's refine the body theory into a new theory that we will call the *biological connection theory*. Let's lay out the features of this theory before we state the theory.

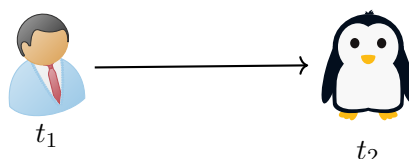
First, let's insist that bodies must be spatiotemporally continuous. That is, if A is Tek's body at t_1 and B is a body at t_2 , then A is the same body as B iff there is a continuous physical connection between A and B .

A few examples. Suppose Tek is in New York at t_1 , it would not be the case that he is the same person as another body in Paris at t_2 if there was no spatial continuity between the two bodies. That is, the person would not be Tek if

³With respect to our bodies undergoing a complete replacement of its parts, Flanagan writes: "It is obvious to the naked eye that humans go through radical physical transformations on the way from birth to old age and death. What is not so obvious to the naked eye, although it helps explain these transformations, is that our bodies are continually cashing in old cells for new ones. Now suppose your average, everyday body cell has a life span of seven years (this is probably on the high side). If the average life of your average cell is seven years, then the average person has ten different bodies in a lifetime." [44, p. 17]

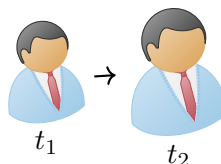
Tek did not travel from New York to Paris. Alternatively, suppose goes to sleep in his bed one night and then wakes up in the morning. The Tek that went to sleep is the same Tek that wakes up in the morning because there is a continuous physical connection between the two bodies.

A second restriction is that in order for a body A to be the same body as body B , both bodies must be of the same kind. To illustrate why this is important, suppose Tek is a human at time t_1 and we gradually replace Tek's cells with the cells of a penguin (or a scrambled egg). Suppose at time t_2 (the end of the process of replacement), Tek's body is entirely composed of penguin cells.



Based our new restriction, Tek would not survive as the penguin since the body at t_2 is not the same kind of body had by Tek at t_1 .

A third restriction on what it means to 'have the same body' is that bodily transformation is gradual (non-rapid). For example, suppose Tek stands 5'11 tall and grows to 6'3 over two years. All other things being equal, since the bodily transformation is gradual, Tek has the same body.



In contrast, suppose Tek is a human, six-feet tall, had brown hair, brown eyes, no tattoos, and his original set of organs. Now suppose that Tek enters machine that quickly transforms him into a person that is ten-feet tall, has blue eyes, green hair, tons of tattoos, and quickly replaces all of his organs. Tek is still the kind of being (a human) but the changes to his body are so rapid that there is the temptation to say he has a 'different body'.

With these clarifications in place, let's improve upon the body theory with a theory that we will call the *biological theory of diachronic personal identity*.

Definition 5.12: the biological theory

If A is a person at t_1 and B is a person at t_2 , then A is the same person as B iff (1) there is a continuous physical connection between A and B (spatiotemporal continuity), (2) A and B are of the same kind (same species), and (3) any change from A to be B is gradual in nature.

Exercise 5.21

Test the body theory in the following cases. For each case, answer whether the person in the scenario, determine whether S refers to one body or different bodies.

1. In the following example, S has a human body and is mostly the same throughout (minor changes): S appears in 2000 in Philadelphia, S disappears in 2021, S reappears in 2025 in New York for 1s, S disappears and then reappears in 1924 in South Africa for 1s.
2. S is a human being. S is transformed into a zebra, then a dog, then back into a human.
3. S at t_1 is a human being. S at t_2 has all major organs replaced (included the brain).

Create an example that violates each of the three restrictions on the biological theory. That is, create an example where you have a person but they undergo a change where:

1. There is no spatiotemporal continuity (e.g., a gap)
2. They no longer have the same kind of body
3. The change is not gradual

5.4.2 Arguments for the biological theory

Let's consider some arguments in support of the biological theory.

5.4.2.1 The argument from common speech

Let's call the following argument, the *Argument from Common Speech for the Body Theory*. Consider some of the ways in which we talk about ourselves.

1. If you punch me in the arm (my body), you have not touched my psychology or soul. Rather, you have punched *me*.
2. "I'm going to the store" means "My body is going to the store".
3. If you hugged *me*, it isn't the case you haven't hugged me because you haven't hugged my psychology, my mind, my consciousness, or my

soul. Rather, I say you hugged *me* when you hugged *my body*.

It appears that whenever we talk about how other objects interact with our bodies, we can substitute “me” for “my body”. If my body has a medical operation, then it is *me* that is having that operation. Alternatively, whenever we talk about how we interact with the world, we can substitute “my body” for “me”. If I walk to the store, then it is *my body* that is walking to the store. So, in sum, we talk about ourselves as though we are our bodies.

The argument then can be phrased as follows:

- P1: We talk about ourselves as though we are our bodies.
- P2: In addition, when talk about ourselves as persisting, we express it in terms of our body persisting.
- P3: How we talk about things tracks reality.
- C: Therefore, we are our bodies.

Exercise 5.22

The Argument from Common Speech for the Body Theory assumes that talk about our bodies can be replaced with talk about ourselves (and vice versa). Can you think of any counter-examples to this assumption? How might someone who supports the argument respond to these counter-examples?

5.4.2.2 Location argument

Let's call the following argument, *the Location Argument*. Suppose Tek and Liz are both busy professionals. They are in a committed relationship but since they are both constantly traveling for work, they often spend time apart. One day, Tek is in New York and Liz is in Paris. Tek calls Liz and asks her where she is. Liz responds “I'm in Paris! It is beautiful here! Where are you?” Tek responds “I'm in New York.” Presumably, it makes sense to say that an object *O* has a spatial location in the world if and only if *O* is a physical object. It makes no sense to say that an invisible, immaterial unicorn is sitting next to me since unicorns are not physical objects. So, given that we can locate ourselves in the world, it is plausible to think that we are physical objects (our bodies).

5.4.2.3 Grounding argument

Let's call the following argument *the Grounding Argument*. Our bodies are the basis for our psychological states. That is, our psychological states are

grounded in our bodies. In other words, our psychological states are dependent upon our bodies such that there is no change to a psychological state without a corresponding change to the body. For example, if I am in pain, then there is some physical state of my body that corresponds to that pain. If I am happy, then there is some physical state of my body that corresponds to that happiness. Another consequence of our psychology being grounded in our bodies is that we cannot exist without our bodies. Given that our psychological states are so dependent upon our bodies, it is plausible to think that we are our bodies.

Exercise 5.23

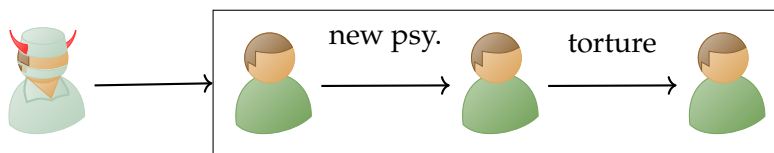
The mere fact that X depends upon Y does not imply that X is Y. A house may depend upon a foundation, but this does not mean that the house is the foundation. Similarly, a statue made of clay may be entirely constituted by the stuff that it is made out of (e.g., clay) but this does not mean that the statue is the clay. Given this, how might someone who supports the Grounding Argument respond to this objection?

5.4.2.4 Mad scientist argument

Let's call our first argument the *Mad Scientist argument*. Suppose Dr. Frank kidnaps Tek. He tells Tek that Tek will undergo the following procedure:

1. he will eliminate all of his psychology (memory, habits, etc.),
2. give him a new psychology (memory, habits, etc.), and then
3. torture him for seven years.

That is,



Next, let's suppose that we are just our psychology (our thoughts, feelings, memories, and so forth). If Tek is his psychology, then he will die after Step 1. Therefore, it would be rational to be worried about Step 1. It would not be rational to worry about steps 2 and 3 since these are events that do not happen to Tek (since Tek is just his psychology).

The mad scientist argument claims that it is rational to worry about step 3. What this indicates then is that we are our bodies. We are worried about step

3 because step 3 involves damage (harm) to you. Let's put this argument in argument standard form:

- P1: If you are your psychology, then it is only rational to worry about yourself in step 1 of the mad scientist procedure (not steps 2 or 3).
- P2: If it is rational to worry about step 3 of the mad scientist procedure, then you are your body (this is because step 3 harms you if and only if you are your body).
- P3: It is rational to worry about yourself in step 3 of the mad scientist procedure.
- C: Therefore, you are your body.

Exercise 5.24

Evaluate the Mad Scientist argument. Are you the worried about Step 3?

5.4.2.5 Simple fetus argument

Let's call the following argument, *the simple Fetus Argument*. [cf. 91] Some people talk about themselves as though they have had continued existence from the time they were a fetus to now. The body theory can accommodate this intuition. Suppose Liz has a child named Renna. Renna asks if she was ever in Liz's belly. Liz shows her a sonogram, points to a fetus (sometime prior to 24 weeks), and tells Renna that "there you are". Intuitively, we think Liz is stating the truth: Renna was once a specific human fetus. However, a fetus prior to 24 weeks lacks any significant psychological capacities so the fact that Renna today is the same person as the fetus cannot be explained by a psychological connection between Renna today and the fetus at an earlier date. Consider even further that the development of the human brain does not begin until the third week of pregnancy. [129] Now suppose that Renna points to the fetus on a date prior to the development of the brain. Again, if Liz says to her "there you are", she is not saying something false. But, given that Renna has no brain at that point in time, continued brain activity cannot explain how Renna today is the same person as the fetus at an earlier date.

So, what does explain why Renna is the same person as the fetus inside Liz? As we saw, there are powerful reasons to deny that Renna is a soul. An alternative solution is to say that Renna as a fetus (at t_1) and Renna today are the same person because there is biological continuity between them. Or, more simply, they have the same body. Let's express this argument more explicitly:

- P1: Renna is a person.
- P2: Renna was once a fetus.
- IC: Therefore, Renna as a person is identical with Renna as a fetus.
- P3: The fact that Renna is identical to a fetus cannot be explained by a psychological connection or having the same brain.
- P4: This fact can be explained by Renna today being biologically connected to said fetus.
- Therefore, what preserves our personal identity over time is biological continuity (having the same body).

Exercise 5.25

Were you ever a human fetus? Why or why not? If your mother pointed to a picture of a fetus and said 'that is you', would your mother be mistaken?

5.4.2.6 Olson's Fetus Argument †

In the previous section, several arguments were presented in support of the biological theory of diachronic personal identity. In this section, let's consider a more complex and developed argument in favor of this theory. This argument is a variation on the *the simple Fetus Argument* that we considered in the previous section. Let's call this *Olson's Fetus Argument* since the argument is raised by Eric Olson.[\[91\]](#).

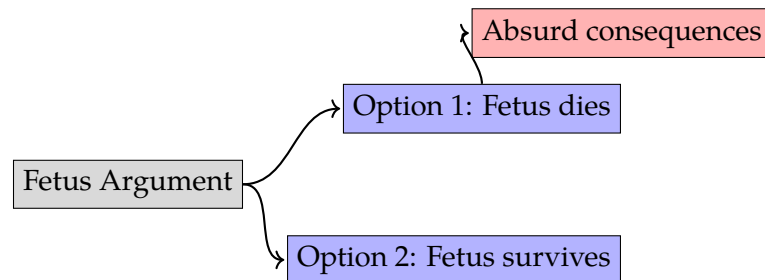
To begin, let's assume that to be a person, you need to have certain a certain level of psychological complexity, e.g., you have to be conscious, to be capable of memory, and perhaps some other traits. If having these psychological traits are required to be a metaphysical person, the human fetus (at least early in its development) is not a person. Now if all of this is true, then the person we are today is not numerically identical with the fetus. That is, we were never a fetus. Instead, what appears to be the case is that at some point concurrent with our organic development as humans (when exactly this occurs is not important here), we appear.

Assume I appear when I become conscious. Given the above account, Eric Olson asks, when the person appears, "what happens to the fetus?" He believes that if we take personhood to be tied to having certain psychological traits, then there are two options.

1. When the person first appears, the fetus (the human organism) dies.
2. The fetus survives as a more developed human organism and it occupies the same space, time, and material as the human person.

Olson rejects both options. On the first option, once the person appears, the fetus (the human organism) does not survive. It fails to continue its development into a baby, child, or human adult. As Olson[91, p. 100] puts it, “the fetus *necessarily* ceases to exist as soon as its nervous system has developed enough to make thought possible. Each human fetus (or infant) must perish in the act of bringing forth a human being.”³

Olson[91, pp. 100–101] characterizes this first option as absurd, saying it “would be one of the most remarkable and one of the saddest facts in all of natural history”. It is remarkable since it would make human beings exceptional since while non-human embryos survive into adulthood, the human organism perishes once it obtains a certain level of psychological complexity. In addition to its exceptional nature, Olson points out that the human organism dying when a person appears is puzzling because it would imply that something dies when it *gains* a power (the power to think, remember, etc.) rather than *lose* a power.



The second option is the more natural one. On this option, the fetus survives and develops into an adult human and the person emerges when the fetus reaches a certain level of psychological complexity. On this second option, there are a few things worth pointing out. First, the person is composed of organic material (the person has a human body). Second, while this body is the same body as the fetus (there is biological continuity between the fetus and the adult human), the person never was a fetus (due to its lack of psychological complexity). Third, the human body (which developed from the fetus) and the person occupy the same space and time. Finally, even though the human body and the person wholly occupy the same space and time, they are distinct for several reasons:

1. the organism is older since it was a fetus
2. the organism can survive in scenarios where the person does not (e.g., permanent loss of consciousness)
3. perhaps the person could survive in scenarios where the organism does not (e.g., your psychology is supported by a new body or artificially).

With these key points in mind, the main consequence of the second option is that (1) there exists a human being that is numerically identical to a human fetus (it once was a human fetus), (2) there exists a person that never was a fetus, and (3) the human being and the person (which are distinct) occupy the same space and time.

Olson finds this second option problematic. First, he[91, p. 101] claims that it entails the odd conclusion “we are not human animals”. Even though we look like an organism, are alive, occupy the same space and time, are composed of animal parts, and are arranged in just the way animals are arranged, we are not animals. Second, he[91, p. 101] also takes this second option to be mistaken because it entails that it is “uncertain whether you and I are people at all.” Let's call the argument rejecting this argument the *Epistemic Argument*.

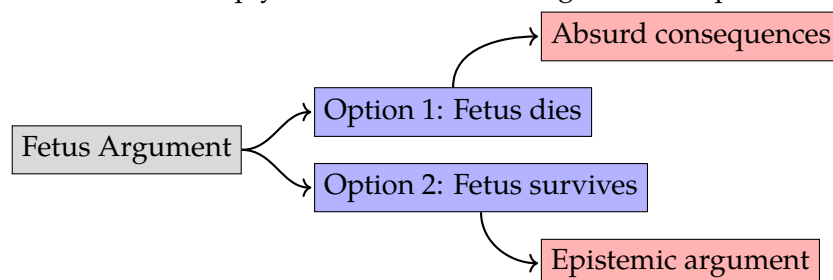
- P1: The human being and the person occupy the same space and time, are composed of the same parts, are composed of the same material, have the same behaviors, and have the same brain states.
- P2: Since the human being and the person are composed of the same material, arranged the same way, have the same behaviors, etc., etc., the human being and the person have the same thoughts (see P1).
- P3: You think you are a person.
- P4: If you think you are a person, then the human being thinks it is a person (see P2).
- P5: On the psychological theory of personal identity, the person and the human animal are distinct and so only one of you can be correct (you both cannot be persons).
- P6: There is no evidence that would decide who is correct.
- P7: A theory of personal identity should not entail the consequence that we don't know whether or not we are persons.
- C: Therefore, we should reject the psychological theory of personal identity.

Parfit[96, p. 7] articulates this argument as follows:

“If there are two conscious beings thinking all my thoughts, the person and the animal, how could I know which one I am? If I think I am the person, Animalists object, I might be mistaken, since I may really be the animal.”

Let me summarize Olson's Fetus argument. If we accept some kind of psychological theory of diachronic personal identity, then we have to answer the puzzling question of whether I was ever a fetus. The psychological theory

seems committed to answering that no person was ever a fetus given that the fetus lacks sufficient psychological complexity. But, at some point the human organism develops to have a sufficiently complex psychology such that we can say a person is present. Once it is on the scene we have the question of what happens to the fetus? It seems absurd to say it dies. On the other hand, if it carries on in the form of a more developed human organism, then we have the puzzling situation where the human organism and the person (which are not the same) occupy the same space and time, are composed of the same material, have the same thoughts, commit the same actions, and think the same thoughts. Since they both think they are persons, it is unclear whether we (who think ourselves to be persons) are really persons. We could simply be the animal thinking that it is a person.

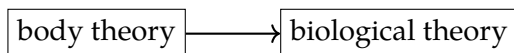


Exercise 5.26

1. In your own words, articulate Olson's Fetus Argument.
2. How does Olson's Fetus Argument differ from the simple Fetus Argument?

5.4.3 Arguments against the biological theory

Recall that we developed the *body theory* into the *biological theory* by offering several clarifications of what it would mean for a person to have the same body over time.



We noted that a person can have the same body over time even if that body undergoes change. In brief, we said that *A* is the same body as *B* (at least for the purposes of preserving our personhood) provided the following conditions hold:

1. Spatiotemporal connectedness: *X* is *Y* only if there is continuous physical connection between the bodies
2. Same kind: *X* is *Y* only if the bodies of *X* and *Y* are of the same kind

3. Gradual change: X is Y only if the change from X to Y (or Y to X) is gradual

In emphasizing biological connection, we can better account for the natural growth and decay of the body that occurs over time,⁴ and for explaining how we can retain our identity while *losing* parts of our body.⁵

Let's consider several objections to the biological theory.

5.4.3.1 From rapid change

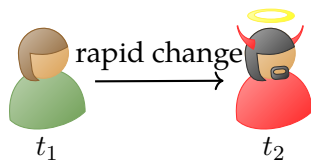
The biological connectedness theory insists that change to the body is gradual. Gradual changes are the normal changes that occur to bodies over time. These would relate to development and decay. For example, as a child I may be 100 lbs and 4'5 but I become someone who weights 160 lbs and is 6'0. The lens in my eye may become less flexible over time, making it more difficult for me to focus on objects that are close to me. This change is not radical enough to constitute a change in my identity (it is not enough to say I have a new body).

But now let's consider the following body modifications:

1. Tattoos and piercings
2. Cosmetic implants: breast / calf implants, subdermal horns, facial fillers
3. Alterations: tongue bifurcation, earlobe stretching, neck elongation, tooth filings, ear pointing (elfin ears)
4. Scarification or branding
5. Organ Transplants: kidney transplant, pace maker
6. Neural implants: cochlear implants, neuralink

Let's suppose that a person can survive any one of the above modifications to their body.

But now suppose that Liz were to undergo all of the above modifications in a single day. That is, Liz gets a tattoo, a haircut, several piercings, and has several cosmetic procedures performed. In addition, Liz has two serious medical operations: a kidney transplant and a heart transplant.

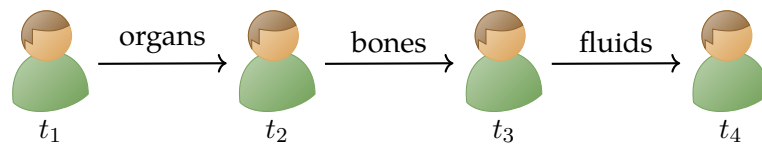


In this case, changes to Liz's body are *not gradual* but Liz's body is still the same body (he survives). Therefore, gradual change is not necessary for having the same body.

Exercise 5.27

What are some other non-natural changes a person could make to their body and they would still be the same person?

However, this result should not be taken to imply that a person can survive any sort of rapid change to their bodies. To illustrate why this is the case, suppose Tek undergoes a full body replacement. Over the course of two days, every organ (heart, liver, kidney, brain), bone, and fluids (blood) is replaced in a successive manner.



Even though there is (1) spatiotemporal continuity and (2) the body at the beginning of the process is the same kind of thing (human) as the body at the end of the process, these rapid changes surely resulted in the death of Tek.

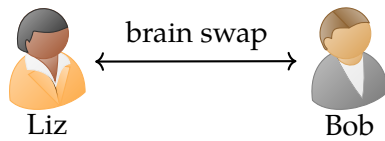
If we are to still accept the biological theory, then it appears we must modify the assumption that biological continuity requires gradual change. Let's revise the assumption as follows:

- (3*) the change is gradual or if the change is rapid, there is no significant modification to the parts of the body that are essential way for preserving a person's identity.

However, the problem with this revision is that our new condition is vague. It does not state what parts of the body are essential for preserving our survival. Let's consider this issue in the next section.

5.4.3.2 From brain transplants

What parts of the body are essential to preserving a person's identity? Or, what parts of the body are such that if we modified in a rapid way would result in our death? To get at what parts are essential to a person, let's consider theoretical brain transplants. While not currently medically possible, suppose that you and a friend were to swap brains. That is, suppose a doctor was able to surgically removes your brain and your friend's brain. Then, the doctor was able to put your brain in your friend's body, and your friend's brain in your body.



If the biological theory only permits the modification of those parts of the body that are not essential to the person's identity (e.g., their finger nails), then the theory must take a stance about where you go in the case of a brain transplant (that is, whether the loss of the brain is a loss of an essential part of the person). In other words, which parts of the body are essential to your identity? Here are some possible answers:

1. you are your torso (the brain is not essential)
2. you are your brain (the torso is not essential)
3. you are your brain and your torso (both are essential)

Let's consider each one of these. (1) says that you are essentially your torso. On this view, you can survive a brain transplant but you go where your torso goes. If you were to receive a new brain (e.g., your friend's brain), you would be the person constituted by your torso and your friend's brain. In other words, receiving a new brain would be similar in kind to receiving a new heart, liver, or kidney. No doubt, the transplant would result in many significant changes. The most significant change would likely be that you would have different memories and beliefs. However, these changes would not be enough to say that you are a new person. You would still be the same person as you were before the brain transplant. *You go where your brain goes!*

(2) says that you are essentially your brain. You can survive a brain transplant, but you go where your brain goes. After the brain transplant, you would have a new torso (new heart, liver, kidneys, etc.). Receiving a new torso would be similar in kind to the normal case of receiving a new organ. The main difference would be that all of the organs would be received all at once and you would receive many parts that are not typically transplanted. No doubt, there would be many changes. Your eye color may be different, your heart rate may be higher than it was, and there will likely be many changes to your emotions given the impact of hormones on emotions. However, these changes would not be enough to say that you are a new person.

(3) says that you are essentially your brain *and* your torso. On this view, your torso and brain are essential and so a replacement of either would result in your death.

Exercise 5.28

In the case of a brain swap, where do you go?

To summarize, one problem for the biological theory is that it needs to provide greater specification about what parts of the body are essential for preserving a person's identity. In particular, the theory must state whether or not a person can survive a brain transplant. And, if they can, it must state where the person goes: do they continue with their torso or their brain?

5.4.3.3 From dicephalic parapagus

A second problem for the biological connection theory is that the theory cannot explain dicephalus. Dicephalic parapagus is an extremely rare case of twinning where there are two heads on a single torso. They occur 1 in every 50,000 to 100,000 deliveries. In most cases, the twins share body parts like the liver, pancreas, gallbladder, rectum (even the heart), but some cases involve two hearts, two pairs of lungs, etc. Normally, surgical separation of the twins is only performed if one of the twins is dying.[\[52\]](#)

The prognosis for survival for dicephalic twins is bleak. In many cases (40%), dicephalic twins are stillborn because of malformations in the vital organs. Twins that are born alive (30%) die within the first 24 hours of life.[\[52\]](#)

Two famous twins are Abigail Hensel and Brittany Hensel (born March 7th, 1990), two American conjoined twins. Each of the twins has a heart, stomach, spine, pair of lungs, and spinal cord, and each has control over one arm and one leg. Because of the separate control of the pair of legs and arms, crawling, walking, and clapping requires coordination while activities like writing and eating can be performed independently. For a short period of time Abby and Brittany were famous as they had their own reality TV series, and they appeared on talk shows and in popular print publications. At least at the time of writing (2024) both are elementary school teachers in Minnesota.

It is not clear what the biological theory says about cases involving dicephalus. The biological theory seems to say that since there is only one body, there is only one person. However, this leads to several problems.

First, the theory would contrast with our intuitions. Most people count there being two persons in one body rather than one person in one body. This is not only reflected in our intuitive judgments but also in how we talk about the twins, e.g., they have separate names.

Second, in cases where twins survive into adulthood, the view that there is only one person (in two bodies) would conflict with the testimony of the twins themselves. The twins regard themselves as two distinct people. Imagine that you underwent an operation where you were stitched together with another person. In such a scenario, you and the person with whom you were stitched would wake up thinking that you were now sharing a single body.

Third, ethical concerns surrounding the termination of one of the twins so the healthy twin survives seems to depend upon their being two people present in one body.

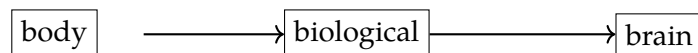
Fourth, successful separation of the twins would result in two surviving human beings. If there are two human bodies, then presumably there are two persons now. But this raises several problems. Was a new person created? Which one of the two people is the original person (if any)?

In short, just as in the case involving brain transplants, the biological theory of personal identity seems committed to saying that not every part of your body is essential to who you are. What is essential is your brain. When we are counting persons, the number of persons will equal the number of brains.

5.4.4 *The brain theory*

Given the arguments involving brain transplants and dicephalus, let's review the same body theory once again. Intuitively, we survive changes to our body that are the result of natural growth and decay. In addition, we can survive the periodic or potentially rapid replacement of body parts (heart, liver, kidney, etc.) that are not essential to our identity. We cannot, however, survive the replacement of our brains. Our brains are essential to preserving who we are: we go where our brain goes. Therefore, what preserves our identity over time is not biological continuity of the body (or the animal) but *biological continuity of the brain*. In addition, we also want to say that when we go about counting the number of persons in a given place, the number of human persons corresponds to the number of living human brains (not the number of bodies) in that place. So, in the case of dicephalic twins, there are *two persons* even if they share a torso.

With all of this in mind, let's formulate a new *physicalist* approach to personal identity. We might call this view the *The Brain Theory*.



According to the Brain Theory, you go where your brain go.

Definition 5.13: Brain theory

If A is a person at t_1 and C is a person at t_2 , then A is the same person as C iff A and B have the same *brain*. Two brains B_1 and B_2 are the same provided there is biological connection between the brains. And, there is biological connection between two brains provided: (1) there is a continuous physical connection between the brains, (2) the brains are of the same kind (e.g., both human brains), (3) the change to the brains is gradual, and (4) the brains are alive.

The brain theory is supported by arguments that support the body theory and the biological connectedness theory. In addition, there are two other considerations in its favor.

First, a common answer is that you go where your brain goes because your brain is responsible for your psychology and so its importance is solely due to the fact that it makes possible your personhood at an instant (your ability to be self-conscious or your ability to be rational). On this view then, the important parts of your brain (the essential part for preserving your personal identity) are the parts responsible for maintaining your psychology, not the parts that are central for other non-psychological aspects, e.g., respiration.

Second, the brain's importance (as opposed to other parts of the body) is central to maintaining the type of physical being that you are: a living human organism. As a living human being, your brain is important for maintaining your *dynamic stability*. That is, your body is undergoing a rapid exchange of material but it retains a stability (a stable form) due to a complex set of internal controls, e.g., the regulation of body temperature, metabolism, etc. The part of your body that is responsible for maintaining, coordinating, and regulating these vital functions is your brain.[90, p. 133][141] On this view then, the important parts of your brain are the parts responsible for maintaining your life (brain stem), not the parts that are central to more psychological aspects of your existence.[see 90, pp. 140–141]

The same brain theory seems to inherit all of the advantages of the earlier biological connection theory. Recall that one problem with the same body theory was that the theory was unclear about what it means for two bodies to be one and the same. If two bodies must be strictly identical in order to be one and the same, then no person can survive any change to their body. In order to overcome this problem, it was claimed that what is important for preservation of the person via their body is that there is the right kind of connection between one body and another. This feature is importantly

imported into the same brain theory since *a person's brain changes* in many ways:

1. At birth you have roughly 2,500 synapses (connections between neurons) per neuron. At age 3, you have 15,000 synapses per neuron while as an adult, you have roughly 7,500 synapses per neuron.
2. You are born with roughly 100 billion neurons and while the number of neurons stays roughly the same, the neurons grow (your brain gets heavier with dendrite growth)
3. Part of your brain consists of gray matter and there is a tiny loss of gray matter each year

In addition to inheriting this feature of the biological connection theory, the same brain theory also has two key advantages: (1) in the case of a brain transplant, you would go where your brain goes and (2) it implies that there are two persons (rather than one) in the case of dicephalic parapagus.

Given that we have winnowed down the body theory (where survival required preservation of the body) to the brain theory (where survival required preservation of the brain), one might wonder whether the brain theory might be winnowed down even further. That is, perhaps we do not need to preserve the entire brain in order to preserve our identity. Perhaps we only need to preserve some part of the brain. For example, perhaps we only need to preserve the hippocampus or the prefrontal cortex. While this is an interesting question, it is not one that we will consider in this chapter.

5.4.5 *Objections to the brain theory*

Let's consider some possible objections to this theory.

5.4.5.1 *Brain fission*

Our path has been from the biological-connectedness theory to the brain theory. One key reason for this transition was that we did not take the whole of the body to be necessary for preserving our identity. Instead, it was argued that only a *part* of the body (the brain) is necessary for personal identity. With that in mind, we might ask whether we need to preserve the *entire* brain in order to preserve our identity. Perhaps we only need to preserve some part of the brain.

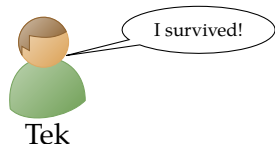
Suppose that Tek has a routine health checkup. He has some blood drawn and a few days later receives a call from Dr. Jon informing him that he has been infected with a rare virus. Dr. Jon informs Tek that Tek's entire body has been infected except the left hemisphere of his brain. Dr. Jon tells Tek that

unless Tek has his left hemisphere removed and transplanted into a healthy body, the virus will spread to the left hemisphere and Tek will die. Dr. Jon assures Tek that medical science and technology has made such a transplant possible. He says that the transplant procedure improves upon (1) corpus callosotomies (severing the corpus callosum and thereby disconnecting the two hemispheres) and (2) hemispherectomies (the entire removal of one of the cerebral hemispheres). Dr. Jon notes that these procedures are performed on patients with severe epilepsy when antiseizure medication has proven to be ineffective. And, while there is some functional limitations, these individuals go on to lead healthy, normal lives.

Tek reasons as follows:

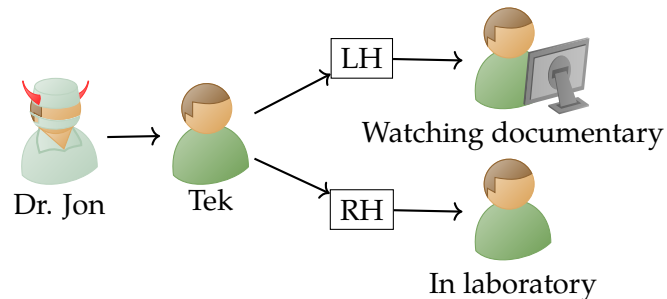
- P1. People can survive a corpus callostomy.
- P2. People can also survive a hemispherectomy.
- IC. Therefore, my *whole* brain is not essential for my survival.
- P3. People can also survive organ transplants.
- P4. I am my brain.
- C. Therefore, I can survive my left hemisphere being transplanted into a new body.

Reasoning that if someone can survive a hemispherectomy and someone can survive a total brain transplant, then Tek believes he ought to be able to survive the transplantation of one of his hemispheres into a new body (assuming there are no complications in the surgery). With that in mind, Tek agrees to the operation. After the operation, Tek is able to live a normal life in his new body with some functional limitations. One such limitation is that since Tek has only retained his left hemisphere, he is unable to name objects in his right visual field (note that visual information from the right visual field is processed by the left hemisphere, while the left hemisphere predominantly handles speech). But besides this limitation (and a few others), Tek is able to walk, talk, and think. He is able to remember his past and plan for his future. He is able to engage in the same activities that he did before the operation.



Years later, Tek is watching a documentary on the brain. The documentary covers how his doctor (Dr. Jon) performed hemispherectomies on patients and the patients were able to live normal lives. However, Dr. Jon was later arrested since (1) he performed the operation on healthy patients and (2) he was transplanting each of the two hemispheres into separate, healthy human

bodies. Dr. Jon would then release one of the healthy patients to the public and keep the other patient in his lab.



Tek was shocked by this revelation and, after some reflection, he began to wonder whether he survived the operation. He reasoned as follows:

- P1: If the brain theory is true, then a person can survive the transplantation of one of the hemispheres of their brain.
- P2: My left hemisphere was transplanted into a new body, so I survived.
- P3: My right hemisphere was transplanted into a new body, so they survived.
- IC: Therefore, the person watching the documentary is the same person that was trapped in Dr. Jon's lab (P2, P3).
- P4: It cannot be the case that I am the same person as the person that was trapped in Dr. Jon's lab (we are in different places, have different thoughts, have different bodies, etc.).
- C: Therefore, the brain theory is false.

The above argument poses a problem for the brain theory. On the brain theory, if a person A_1 goes where their brain goes, then if the brain B can be divided into two parts B_1 and B_2 , and the person can survive as both of these two parts, then person A_1 survives in two different bodies, in two different locations, with two different sets of thoughts. This seems to be an implausible consequence.

Cases like the one involving Tek and Dr. Jon are sometimes referred to as "fission" cases. The fission case is a problem for the brain theory since it seems to imply that a person can survive in two different bodies or that a person cannot survive the transplantation of one of their hemispheres.

Exercise 5.29

We will look at another case like this later. For now, how would you go about solving it?

One possible revision is to say that A_2 is the same person as A_1 provided (1) A_2 has the same brain as A_1 and (2) there is no other person that has the same brain as A_1 . This would imply that (1) a person who has a hemisphere transplanted into a healthy body would survive if the other hemisphere was destroyed but (2) not survive the transplant if the other hemisphere were transplanted into another healthy body. Here is how Swinburne puts the point:

“Suppose P_1 's left hemisphere is transplanted into some skull and the transplant takes. Then, according to the theory, whether the resulting person is P_1 , i.e., whether P_1 survives, will depend on whether the other transplant takes. If it does, since both resulting person will satisfy the [...] brain continuity criteria equally well, neither will be P_1 . But if the other transplant does not take, then since there is only one person who satisfies the criterion, that person is P_1 . So, whether I survive an operation will depend on what happens in a body entirely different from the body which will be mine, if I do survive.” [116, pp. 15–16]

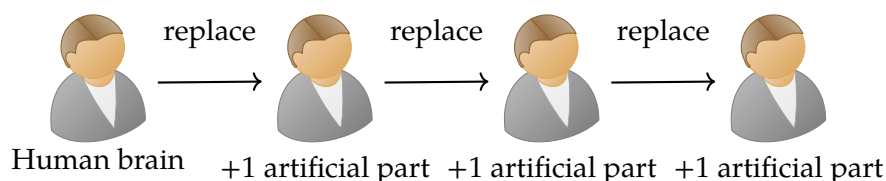
But this result seems implausible since presumably whether person P_1 continues to exist should depend solely upon the intrinsic properties of P_1 not whether some other entity P_2 exists.

5.4.5.2 *Artificial brain replacement*

One of the assumptions that we have not critically examined is that B_1 is the same body (brain) as B_2 only if B_1 and B_2 are the same kind of thing. In the case of human persons, this implies that if A is a human and B is a dog, then A and B do not have the same body. With respect to brains, if A was a human brain and B was a dog's brain, A and B would not be one and the same brain. More to the point, we don't preserve our personhood if we switch to a different kind of organism.

Let's call the following the *Artificial Brain Replacement Objection*. Suppose some future technology allows for us to reconfigure our brains by replacing brain material with artificial material that accomplishes the same function as a human brain. For example, the hippocampus is a part of the brain that plays an important role in memory. Suppose it is discovered that Tek has Alzheimer's disease and it is also discovered that his Alzheimer's could be cured by replacing his hippocampus with artificial yet biocompatible materials. Presumably, we would say that provided Tek's body survives the surgery and he behaves in a similar (or better) way than he did before the surgery, then Tek is the same person. However, suppose ten years later, Tek's has

several problems with his vision and the problem is located in his occipital lobe. Similar to before, Tek has an occipital lobe replacement. And, similar to before, Tek is cured. Again, provided Tek behaves in much the same way (or better) we would regard Tek as the same person.



Now consider that Tek's life is extended another hundred years and over that hundred years, each part of his brain is replaced with an artificial part. Biological theorists (and the brain theory) would reject the idea that you could survive with an artificial brain. As Olson[90, p. 125] puts it,

“It may be possible to replace all of your parts, including your brain, gradually and piece by piece, with inorganic prostheses in such a way that your mental capacities were preserved throughout. [...] The result would be a wholly non-biological person---with rationality, consciousness, free will, the works---who was both psychologically and materially continuous with you. Nevertheless, according to the Biological Approach that being would not be you, for you are a biological organism, and no organism could come to be a non-organism.”

Presumably, we would say that (1) Tek's brain is no longer a human brain but (2) Tek is the same person. More generally, it seems to suggest that the type of material that makes us up is not essential to who we are.

To illustrate this point in a different way, we might consider why we think that the human brain is important for preservation of our continued identity over time? Presumably, we think this because it is responsible for our consciousness, our memories, our psychology. This suggests that what really matters for our continued survival is not the brain itself or being a particular type of animal but *what our brains do for us*. And, if there were another way of fulfilling the function of our brains (e.g., holding our memories) without the use of the human brain, we would no longer take the preservation of our human brains to be the only means of preserving our identity. Locke[72, Bk.II, xxvii, 17] considers an extravagant illustration of this intuition when he says that we tend to see our bodies as a part of ourselves but if we lose certain parts and these parts are no longer connected with our consciousness, then we regard it as no longer being part of who we are. For example, if you have

a little finger, you regard it as a part of you, but if you were to lose that finger, you would regard it as a prior part of your body but not part of who you are now. But suppose now

“Upon separation of this little Finger, should this consciousness go along with the little finger, and leave the rest of the body, 'tis evident the little Finger would be the *person*, the *same person*; and *self* then would have nothing to do with the rest of the body.”

Locke's point here is that we would regard the person as going with the consciousness no matter the substance in which it resides: be it the brain, a finger, a soul, or something else. We are not essentially bodies even though we have these bodies.

So, while Tek may not be able to survive having no brain at all, it isn't clear that Tek must survive as a human brain in order to retain his identity.

Exercise 5.30

Suppose you find yourself in a situation where part of your brain needs to be replaced with an artificial component. Would you undergo the operation (replacing the biological material with artificial material)?

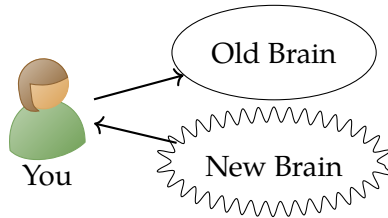
5.4.5.3 Spatiotemporal continuity objections

One of the key claims of both the biological continuity and brain theories is that to preserve one's identity their needs to be biological continuity between the two persons, e.g., continuity of the brain. At least here, biological continuity is understood in terms of mere *spatiotemporal continuity* between the body. That is, there cannot be gaps in space and time between the consecutive moments of the object's existence. Let's challenge this claim by considering whether a person can retain their identity without biological continuity. We will consider two cases.

5.4.5.4 Lab-grown Brain Objection

Let's call this objection the *Lab-grown Brain Objection*. Suppose your brain is beginning to die. Worried about the ultimate decay of your brain, a medical professional asks if you are interested in having your psychology transferred. First, your brain is mapped using the latest 3D-imaging software. Using anesthesia, you are put to sleep (temporarily) while new brain material is grown. Once the brain is grown, your brain and the lab-grown brain are compared. Once it is verified that the brain configurations are identical,

your brain is removed from your body and the lab-grown brain is inserted into your body. Anesthesia is then stopped. When your body with its new brain wakes up, your old brain is destroyed.



If the brain theory is true, then you do not survive this operation. However, some might argue that you do survive the operation since the resulting person has the same memories, goals, and personality that you did before you underwent the operation. It is just like an organ transplant and the surviving person is like you in every way (they talk like you, act like you, etc.). What this objection would further reveal is that biological continuity is not necessary for preserving personal identity.

Exercise 5.31

Consider the Lab-grown Brain Objection. Do you think you would survive the operation? Explain your reasoning.

5.4.5.5 *Body-Swapping Objections*

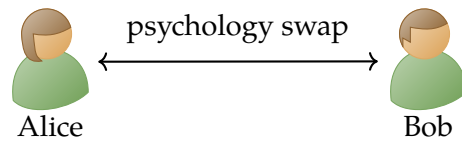
Let's call this objection *Williams' Body-Swapping Objection* after an example given by Bernard Williams[152]. This example proposes a hypothetical example aimed at showing that bodily continuity is not a necessary condition for preserving identity. That is, it propose an example where personal identity is preserved but spatiotemporal continuity is not.

Suppose two people *A* and *B* are in a room with a mad scientist.



The mad scientist has developed a machine that can take two people and swap their bodies. The machine does this by extracting information from the brains of *A* and *B*, storing it in a device, and then reimplanting the information into the other person's brain.[152, p. 162] Let's suppose the machine does this by taking the information from person *A*'s brain and then incre-

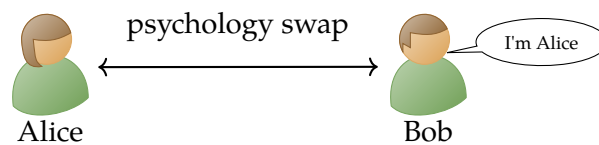
mentally changing person *B*'s brain so that it is virtually identical to person *A*'s brain. The same is done for person *B*'s brain.



Let's note two things about the result of using the machine. First, note that body-*A* at t_1 (before the running of the machine) is biologically continuous with body-*A* at t_2 (after the machine runs). Similarly, body-*B* at t_1 is biologically continuous with body-*B* at t_2 . If the brain theory is correct, then, after the machine runs, person *A* is in body-*A* and person *B* is in body-*B*.

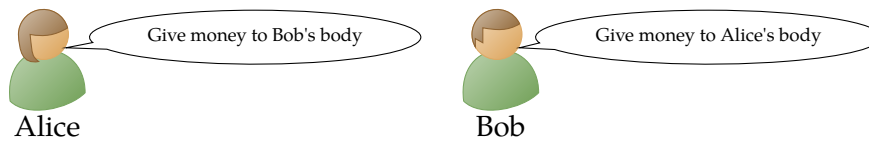
Second, what we want to see is whether spatiotemporal continuity of the body is a necessary condition for preservation and so whether it is possible to preserve identity without spatiotemporal continuity of the body.

The third thing to note is that the utterances, character, and actions coming from Body-*B* are expressive of the memories, character, and behaviors associated with *A*. In other words, the personality of Body-*B* appears to be the personality of *A*. For example, Body-*B* appears to recall experiences that *A* had. The same is true for body-*A*. The personality of body-*A* is the personality of the person *B*. As Williams notes, after the machine runs "persons familiar with *A* and *B* are just *overwhelmingly struck* by the *B*-ish character of the doings associated with what was previously *A*'s body, and conversely." [152, pp. 161–162]



The question then is even though body-*A* has personality *B* and Body-*B* has personality *A*, it is an open question whether body-*A* (or body-*B*) is person *A*, person *B*, or some new person *C*. To test this, Williams considers several thought experiments involving *A* and *B*. Let's consider two of these thought experiments.

First, suppose, *before the psychology swap*, the mad scientist tells *A* and *B* that after the machine runs, he plans to give one of the bodies (Body-*A* or Body-*B*) a large sum of money and the other body a painful electric shock. The scientist then asks both *A* and *B* for their preference about which body should receive the money and which should receive the shock. Suppose *A* says that Body-*B* should receive the money while *B* says that body-*A* should receive the money.



Now the scientist cannot act on the preferences of both *A* and *B* (since they conflict) so the scientist proceeds to award it to body-*A*. What would be the response from both of the body-persons? Body-*A* would be pleased with the money. They would say that this is outcome they desired. Recall that it was *B* who wanted the money to go to body-*A*. But, Body-*B* would act upset. They would say that they wanted the money (not the electric shock) to go to Body-*B*. But recall that it was person *A* who wanted the money to go to Body-*B*. From this, Williams concludes that

"These facts make a strong case for saying that the experimenter has brought it about that *B* did in the outcome get what he wanted and *A* did not. It is therefore, a strong case for saying that the *B*-body-person really is *A*, and the *A*-body-person really is *B*" [152, p. 164]

Second, suppose *A* has anxiety and *B* has a bad memory of a car accident. Each wishes to escape their respective condition by swapping bodies. Their thought is that if you go where your body goes, then the machine will replace their personality with the personality of another person. After the machine runs, the mad scientist asks body-*A* if their anxiety is gone. Body-*A* says that they never had anxiety and express displeasure with the scientist since the machine did not rid them of their troubling memory of the car crash. The scientist then asks Body-*B* if they remember the car accident. Body-*B* says that they have never been in a car accident and express annoyance with the scientist since their anxiety has not been removed. Williams concludes that this experiment suggests that "the only rational thing to do, confronted with such an experiment, would be to identify oneself with one's memories, and so forth, and not with one's body. The philosophical arguments designed to show that bodily continuity was at least a necessary condition of personal identity would seem to be just mistaken." [152, p. 167]

In sum, we can potentially draw two conclusions. First, spatiotemporal continuity is not sufficient for preserving our identity since *A*'s brain survives but *A* is not in *A*'s brain. Second, spatiotemporal continuity is not necessary since you can survive in another body.

Exercise 5.32

Consider Williams' Body-Swapping Objection. Suppose you and a friend swap bodies as Williams has described. Would you survive the operation? Which body would you go with? Explain your reasoning.

5.4.5.6 End of life and degeneration issues

Let's consider an example where there is biological continuity between two brains but where we intuitively think that the person no longer exists. In such a case, continuity of the brain is not sufficient for preserving personal identity.

One attractive feature of physical theories of diachronic identity is their ability to account for the fact that we were once a fetus. That is, they can connect the person that we are now with the preconscious "person" that we once were. But what do they say about when we die. Consider the following three scenarios:

1. A human brain in a persistent vegetative state (no capacity for consciousness, no brain activity)
2. A human brain in a deep coma (the parts of the brain that are responsible for consciousness are intact but they are not active or functioning)
3. A human brain in a state of advanced Alzheimer's disease (significant dementia, loss of memory, inability to use language, loss of basic problem-solving abilities)

One reason we take the brain to be so essential for preserving our personal identity is its ability to maintain our psychology (ground our consciousness, keep our perspective on things, maintain a stockpile of memories). However, while the brain may continue to exist, when it no longer supports these critical functions, we may intuitively think we no longer exist. If that is the case, then (1) brain B_1 at t_1 is biologically continuous with brain B_2 at t_2 but (2) while we exist at t_1 , we no longer exist at t_2 (contradicting the brain theory).

Exercise 5.33

Would you still exist (1) in a persistent vegetative state, (2) in a deep coma (that you will never wake up from), or (3) in an advanced state of Alzheimer's disease?

5.5 PSYCHOLOGICAL APPROACHES

In previous sections, we considered two major approaches to answering the question of diachronic personal identity: soul approaches and physical approaches.

There are a set of approaches that we will call “psychological approaches” to diachronic personal identity.⁶ The core characteristic of a psychological approach is that there is some feature concerning a person's psychology that determines whether a person at t_1 is the same person at t_2 . In this section, we consider several instances of this general approach.

5.5.1 *Same psychology theory*

Let's try to solve the problem of diachronic personal identity by relying upon the idea that a person A at t_1 is the same person as B at t_2 iff B at t_2 has the same psychology.

Definition 5.14: Same psychology theory of personal identity

If A is a person at time t_1 and B is a person at t_2 , then B is the same person as A iff B at t_2 has the same psychology as A at time t_1 .

There are two problems with this theory. First, it is unclear what it means for a person to have the “same psychology” as another person. The theory then is too vague. Second, our psychologies change over time. For example, Tek may enjoy cartoons as a child but now, for enjoyment, only reads books of poetry. Similar to physical approaches to personal identity, it seems then that a psychological approach must be able to accommodate cases where A and B are the same person even though their psychologies are not numerically identical.

5.5.2 *Memory based approaches*

As mentioned, one problem with the same psychology theory of personal identity is that the conditions under which two persons have the same psychology is vague. Let's fix this problem by specifying what part of the psychology is important for survival. Let's suppose that A and B are the same person iff B can remember the experiences of A , or vice versa. In doing this, we regard the person as essentially being their mind (or consciousness). This consciousness has experiences and thoughts at various times. What preserves the same person over time is (1) continued consciousness but also (2) the ability to remember prior experiences and thoughts.

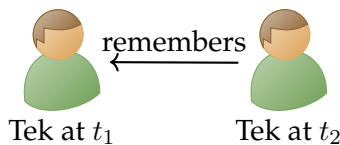
Definition 5.15: Simple memory theory of personal identity

If A is a person at time t_1 and B is a person at t_2 , then B is the same person as A iff B at t_2 has the memories of the experiences of A at time t_1 .

Here is how Locke[72, Bk.II, xxvii, 16] expresses this view:

“plain consciousness, as far as ever it can be extended, should it be to ages past, unites existences, and actions, very remote in time, into the same person, as well as it does the existence and actions of the immediately preceding moment: So that whatever has consciousness of present and past actions, is the same person to whom they both belong.”

Thus, Tek at t_2 is the same person as Tek at t_1 if and only if Tek at t_2 remembers an experience had by Tek at t_1 .



Let's make a few clarifications to the theory. First, we don't expect Tek at t_2 to have a perfect memory of every detail of what happened at t_1 . We just expect Tek to remember something. So, if Tek at t_2 remembers going to a carnival yesterday, he only needs to remember the experience of going to the carnival, not every person, ride, and food vendor present. In short, Tek at t_2 is the same person as Tek at t_1 provided Tek has at least a *fuzzy memory* of an experience or action he performed at t_1 .

Second, an *occurrent memory* is a memory a person is actively having. For example, suppose Tek is sitting in a chair and currently thinking about what he ate yesterday. In this example, Tek's memory is *occurrent*. In contrast, a *non-occurrent memory* (also called “dispositional memory”) is a memory that Tek is capable of having. This type of memory is also called “standing” or “dispositional”. For example, suppose Tek is thinking about eating a sandwich. If we asked Tek to remember a painting that he painted years ago, he *could* recall that painting. His memory of the experience of painting that painting is *non-occurrent* as it is not occurring in Tek's mind but it could occur in his mind if we prompted him.

The memory theory says that the memory that makes Tek at t_2 the same person as Tek at t_1 need only be *non-occurrent*. That is, the persistence of Tek as a person over time doesn't require that Tek is constantly thinking of

every prior experience Tek had. Instead, we only expect Tek at t_2 to be *capable* of remembering experiences had by Tek at t_1 .

Third, there is a distinction drawn between (1) factual memory, (2) procedural memory, and (3) personal (episodic) memory.

Factual memory refers to remembering facts about what one has learned about the world. Suppose I learn that the capital of Pennsylvania is Harrisburg. Later on, I recall this fact. My memory is about some piece of information about the world. Or, suppose Tek, who was born in 2000, learns that The Great Depression began on October 24, 1929. A day later, Tek takes an exam and is asked about when The Great Depression began. Tek's factual memory refers to his ability to recall a fact about some event in the world.

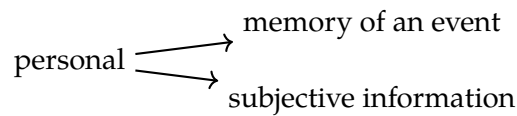
Factual memory is not the type of memory that makes me the same person as a person in the past. The fact that I remember about what someone did during The Great Depression does not make me the same person as that person. I don't remember doing what that person did, nor do I believe myself to be remembering experiences of a person that is identical with me. My memory of them is not drawn from a belief that I had those experiences or committed those actions.

This point is important since Williams[151, p. 233] points out that if the memory thesis is "if X can remember Y's doing such-and-such, then X and Y are the same person", then the thesis is mistaken. He notes that we constantly say things like "I remember my brother joining the army" but there is no implication that you and your brother are the same person.

Procedural memory refers to being able to remember *how* to perform some task (procedural knowledge simply being knowing how to do some act). For example, I may have learned to ride a bicycle, write in cursive, or create a cereal box using some 3D software. Suppose many years pass and I don't remember how I ever learned these things. I don't recall any instructions, any trial or error, or if I even know how to perform these tasks. Nevertheless, when I hop on the bicycle, I am able to ride it. My riding the bicycle indicates I remember *how* to perform this task.

Finally, *personal (episodic) memory* is a type of declarative knowledge and it refers to memories a person has about events (when, where, and what occurred) that the person believes themselves to experienced along with subjective information derived from having experienced that event, e.g., thoughts or emotions that were had during the experience of that event. For example, suppose Tek sees a tall person with brown hair outside of his residence. He feels fear and considers whether or not he should call the police. Later that

day, he recalls seeing a tall man outside of his residence and that he felt fear. In this example, Tek's recollection is a form of personal (episodic) memory.



Here is another way of describing personal memory. When a person has an episodic memory of an event, they are thought to be able to travel back in time and *reexperience* the event in their mind. This is in contrast to factual memory (of say The Great Depression) where we are only *imagining* what it would be like to experience a given event.

The memory theory says that the memory that makes Tek at t_2 the same person as Tek at t_1 is personal memory. That is, the persistence of Tek as a person over time does not involve the recollection of facts about the past but a person's experiences had in the past. To put this in more Lockean terms, our continued persistence requires not only consciousness of the present moment but a consciousness of our consciousness in the past.

Finally, a distinction is drawn between genuine memory (strong memory) and apparent memory (weak memory). If Tek has a *genuine memory* of doing an action A , then (1) A had to happen and (2) it had to be Tek that did A . In other words, you can only have a genuine memory of action that you, in fact, did. For example, Tek can only remember eating a sandwich yesterday only if he ate a sandwich yesterday. An *apparent memory* refers to the belief that one had a certain experience. In contrast to a genuine memory, having an apparent memory does not entail that what one "remembers" actually occurred. For example, Tek may have an apparent memory that he ate a sandwich yesterday (he may feel strongly that it was even a ham sandwich) but he had soup instead. Here is how Swinburne puts this distinction:

"In the normal or strong sense of 'remember', one can only remember doing something if one really did it. I may say that I 'remember' going up the Eiffel Tower, but if I didn't do it, it seems natural to say that I cannot really remember having done it. In this sense, just as you can only know what is true, so you can only remember what you really did. However, there is also a weak sense of 'remember' in which a man remembers whatever he believes that he remembers in the strong sense. One's weak memories are not necessarily true ones." [116, p. 9]

Persistence requires apparent rather than genuine memory. To see why, con-

sider the following explanation:

To “be good” is to *act good*. What then does it mean to “act good”?
Ah! To “act good” is just to *be good*.

The above explanation is *circular*. X is explained by Y and Y is explained by X . Not only are some terms defining circularly but various claims about the world have an argument structure that is circular. To illustrate, consider the following argument:

America is the best country in the world. What makes it the best is it being the number 1 country. What makes it the number 1 country is its being the best in the world.

Notice that this argument is valid but circular. This circularity is problematic because, in uttering the controversial claim (America is the best country) there ought to be evidence in support of the claim that does not presuppose the truth of the claim. In the above argument, the evidence (America is the number 1 country) already assumes the truth of the conclusion.

Now let's consider a claim made by the simple memory theory:

What makes X at t_1 and Y at t_2 the same person is Y having a genuine memory of some action performed by X .

So far so good. But when does Y have a genuine memory of an action performed by X ? We said that a genuine memory requires two things. First, the memory must be of an event that actually happened. If Tek remembers a bank being robbed but the bank was never robbed, then Tek does not genuinely remember this event. Second, the memory must be a personal memory, so it must be of an event that Y actually performed or experience. So, if Tek remembers robbing a bank but it was Tek's friend (rather than Tek) who robbed the bank, then Tek does not genuinely remember this event. Tek's memory must be of an event he experienced. To summarize:

Definition 5.16: genuine memory

Y has a genuine memory of an action performed by X if and only if (1) the action was performed by X and (2) Y performed the action.

But notice that the definition of a genuine memory presupposes that Y and X are identical. That is, for Y to have a genuine memory of an experience had by X , Y must be X . To summarize:

1. What makes X and Y one and the same person is Y 's ability to have a genuine memory of some action performed by X .

2. Y 's ability to have a genuine memory of an action performed by X presupposes that X is one and the same person as Y .

Let's consolidate this even further to make the circularity apparent:

X and Y are the same person if Y genuinely remembers what X did, but Y only remembers what X did if X and Y are the same person.

With all of these clarifications, let's restate the memory theory to reflect that the type of memory a person Y must have to be the same person as X is a fuzzy, personal, non-occurrent, apparent memory of X .

Definition 5.17: Simple memory theory of personal identity

If A is a person at time t_1 and B is a person at t_2 , then A and B are the same person iff B at t_2 can have a fuzzy, personal, non-occurrent, apparent memory of some event or experience had by A .

On this account, a person is not identical to the person who I believe myself to be.⁷ I might believe that I am the biblical Noah or Genghis Khan but this belief is based upon, not on any memory of their experiences, but on an interest in these characters and a host of other irrational beliefs. Rather, on the memory theory, a person is identical to all of those people they believe themselves to be on the basis of beliefs about the actions and experiences that they believe themselves (rightly or wrongly) to have performed or had.

At least at first glance, the memory theory appears to give us the right results. Take Tek and Liz. Tek has a great many beliefs about experiences he has had and actions he has performed. He may believe he remembers going to school at such-and-such school, married Sal on such-and-such date, and being honored with an award in his workplace. Insofar as Tek has these apparent memories, these prior persons are (according to the theory) one and the same person as Tek today. Since Liz does not have any of these beliefs, these prior persons are not Liz.

Liz, in contrast, has her own beliefs about what she remembers. In contrast to Tek, she apparently remembers going to a different school, turning down a marriage proposal from Sal, and being snubbed for an award. Since she apparently remembers these events, these prior people are one and the same person as Liz today. Since Tek does not have any of these beliefs, these prior persons are not Tek.

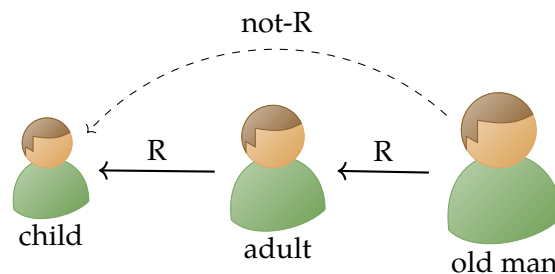
5.5.3 Objections

Even with the memory theory appropriately qualified in terms of the type of memory a person must have in order to be one and the same person as someone in the past, there are several remaining problems. In this section, we consider one such objection, postponing others that could apply to a more refined version of the memory theory. The example we will consider is from Thomas Reid. [107, p. III.6] Let's call this objection *The Objection from the Transitivity of Identity*.

Reid asks his reader to suppose a military man at three different stages of life:

1. As a child, he was flogged at school for robbing an orchard
2. As an adult, he captured the enemy's flag (standard), thereby indicating their total defeat
3. As an older man, he was promoted to the position of General.

As an adult, he could recall having been flogged as a child. As an older man, he could recall having captured the enemy's flag but could not recall having been flogged as a child.



According to Reid, it follows that

“he who was flogged at school is the same person who took the standard (the flag), and that he who took the standard is the same person who was made a general. Whence it follows, if there by any truth in logic, that the general is the same person with him who was flogged at school. But the general's consciousness does not reach so far back as his flogging, therefore [...] he is not the person who was flogged. Therefore, the general is, and at the same time is not the same person with him who was flogged at school.” [107, III.6, p.488]

Let's summarize

- Assume the memory theory is true.

- P1. The general g remembers being an adult a so $g = a$.
- P2. The adult a remembers being a child c so $a = c$.
- P3. The general does not remember being a child so $g \neq c$.
- P4. But, Identity is transitive so if $g = a$ and $a = c$, it follows that $g = c$ (the general is one and the same person as the child).
- IC. $g \neq c$ and $g = c$ (from P3 and P4). Contradiction!
- C. Therefore, the memory theory is not true.

Exercise 5.34

Create your own version of Reid's famous objection.

5.5.4 Modified memory theory

It appears that the primary problem with the simple memory theory is that it requires *direct psychological connections* between persons. That is, in order for A to be B , person A must directly remember something about what B has experienced. One way to fix this issue is to say that what is required are not direct psychological connections but indirect psychological connections (this is sometimes referred to as *psychological continuity*). Let's define a new theory that makes use of this idea and then articulate the idea of psychological continuity more fully.

Definition 5.18: Modified memory theory of personal identity

If A is a person at time t_1 and B is a person at t_3 , then B is the same person as A iff either (i) B at t_3 apparently remembers an action or experience of A at time t_1 or (ii) there is some person C at t_2 such that (a) B remembers apparently remembers the actions / experiences of C and (b) C apparently remembers the actions / experiences of A .

In short, the modified memory theory of personal identity says that two persons A, B are the same person if there is a *continuity* of memory between A and B . In other words, while you might not be able to directly remember your experiences five days ago, three months ago, or five years ago, you are the same person if you were able to remember the experiences of a prior-you who is able to remember said experiences.

5.5.4.1 Objections

There are several objections to the modified memory-theory approach.

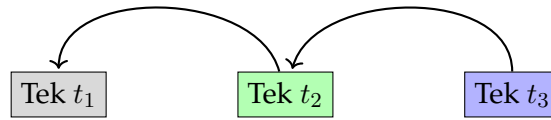


Figure 5.1: In the modified memory theory, Tek at t_3 is the same person as Tek at t_1 even if Tek at t_3 cannot remember having an experience as Tek at t_1 . Rather, it is sufficient that Tek at t_3 remember having an experience as Tek at t_2 and that Tek at t_2 remembers having an experience as Tek at t_1 .

5.5.4.2 Gaps

The modified memory theory suffers from the fact that if there is a gap (missing link) in your memory, then you cannot claim identity with a person with whom you don't have any memory of their experiences. There are several examples of gaps, including (but not limited to) being asleep, intoxicated, spacing-out, being unconscious, being in a deep meditative state, suffering from a brain injury or amnesia.

First, suppose Tek suffers a brain injury, recovers, but there is a significant gap in his memory. It is 2017 and while he remembers what he did prior to 2008, he does not remember anything between 2008-2016. According to the modified memory theory, Tek is the same person as the person living before 2008 but he is not the same person as the person who was lying in the hospital bed and subsequently recovering between 2008-2016. brain injury

Second, suppose you are out drinking (heavily). You do some embarrassing things that you don't remember. This creates a gap in your memory. You can remember what you did yesterday and so you are the same person as that the person yesterday but you don't remember what you did last night. According to the modified memory theory, you are not the same person who was out drinking. Third, we can think of other temporary gaps in memory, e.g. when we are sleeping, being knocked unconscious, daydreaming. heavy drinking

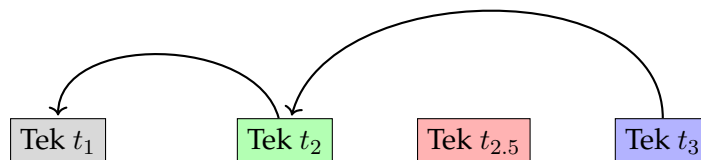


Figure 5.2: In the modified memory theory, if Tek at t_3 cannot remember Tek at $t_{2.5}$ and there is not an individual before Tek t_3 but after Tek at $t_{2.5}$ that remembers having an experience as Tek at $t_{2.5}$, then Tek at $t_{2.5}$ is not the same person as Tek at $t_{2.5}$.

Exercise 5.35

Articulate a time in your life that you have absolutely no memory of when it occurred.

5.5.4.3 Unity objection †

Consider that what makes a person the same person over time is the capacity of the person at a given time to apparently remember (at least in a fuzzy way) what they did at an earlier time. What ties a person together over time then are not individual *acts* performed by the person but a *power* performed by a person. One possible objection (raised by Henry Lee):

“Consciousness can be only a Power, or the repeated acts of a Power; but neither Power or repeated acts of a Power can be the Thing or Substance itself that has that Power or exerts those Operations; and consequently can't be that from which a Person is denominated the same with himself at different times, unless we could suppose the Power to be the Substance of which it is a Power, or an Action to be the *Agent* itself; which is impossible.”[69, p. 172]

For Lee, consciousness is a power or the repeated activity of a power. If it is a power, then it is a power of a person. But, if it is a power of a person, then the person is logically prior (exists before) the power and so it is not the power that makes the person. It is thus not the ability to remember or repeated acts of remembering that allow for persistence. Instead, it is the existence of the thing (the soul or mind) that has that power to remember that allows for persistence.

5.5.4.4 Fission problems

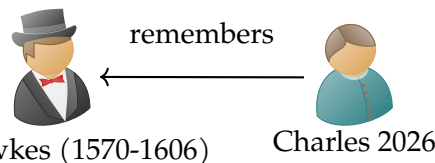
Consider that if person *B* claims to remember what *A* did, then *B* and *A* are the same person. In many cases, this appears to get the right results. Suppose that today I remember that yesterday I wrote a few sentences on metaphysics. My consciousness today is able to remember what a consciousness yesterday experienced. However, suppose I remember seeing Noah build his ark before the great flood in 2350 BC (let's assume this occurred). On this account, I would be one and the same person that saw the building of the ark.⁸ But that cannot be correct.

Let's consider a variation on this objection from Bernard Williams.[151] Consider a person named Charles wakes up one morning and claims to be Guy

Fawkes (1570-1606).

“all the events he claims to have witnessed and all the actions he claims to have done point unanimously to the life-history of some one person in the past--for instance, Guy Fawkes. Not only do all Charles' memory-claims that can be checked fit the pattern of Fawkes' life as known to historians, but others that cannot be checked are plausible, provide explanations of unexplained facts, and so on. Are we to say that Charles is now Guy Fawkes, that Guy Fawkes has come to life again in Charles' body, or some such thing?”[151, p. 257]

To flesh out this case some more, let's suppose that Charles is not merely claiming to be Guy Fawkes. Suppose he is also able to support his claim that he is Guy Fawkes with substantial evidence. For example, Charles' provides evidence that (1) is seemingly not obtainable via the historical record and (2) provides “explanations of unexplained facts”. Let's further add that Charles' current memory claims also (3) result in the discovery of new historical artifacts whose whereabouts would be known by Guy Fawkes. For example, suppose it is known that Guy Fawkes, before his execution, was in possession of a secret diary that outlined a plot to assassinate King James I. Charles, in an effort to support the claim that he is Guy Fawkes, provides researchers the location of where the diary is buried.



Given this evidence the memory theory, Charles is Guy Fawkes.

- P1: Charles claims to be Guy Fawkes.
- P2: The events Charles claims to have witnessed and the actions he claims to have performed point only to the life of Guy Fawkes.
- P3: The events and actions that Charles claims to have performed and that can be verified are checked by historians and these events and actions correspond to the life of Guy Fawkes.
- P4: Charles' claims also provide explanations for several previously unexplained historical facts (historical mysteries).
- P5: Charles' claims also result in several new historical discoveries.
- P6: The best explanation for what has happened to Charles is that he has become Guy Fawkes (not that he has become clairvoyant).
- C: Therefore, Charles is Guy Fawkes.

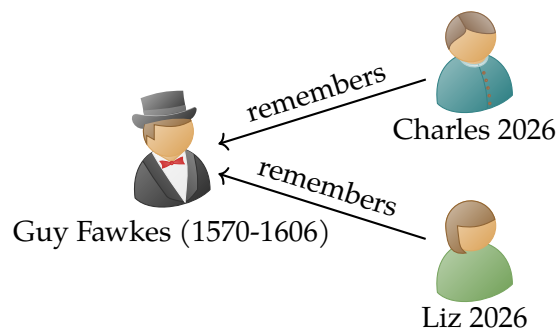
Exercise 5.36

Create your own example where someone *B* today claims to be a previous historical figure *A* and provides so much evidence through various memory claims that it is reasonable to believe that *B* is the same person as *A*.

Here is the objection from Williams.

“If it is logically possible that Charles should undergo the change described, then it is logically possible that some other man should simultaneously undergo the same changes, e.g., that both Charles and his brother should be found in this condition. What should we say in that case? They cannot both be Guy Fawkes; if they were, Guy Fawkes would be in two places at once, which is absurd. Moreover, if they were both identical with Guy Fawkes, they would be identical with each other, which is also absurd. Hence we could not say they were both identical with Guy Fawkes.” [151, p. 257]

In short, we have another fission case. It is possible that Charles and Liz (who occupy different bodies at the same time) can both remember the experiences of Guy Fawkes. If memory appears to be sufficient for preservation of personhood, then both of these individuals would be Guy Fawkes.



If personhood is transitive, then Charles = Liz. But intuitively, Charles $\neq$ Liz. Therefore, if some psychological approach to personal identity is correct, then it is not based on memory.

Exercise 5.37

Create your own variation of the Guy Fawkes objection.

5.5.5 The personality theory

Memory might be *sufficient* for survival but *not necessary*. For instance, we can imagine an individual Tek who loses all of his memories but nevertheless retains much of his personality traits. Tek still tells the same stupid jokes. Tek still has many of the same hobbies. Tek is still an expert at many tasks (despite failing to remember how he acquired these skills). Given this sameness of personality, we might be inclined to say that Tek at t_2 is still Tek at t_1 even though Tek at t_2 fails to remember anything about Tek at t_1 . In short, to remedy the issue, we might rely upon other facets of one's psychology in order to state the criteria for A being the same person as B.

Definition 5.19: Personality theory

If A is a person at time t_1 and B is a person at t_2 , then B is the same person as A iff either

1. B at t_2 can apparently remember the actions or experiences A at t_1 , or
2. there is some person C who between t_1 and t_2 such that A can apparently remember the experiences of C and C apparently remembers the experiences of B, or
3. A has the same personality as B.

The personality theory of personal identity takes the memory-conditions from the memory theory and adds a further condition in the case that there is a memory gap between A and B. Namely, we will say that if A does not remember B, but A and B have the same personality then A and B are the same person.

The initial problem with this theory is that the property of *having the same personality* is extremely unclear.⁹ Intuitively, we think that a person's personality consists of a wide variety of things: beliefs, goals, loyalties, habits, personality quirks, interests. And so having the same personality would consist of these things staying the same over time.

However, the problem is that an individual's beliefs, intentions, habits, etc. change over time. And so, there is hesitation to say that any change in personality makes you a new person. Instead, one desiderata for the theory seems to be that any theory of personal identity over time is capable of handling some amount of change.

Exercise 5.38

Identify one way that your personality has significantly changed over time. If you can't think of one, who is someone you know whose personality has significantly changed.

5.5.6 Psychological-connectedness theory

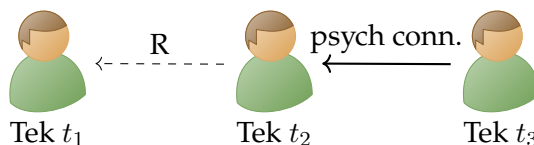
We can solve the primary problem with the same personality theory in the same way we solved the problem that faced the same psychology theory. Rather than requiring direct psychological connections between persons, we only need to be an **overlapping chain of psychological connections**. That is, A and B are the same given all of the memory conditions hold or there is psychological continuity between A and B .

Definition 5.20: psychological-connectedness theory

If A is a person at time t_1 and B is a person at t_2 , then B is the same person as A iff either

1. B at t_2 has the memories of the experiences of A at t_1 , or
2. there is some person C who between t_1 and t_2 such that A remembers the experiences of C and C remembers the experiences of B , or
3. there is psychological continuity between B and A .

In other words, A is B iff there is psychological connectedness between one complex set of beliefs, habits, memories, etc. and another.



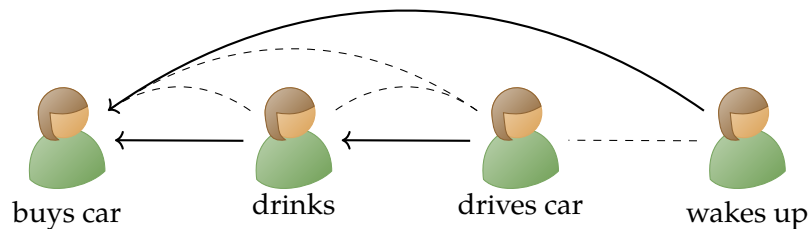
What is a psychological connection? A psychological connection would be direct causal or informational links between mental states at different times. Some examples include:

1. Memory: Tek's memory (mental state) of his high school graduation is caused by an earlier mental state (experience of high school graduation).
2. Belief formation: Tek's belief that it is important to be kind to neighbors is based on experiences as child.
3. Belief continuity: Tek's belief that it is important to be kind to neighbors is based on an earlier belief that it is important to be kind to neighbors.

4. Trait carryover: Having consistent personality traits, e.g., tendency to tell inappropriate jokes.
5. Desire chains: Tek wants to sign up for a marathon. This desire is causally connected to his desire to run a marathon.
6. Information retention: Tek has the experience of understanding French. This mental state is causally connected to thousands of experiences he has had studying and practicing French.

Let's illustrate this theory with two examples.

Suppose Liz desires to buy a new car. She goes to the dealership and buys the car. To celebrate, Liz begins drinking heavily and blacks out. She then, quite inappropriately, drives home. When she wakes up, she does not remember driving home but does remember buying the car.



According to the psychological connectedness theory, Liz survives the entire chain of events because at each stage there are psychological links connecting each person. These connections exist even though Liz there is a gap in her memory.

Here is a second example. Suppose Tek wants to go to Liverpool to visit his sister. He misses her. In order to visit her, Tek buy a ticket from Chicago to Liverpool. Tek arrives early to the airport in Chicago. However, as he is walking toward his airport gate, he momentarily forgets the initial reason why he wanted to go to Liverpool, that he bought the ticket, or how he got to the airport. Nevertheless, Tek sees the ticket in his hands and thinks "Ah! I really want to see my sister in Liverpool. I don't remember buying this ticket but I must have bought it since it is in my hands. So, I will use this ticket to see her."

On the memory theory, Tek would not be identical with the person who (1) originally desired to visit his sister, (2) bought the ticket, or (3) traveled to the airport. But notice that despite Tek's inability to remember these events (his lack of personal memories of these events), aspects of Tek's psychology cut across the dark gap in Tek's memory. Tek's desire to visit his sister in Liverpool still remains, his willingness to act on this desire still remains, and,

it is very likely, countless other psychological features of Tek carry forward as well, e.g., his sense of humor, his light-heartedness about his forgetfulness, his values, the things that Tek finds beautiful or interesting, and so on.

Intuitively, this gets things right, Tek on the airplane en route to Liverpool is the same person who bought the ticket in Chicago.

Now, besides the intuitive nature of the theory, and how it appears to be preferable over memory accounts, the psychological theory seems to deliver the right result in several scenarios.

1. Fixes the memory gaps problem facing the modified memory theory
2. Gives the intuitively correct result for brain transplants
3. Gives the intuitively correct result for persons with dicephalus
4. Gives the debatably intuitively correct result for artificial brain replacement

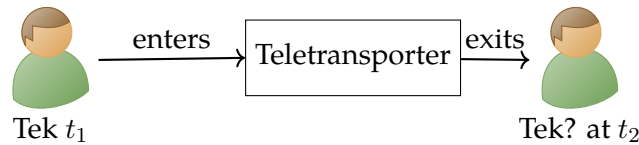
With respect to brain transplants, suppose that you are skeptical that the psychological-connectedness theory is true. You are inclined to think that what makes an person the same person over time is having the same body. If this is the case then consider the following example. While not currently medically possible, suppose that you and a friend were to swap bodies. That is, a doctor removes both of your brains and puts your friend's brain in your body and your brain in your friend's body. If this were to occur, you would likely say that *you* would continue to exist but your existence would take place in your friend's body. To put this more precisely: if *A* and *B* were to swap bodies at t_2 via brain transplant, *A* would continue *A's* existence in *B's* body while *B* would continue *B's* existence in *A's* body.

5.5.7 *Split personality*

TBA

5.5.8 *Teletransportation Device*

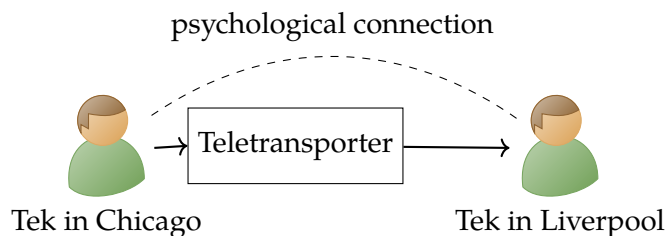
Let's consider a thought experiment. Suppose it is 2075 and current forms of transportation are obsolete. People no longer wish to drive or take the bus and instead use a new way to travel. The primary way to travel is to use the tele-transporter 9000. This machine takes a 3D scan of your body, stores this information, incinerates your body, then sends the information to another location, where a person is assembled using new materials and the data scanned by the tele-transporter 9000.



Let's walk through a sample teletransportation involving Tek:

1. Full 3D model of Tek's brain is scanned at t_1 in Chicago. Call this person Tek_1 .
2. Data from the model is sent to Liverpool.
3. Data from the model sent to Liverpool is used to create a new body in Liverpool that has a personality indistinguishable from Tek. Call this person Tek_2 .
4. Tek_1 's body in Chicago is destroyed and Tek_2 's body in Liverpool wakes up.

Now let's suppose Tek_2 (1) believes they are Tek_1 , (2) has detailed memories of Tek_1 's life, (3) has the same beliefs and habits as Tek_1 , and (4) has the same goals as Tek_1 . In fact, Tek_2 recalls buying a ticket at the teletransporter station, being scanned in Chicago at t_1 , falling asleep, and then waking up in Liverpool at t_2 . Since there appear to be strong psychological connections between Tek and the person who wakes up in Chicago, we are tempted to say that Tek_1 is one and the same person as Tek_2 .



Why wouldn't we? First, Tek believes that he is that person. Generally, unless we have strong evidence to the contrary, we do not generally doubt people are who they say they are. Second, Tek acts just like the person who entered the teletransporter. Again, he has all of the same memories, same jokes, same beliefs, same values, etc., etc. If he were asked where he was before he arrived in Liverpool, Tek would say "Ah! I entered the teletransporter in Chicago. I'm quite busy since I'm on my way to visit my sister." So, intuitively, Tek survives teletransportation.

Some people might worry about whether to engage in teletransportation due to its novelty. To address these worries, let's suppose that tele-transportation is rampant. Reviews of the teletransportation procedure are all positive (10,000 five-star reviews), with comments reading "I can't believe we used to drive.

Teletransportation is so much easier'. Tek has undertaken teletransportation many times. His friends and family have undertaken this mode of transportation, and it has greatly increased the productivity and joy of society. Individuals no longer miss out on important family events. Father dying? No longer must you take off a week of work to travel to your father's home in rural Montana. Instead, after work you use the teletransporter to materialize in the hospital. Thus, travel by teletransporter has become a commonplace in society.

With respect to the theories that we have so far considered, only the psychological theory would have Tek survive. He surely perishes on the body and brain theories since no physical part of him that goes into the teletransporter in Chicago is spatiotemporally continuous with any part that comes out the other teletransporter in Liverpool. And, it is generally thought that while on earth, the soul is tied to its body, the destruction of the body would suggest that the soul no longer exists here on earth. So, in short, the psychological theory gives us the intuitively right result.

Exercise 5.39

Suppose people make use of consciousness travel on a regular basis. They describe being scanned, feeling a tingle, and waking up in a new location with the memories of the past, with the same habits, the same skills, the same character traits and so forth.

1. Would you undertake consciousness transfer?
2. Do you think people retain their personal identity through the travel (or do they die)?
3. Draw a picture to illustrate your view on the matter.

Exercise 5.40

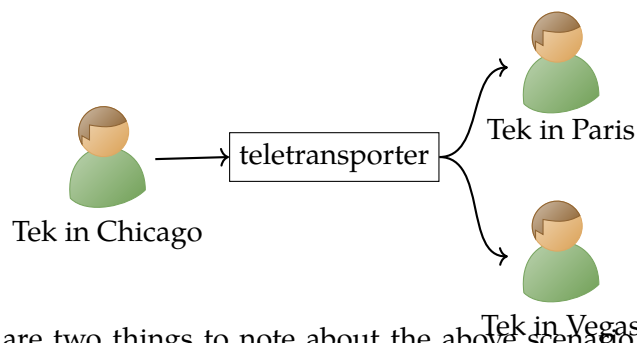
This sort of consciousness transfer appears in fiction. For example, in the television show *Black Mirror* (San Junipero) people upload their personalities to a VR world. Many of the individuals have terminal diseases, others had diseases and are now dead. In *Neon Genesis Evangelion* (S1:E13 Angel Infiltration), a partially organic supercomputer (MAGI System) is built using portions of a person's personality. Create your own version of a consciousness transfer or upload.

5.5.9 Fission (Reduplication) Problem

Let's consider a rather serious problem for the psychological theory. Reduplication problems occur when Tek is multiply duplicated. So, let's consider a scenario where Tek is rematerialized twice. Suppose the following:

1. Tek is scanned in Chicago (he wishes to be sent to Toronto)
2. Information is sent to Paris and Las Vegas
3. Tek is rematerialized in Paris and Las Vegas

In this scenario, Tek is rematerialized in two different locations (suppose that his body is destroyed in Chicago after the information is sent to both Paris and Las Vegas).



Now, there are two things to note about the above scenario. First, in the duplication scenario noted above, we said that Tek in Chicago is the same person as Tek in Liverpool. We take this to be the case because Tek in Chicago and Tek in Liverpool are psychologically connected. Second, notice that both Tek in Paris and Tek in Las Vegas are psychologically connected to Tek in Chicago since:

1. Tek in Paris and Tek in Las Vegas remember entering the teletransport machine in Chicago; they remember being anesthetized; they remember what meetings they must keep in Toronto
2. both remember a multitude of facts about their prior experiences (what they did at summer camp, etc.)
3. both feel like they are Tek (they feel as though they are Tek just as strongly as you feel that you are you)
4. both remember who they love, what their bank account password is, what they like to do, etc.

Given that psychological connectedness ensures survival, we have:

- Tek-Vegas = Tek-Chicago
- Tek-Paris = Tek-Chicago
- Therefore, Tek-Vegas = Tek-Paris

However, there is a strong case to be made that Tek in Paris and Tek in Las Vegas are not the same person:

1. Tek in Paris and Tek in Las Vegas are in two different locations
2. One of them can exist while the other one perishes
3. Neither can tell what the other one is thinking
4. A person only wholly exists in one distinct, wholly unseparated space and it is presumably impossible that a person can be wholly found in two distinct, separated places

Therefore, in contrast to Therefore, Tek-Vegas = Tek-Paris, there is an equally strong intuition suggesting that Tek-Vegas $\neq$ Tek-Paris. But this is inconsistent.⁴

5.5.10 Solutions to the Branching Problem

Let's consider some possible responses to the reduplication problem.¹⁰ Each of these responses should be understood as a response that aims to save the psychological theory but deal with the problematic consequence of reduplication.

1. *Pick one*: Tek in Paris, not Tek in Las Vegas, is identical to Tek in Chicago
2. *Pick one*: Tek in Las Vegas, not Tek in Paris, is identical to Tek in Chicago
3. *Neither one*: Neither Tek in Las Vegas nor Tek in Paris are identical to Tek in Chicago
4. *Both*: Both Tek in Las Vegas and Tek in Paris are identical to Tek in Chicago
5. *No answer*: There is no fact to the matter
6. *Gradual separation*: Tek-Vegas and Tek-Paris are Tek-Chicago, but become distinct from each other.
7. *Reject reduplication*: I don't need to worry because reduplication is impossible.
8. *Reject the psychological approach*: There is no solution and so the psychological theory should be rejected.

Let's consider these options.

⁴It is worth pointing out that simply performing many of these thought experiments would not solve the problem. As Parfit points out: "Suppose, for example, that I do survive as one of the resulting people. I would believe that I have survived. But I would know that the other resulting person falsely believes that he is me, and that he survived. Since I would know this, I could not trust my own belief. I might be the resulting person with the false belief. And, since we would both claim to be me, other people would have no reason to believe one of us rather than the other. Even if we performed this operation, we would therefore learn nothing." [95, p. 258]

5.5.10.1 *Pick one*

Options 1 & 2

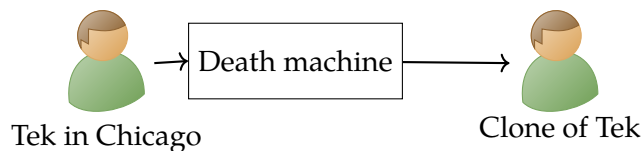
First, note that options 1 and 2 are inconsistent. If (1) is true, then (2) is false. If (2) is true, then (1) is false. However, the choice between the two seems entirely arbitrary. There is nothing about being in Las Vegas or being in Paris that is relevant to determining which individual is identical to Tek. Both individuals are exactly similar from a psychological point of view.

5.5.10.2 *Neither one*

Option 3

Second, option (3) contends that neither person who exits the teletransportation device is the same person as Tek. There appears to be two reasons to support this view. First, no one ever survives teletransportation. Second, people do survive teletransportation but they never survive it when branching occurs. Let's consider both reasons.

First, one might argue that Tek does not survive teletransportation because no one survives teletransportation. That is, the teletransportation device is a death machine. The person (or being) that exits the teletransportation device is not us.



However, compare the following two cases:

- Case 1: Tek stays in Chicago and wanders around the airport.
- Case 2: The single duplication case where Tek is teletransported from Chicago to Liverpool.

If the psychological theory is true, then there does not appear to be any relevant *psychological* difference between the psychological-connectedness of Tek in the two cases. In Case 1, Tek maintains his identity because he is psychologically continuous with the Tek that is wandering around the airport, not because he has the same body or soul. But this is the same grounds for saying Tek survives teletransportation: he is psychologically continuous with the Tek in Liverpool. The only relevant difference between the two cases is that in the former, Tek has bodily continuity, but in the latter, Tek does not. But since bodily continuity is not relevant for survival (according to the psychological connectedness theory), this cannot be the difference that explains why Tek would perish if he entered the teletransporter.

Let's consider the second option that Tek does not survive teletransportation

non-branching

when branching occurs. The relevant difference between cases 1 and 2 and the branching case is that there are two Tek's psychologically continuous with an earlier Tek. So, rather than abandon option 3 too quickly, one approach to salvaging the psychological continuity theory of personal identity is to add a non-branching condition to the theory.

Definition 5.21: non-branching condition

If A is a person at time t_1 and B is a person at t_2 , then B is the same person as A iff there is psychological continuity between B and A and there is no other person C who is psychologically continuous with A.

This revision seems to get us the right results with respect to the teletransporter case. That is, if Tek in Chicago enters the transporter and is sent to Paris, assuming there are no other individuals continuous with Tek in Chicago, then Tek in Paris is identical to Tek in Chicago. However, in the case of reduplication, personal identity is not preserved. That is, in cases that involve branching of psychological continuity, Tek in Chicago does not survive.

ad hoc There are several problems with the addition of this condition. First, this revision is completely arbitrary. There is no other motivation for why personal identity should be non-branching except to combat the objection.

denies the only x and y principle

Second, this condition denies the "only x and y principle".

Definition 5.22: only x and y principle

The only x and y principle states that if x and y are identical, then this identity only depends upon x and y, and not some third item z.

Let's illustrate this principle with an example. If Superman is Clark Kent, then this depends upon facts about Superman and Clark Kent, and not on some other fact. The non-branching psychological continuity theory denies this principle since whether Tek is identical to another individual that he is psychologically continuous with always depends upon whether there is some other individual that is in that same relation.

5.5.10.3 *Both*

Option 4 Option 4 contends that Tek in Las Vegas and Tek in Paris are *both* Tek in Chicago. On this view, Tek-Vegas is one and the same person as Tek-Chicago *and* Tek-Paris is one and the same person as Tek-Chicago. To understand this

option, notice that in non-branching (normal) cases, Tek survives moment to moment as one person. When we talk about Tek at t_1 and Tek at t_2 , there is *one person* that is located in two different *temporal* locations. In the branching case where we say that Tek-Chicago survives as both Tek-Vegas and Tek-Paris, there is *one person* that is located in two different *spatial* locations.

As mentioned, there are several reasons that this result seems absurd.

First, persons are spatially contiguous. But, in this case, one person is not spatially contiguous. Tek is in two spatially separated locations: Vegas and Paris.

Second, a single person knows what they are thinking but they don't know what other persons are thinking. For example, I know what I'm thinking, but I don't know what you are thinking. Based on this fact, I know that I am not you. Tek in Las Vegas does not know what Tek in Paris is thinking, feeling, or experiencing, and vice versa.

Third, if a person dies, they don't continue living. It is possible that people can 'come back from the dead' but in this scenario, a person can die but keep living like nothing has happened. For example, if Tek (Tek-Paris) dies, then Tek (Tek-Vegas) continues to live.

Fourth, People can recognize themselves. But, suppose that the two Tek-Vegas and Tek-Paris (who were once connected) moved to different sides of the world and developed various experiences, memories, etc. If they met each other, "they might even fail to recognize each other".[94, p. 7]

5.5.10.4 No answer

Option 5 says there is no fact to the matter. Consider that we have been Option 5 operating under the assumption that for questions of the form "Is X the same person as Y" always has a true answer. However, several philosophers have suggested that this is not the case. According to Parfit,[94, p. 8]

"If all the possible answers are implausible, it is hard to decide which of them is true, and hard even to keep the belief that one of them must be true. If we give up this belief, as I think we should, these problems disappear."¹¹

So, for Parfit, the logic is that since all answers are unsatisfactory, there is not always a true answer to the question of "Is X the same person as Y".

Kind[66] motivates this sort of answer by pointing to the objective indeterminacy of vague predicates. That is, suppose you are God. You have all

the information about the world. Now suppose you are asked to determine whether "Tek is tall" is true or false. But, Tek is a borderline cases of tall. He isn't clearly tall, and he isn't clearly not tall. You know Tek's height exactly. You know how tall Tek is in comparison to all of the other people in his comparison class (that is, relative to all other human beings). Even with all of this information, you still not know whether Tek is tall or not. There may simply be no fact to the matter concerning this borderline case. Applied to personal identity, there may simply be no fact of the matter concerning whether Tek in Las Vegas or Tek in Paris is identical to Tek in Chicago. Even if you knew all of the facts, there is simply nothing that determines things one way or another.

The objection to this claim is that intuitively we think there is an answer to a number of questions that depend upon precisely this issue. Namely, Tek may wonder whether or not he is going to die or not if he uses the teletransporter machine. He may ask "will I die if I enter the machine?". Intuitively we think there is an answer to this question, but such an answer depends upon whether his personal identity is preserved.

Exercise 5.41

So, which one of the five options is correct?

5.5.10.5 *Gradual separation*

In my experience, option (6) is the most common response from people who first encounter the problem. This position is a type of *hybrid* solution as it combines option (4) and options (1) and (2).

"at the initial moment of fission, Tek-Vegas and Tek-Paris are the same person as Tek-Chicago (option 4). After the two bodies go their separate ways, Tek-Vegas and Tek-Paris are the same person as Tek-Chicago but they are not the same person as each other due to their differing experiences."

This view is problematic for several reasons.

First, while it is clear that this position accepts Option (4), it seems to assert that if a person *A* has an experience not had by person *B*, then $A \neq B$. But, if this were true, then it is unclear why Tek-Vegas and Tek-Paris are the same person as Tek-Chicago since the former two people have experiences not had by Tek-Chicago. In other words, the position seems to assume that no one could ever survive any change since we are constantly having new experi-

ences, e.g., Tek-2025 has experiences that Tek-2024 has not had. Therefore, Tek-2025 and Tek-2024 are not numerically the same person.

Second, option (6) seems to imply that identity is not transitive. That is, typically we think that if $A = B$ and $A = C$, then $B = C$. However, in this case,

1. Tek-Chicago is the same person as Tek-Vegas.
2. Tek-Chicago is the same person as Tek-Paris.
3. But, Tek-Vegas is *not* the same person as Tek-Paris.

Third, in any given location there is, in that location, one and only one object wholly occupying that space. If I am looking at a chair, there may be multiple legs occupying some of the space of that chair, but there is one (not two or three or four) chairs wholly occupying the space of that chair. But, if Tek-Vegas and Tek-Paris are different people and they are one and the same person as Tek, then it appears to follow that there were two distinct people wholly occupying Tek-Chicago. That is, this view appears to multiply the number of people occupying a single space. These persons had the same thoughts, beliefs, and memories until they were separated.

5.5.10.6 *Reject reduplication*

Rather than rejecting the psychological-connectedness theory, Option 7 rejects the falsifying case by saying that teletransportation is impossible.¹² If the psychological theory is true, then it seemingly entails (in reduplication examples) the contradictory conclusion that two distinct people are the same person. However, if reduplication cases are impossible, then no such contradictory conclusion follows. There are three problems with option.

Option 7

First, while fission cases are not *presently* technologically possible, this does not mean they are not technologically possible in the future. A supporter of Option 7 thus needs to either assert (1) future problems are not problems or (2) fission cases are not technologically possible in the future.

Second, even if fission cases are not technologically possible in the future, this does not mean that they are not logically possible, physically possible, nor technologically possible. So, someone who asserts that fission cases can be dismissed in this way needs to show why the only metaphysical issues worth considering are those that are technologically possible, while other possible scenarios can be dismissed.

Third, the examples of fission cases presented here are canonical and extravagant. This does not imply that other, less extravagant, cases that are tech-

nologically possible cannot be presented. For example, rather than a tele-transportation machine, consider some form of biomedical engineering that involved the fission of human beings. Another would involve a dystopian future where global warming is out of control and the earth is no longer hospitable. In order to escape certain death on Earth, individual upload their consciousness to a device. However, in some rare cases, there are multiple uploads. That is, the same individual is uploaded twice.

Fourth, it is possible that a supporter of option (7) misunderstands the purpose of the fission case and thought experiments more generally. One reason thought experiments are employed is to test metaphysical theories that, typically, are not testable by more grounded empirical means. A common approach to testing a theory is to suppose that the theory is true and see whether the theory leads to any absurd (contradictory) results or results that clash with the sciences. Fission cases are problematic since they appear to lead to an absurd (contradictory) results, where the theory seems committed to saying that B and C at t_i both numerically identical to person A but numerically distinct from each other.

5.5.10.7 *Reject the psychological approach*

A final option, option (8), is to reject the psychological approach entirely. Since it is unable to handle problems relating to fission, one ought to reject the theory. However, one problem with this approach is that there appears to be no better alternative.

5.6 HYBRID VIEWS

If the soulful theory, various physical theories, and psychological theories do not yield satisfactory results, then perhaps either (1) there is simply no answer to the question of diachronic personal identity (skepticism) or (2) the solution is found in combining the various approaches we have thus far considered. There are at least two ways to do this.

The first way is to say that survival requires both bodily (brain) continuity *and* psychological continuity. There seems some attractiveness to this idea since we identify ourselves with our psychology (memories, beliefs, character traits, etc.) but also acknowledge the important relation our bodies play in determining our psychology.

Definition 5.23: conjunctive psycho-physical theory

If A is a person at t_1 and B a person at t_2 , then $A = B$ iff there is psychological **and** physical connectedness between B and A

In some discussions about personal identity, it is claimed that a person retains their identity over time provided there is a relation R between persons.

Definition 5.24: Relation R

R is a relation between two persons X and Y consisting of direct psychological connections *and* psychological continuity that is produced by a “normal cause”, namely the “continued existence of the same brain”.¹³

The conjunctive theory faces problems that the psychological and physical theories face independently.

First, there are *fission problems*. With respect to the physicalist theories, we saw that the brain theory faced a fission problem regarding the two hemispheres of the brain being transplanted into two different bodies. In such a case, if A is the whole brain person and B and C are bodies that receive each of the hemispheres, then $A = B$ and $A = C$ but $B \neq C$. This is a contradiction. Since the conjunctive theory requires both psychological and physical connectedness, it would seem that the conjunctive theory would also face the same problem.

Second, consider the claim that you were once a fetus. One of the reasons some people find purely physicalist theories attractive is because the theory is consistent with this claim. However, if personal identity requires both continuity of body the body and the psychology, then you were never a fetus since there is no psychological connectedness. fetus problem

Exercise 5.42

In saying that you need both the bodily and psychological continuity to survive, the theory is made stricter (you need two meet two conditions to survive). What other problem might this theory face?

Rather than saying that survival requires both bodily and psychological continuity, another hybrid approach is to say that survival requires one of the two. That is, A survives as B if and only if B is either bodily or psychologically continuous with A . Such a theory is known as a “disjunctive theory of personal identity”.

Definition 5.25: disjunctive psycho-physical theory

If A is a person at t_1 and B a person at t_2 , then $A = B$ iff there is psychological **or** physical connectedness between B and A

The disjunctive theory seems to be consistent with the claims that you preserve your identity in cases involving brain transplant, the upload/teletransportation cases, and the fetus case. One problem with this approach is that the theory inherits several problems of each theory.

First, we saw that the psychological theory faced a fission problem involving *reduplication* using a teletransporter. If A is a person that undergoes reduplication as B and C , then $A = B$ and $A = C$ but $B \neq C$. The disjunctive theory would also face the same problem.

Second, the theory inherits new problems. Consider that the disjunctive theory contends that $B = A$ iff there is either physical or psychological connectedness between A and B . Imagine two individuals Tek and Liz. Now suppose we put Tek's brain (or psychology) in Liz's body and Liz's brain (or psychology) in Tek's body. According to the disjunctive theory, Tek is the following:

1. the Liz body that has Tek's brain (psychological connectedness)
2. the Tek body that has Liz's brain (physical connectedness)

But this cannot be the case, because the Liz body with Tek's brain is not the same as the Tek body that has Liz's brain. Contradiction!

Further complicating the issue, according to the disjunctive theory, Liz is the following:

1. the Liz body that has Tek's brain (psychological connectedness)
2. the Tek body that has Liz's brain (physical connectedness)

But this cannot be the case, because the Liz body with Tek's brain is not the same as the Tek body that has Liz's brain. Contradiction!

Exercise 5.43

Hybrid views Is there any way to modify a hybrid reductionist view to solve the problem of personal identity?

5.7 PARTIAL IDENTITY

Another approach to the problem of personal identity is to reject the idea that personal identity is an all-or-nothing relation. Rather than saying that two

individuals are either identical or not identical, we might say that whether A is identical to two future persons B and C is a matter of degree.

TBA

5.8 WILD COHERENCIES

TBA

5.9 SKEPTICAL THEORIES

TBA

5.10 THE SIMPLE VIEW AKA THE SOUL THEORY AGAIN

TBA

5.11 FUSION OF PERSONS

In our discussion of personal identity, we have considered various fission cases: (1) the separation and transplanation of each of a person's hemispheres into a different body and (2) the reduplication of a person via a teletransporter. Fission cases pose a challenge for both physicalist and psychological theories in that the theories seemingly permit that a person can become two people. However, intuitively, we reject the idea that there could be two Teks, walking around in different parts of the globe, thinking different thoughts, having different lives. In essence, fission cases help to test the limits of these theories.

Rather than fission, let's consider the *fusion* of persons and consider whether fusing persons can provide insight into which theory of diachronic personal identity is true. The question of fusion is whether it is possible for two or more persons to survive as a single person.

Let's start with a physical fusions before considering psychological cases.

5.11.1 *Body fusions*

In considering the fusion of persons, we need to consider two questions:

1. Has fusion occurred? Was there two objects at t_1 and one object at t_2 ?
2. Have both persons survived?

Let's call this the *Contact* case. Suppose I grab your arm. I claim that because we are in contact, we are one body. And, because I believe in the body-theory

of personal identity, I claim that we are one person. Surely fusion has not occurred but both persons survive.

	fusion?	survival?
contact	x	2

Next, let's call this the *Contact and Weak Fastening* case. I have some superglue and I glue my hand to your arm. I claim that because we are in contact and there is a force binding us together that we are one body. And, because I believe in the body-theory of personal identity, I claim we are one person. Surely fusion has not occurred but both persons survive.

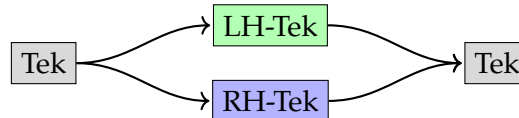
Next, a *Contact and Strong Fastening* case. It has been many years since I glued myself to you. Lucky for you, we were pulled apart using various home remedies (oils, acetone). However, the event got me thinking. Perhaps I could join our bodies in a way where we are not so easily separated. I say to myself, if a wound can be stitched closed, then perhaps I can stitch our bodies together. Again, since I hold the body-theory, and our body is both in contact and a strong force holding us together, I claim that since we have the same body, then we are the same person. Surely fusion has not occurred but both persons survive.

Next, a *Shared organs* case. Suppose a surgeon has removed my stitches and we both received skin grafts to cover the damage to where we were stitched together. I say to myself, "I've got it! Rather than merely stitch our skin together, what about sharing organs!" I prepare an operation where we alternate whose organs we will use, e.g., we will use my heart but your kidneys, your liver but my lungs. At the end of the operation, we are one body with two heads. Again, since I hold the body-theory, and our body is both in contact and a strong force holding us together, I claim that since we have the same body, then we are the same person. Surely fusion has not occurred but both persons survive.

Next, a *Bodily control* case. Suppose a person *A* that is 40 meters tall. A cockpit-like device is crafted onto person *A*'s body. This cockpit allows person *B* to get inside of person *A*. In addition, cybernetic components are grafted onto *A*'s body so that, when *B* is inside of *A*, *B* can control *A*'s body. When inside the cockpit, a fluid (from *A*) fills the cockpit and then fills the lungs of *B*. After less than a minute, *B* finds that they can alternate control of *A*'s body (the biological interface must be powered by a battery). In this case, it is unclear whether this is a case of fusion, but it seems as though both persons survive.

Consider a *Hemisphere transplant-and-fusion* case. Suppose we were to re-

move both of the hemispheres of Tek and place them into different bodies LH-Tek is the left hemisphere of Tek and RH-Tek is the right hemisphere of Tek. Suppose these individuals go on and lead separate lives. LH-Tek becomes a writer, gets married, and moves to Paris. RH-Tek becomes a painter, never marries, and moves to Chicago. This appears to be a case of fission. But now suppose that the two hemispheres are reconnected. According to Parfit, once reconnected, we would take the two experiences and calculations both to belong to the reconnected person. [94, pp. 6–7]



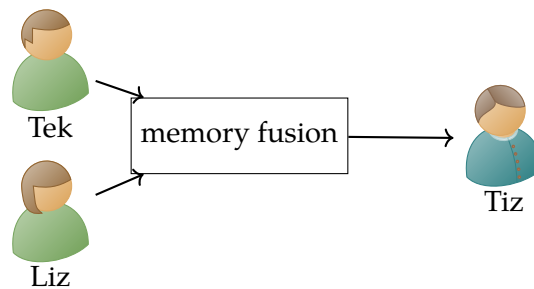
This appears to be a case of fusion since Tek was a person at t_1 , then split into two persons LH-Tek and RH-Tek at t_2 , and then was recombined into Tek at t_3 .

Exercise 5.44

First, create your own case of bodily fusion (be detailed and include a drawing). Second, does your case fall under one of the above kinds of cases? Third, consider each of the above cases of fusion (and your own) and determine whether or not we survive fusion? What, if anything, does this tell you about who you are?

5.11.2 Psychological fusions

Let's consider psychological fusion. Parfit says that when two people are fused together, it is not necessary that any of the memories are lost but “some things must be lost.” [95, p. 298] Let's consider an illustration. Suppose Tek and Liz are fused together as Tiz.



Suppose Tek remembers going to a dance in the 5th grade while Liz has no such memory. Their resulting fusion would not result in a loss of memory:

only the addition. Tiz would now remember going to the dance in the 5th grade. Also, suppose Tek vividly remembers visiting the Grand Canyon as a child while Liz remembers summers at home. Tiz would have the rather confusing memory that they both went to the Grand Canyon and stayed home during the summer.

So surely it is possible to fuse memories. What about other aspects of a person's psychology. Parfit claims that when *conflicting* "characteristics, different desires, and different intentions" of the persons are fused, something of the pre-fused person can get lost. Parfit writes,

"Some of these features will be compatible. These would co-exist in the one resulting person. Some will be incompatible. These, if of equal strength, would cancel out, and, if of different strengths, the stronger would become weaker. These effects might be as predictable as the laws governing dominant and recessive genes." [95, p. 298]

For Parfit then,

1. some character traits can coexist
2. equally dominant traits cancel each other out
3. dominant traits would be weakened by weaker traits (and likely weaker traits even weaker)

Some examples of non-compatible traits:

1. I love Liverpool Football Club. I am going to be fused with someone who equally loves our rival. The result is someone who waffles between clubs.
2. From Parfit: "I admire Palladio, and intend to visit Venice. I am about to fuse with someone who admires Giotto, and intends to visit Padua. The one resulting person will have both tastes and both intentions. Since Padua is close to Venice, both can be easily fulfilled. Suppose next that I love Wagner, and always vote for a Socialist. The other person hates Wagner, and always votes for a Conservative. The one resulting person will be a tone-deaf floating voter." [95, p. 298]

Exercise 5.45

Give an example of psychological fusion involving conflicting characteristics.

What should our attitude about survival be in cases of psychological fusion? In contrast to cases of fission where the duplicate person is nearly psycho-

logically identical to us, in cases of fusion, the intuition is that we would *not* survive. Parfit writes,

“If we were about to undergo a fusion of this kind, some of us might regard this as equivalent to death. And this is less absurd than regarding division as equivalent to death. When I divide, the two resulting people will be exactly like me. When I fuse, the one resulting person will not be wholly similar. This makes it easier to think, when faced with fusion, 'I shall not survive', thus continuing to regard survival as being all-or-nothing.”[95, p. 288]

Let's call the thesis that no one ever survives fusion the *No Survival Thesis*.

Next, let's consider two reasons why this thesis is false. First, Parfit and others point out that fusion (and the psychological relations between the pre-fused and post-fused person) could come in various degrees. For example, suppose that Tek and Liz were fused but the resulting fusion (Tiz) consists of all of the important psychological components of Tek and one insignificant psychological component of Liz. Here we might be inclined to say that Tek survived but Liz did not. So, not all cases of fusion clearly result in death.⁵

Small degree fusion

Second, consider a case where the degree of psychological connections between two persons *A* and *B* is low but the important psychological connections remain (important memories, beliefs, values, characteristics). Let's call this the *Improvement Scenario*. Suppose an experimental therapy that would *remove* several of your bad memories and habits (the death of a family member, failing an exam, smoking, inattentiveness, picking the skin near your nails, etc.). Some people might be willing to undertake such a medical procedure as they would take it to *fix* rather than *destroy* who they are.

Improvement Scenario

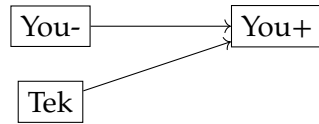


Now suppose that in addition to *removing* several of your bad memories and habits, the procedure can also (for an additional fee) *add* several memories and traits (e.g., pleasant memories of sunsets, vacations, improved concentration). Again, people might be willing to undertake such a medical procedure as they would take it to *improve* rather than *destroy* who they are.



⁵Consider the following example from Ehring.[38, p. 332] Suppose we scan the brain states of Tek and Liz. Rather than fusing them into a single body, imagine they are fused into a spectrum of bodies, ranging from those where Tek's characteristics dominate to those in which Liz's characteristics dominate.

Finally, suppose the fixing-and-improving procedure is performed via fusion with a suitable candidate (Tek). Tek does not have the bad memories (and habits) you have and also has several of the memories (and habits) you desire. The new and improved “you” is formed by fusing you with Tek.¹⁴



The only difference between the non-fusion and fusion cases concerns the causal scenario. In both cases, $A-$ is transformed into $A+$. So, if $A-$ survives as $A+$ in the non-fusion case, then what grounds is there for saying that $A-$ does not survive in the fusion case.

Exercise 5.46

Under what scenario, if any, would you survive fusion in the Improvement Scenario? What matters to you for survival in fusion cases?

ARE THERE INTERMITTENT OBJECTS?

Do ordinary material objects ever exist intermittently? Suppose the following:

1. At t_1 , Tek hands his watch to a watchmaker.
2. The watchmaker then disassembles the watch.
3. At t_2 , there are the watch parts lying on the table.
4. The watchmaker then reassembles the watch.
5. Let t_3 refer to the time that the parts from t_2 have been reassembled into a watch where the parts are in the same relative places as the parts of the watch at t_1 .

There are two questions:

1. Is the watch at t_3 one and the same watch as the watch at t_1 ? When Tek is handed a watch, is the watch the he is handed one and the same watch as the one he handed the watchmaker?
2. Does the watch stop existing at t_2 when it is disassembled?

There are intermittent objects if the answer to both of the above questions is affirmative.

Definition 6.1: intermittent object

An intermittent object is an object that exists at t_1 , does not exist at t_2 , and then exists again at t_3 .

There are at least three plausible answers to the question of intermittent objects:

1. No. The watch exists, ceases to exist in its disassembled state, and then, once reassembled, is not the original watch.
2. No. The watch exists, continues to exist in its disassembled state, and then, once reassembled, is one and the same watch.
3. Yes. The watch exists, ceases to exist, and then exists again as one and the same watch.

In this chapter, we'll consider each of the above three answers in order. Finally, we'll conclude this chapter by considering some implications of the possibility that there are intermittent objects.

6.1 THERE IS A NEW OBJECT

With respect to the watch, one might claim that the watch exists, ceases to exist in its disassembled state, and then, once reassembled, is not the original watch. What support is there for this view?

Locke writes that “one thing cannot have two beginnings of existence”.¹ If we accept that the watch begins at t_1 , is not a watch at t_2 (just the parts), and is a watch at t_3 , then there is a seeming violation of Locke's principle: the watch has two points in time when it begins to exist as a watch. If we accept that there is a new watch at t_3 , this problem goes away.

There are two main problems with the above view.

First, this view clashes with how we interact with and understand the object in its *reassembled* state. It is therefore unintuitive. When Tek receives the reassembled watch from the watchmaker at t_3 , both he and Tek believe it to be Tek's watch from t_1 , not a *new* watch. Furthermore, suppose the watch was a gift from his grandfather. If the watchmaker, instead of disassembling it and cleaning it, simply handed Tek a brand new watch, Tek would argue that the watchmaker was still in possession of his grandfather's watch.

Second, this view also clashes with how we interact with and understand objects in their *disassembled* state. As Lowe[74, p. 222] puts it,

“Someone entering the workshop and seeing the pieces [of the watch] laid out carefully on the watchmaker's bench would quite properly be told “That is [Tek's] watch”; and if such a person were, say to stamp on these delicate bits of machinery, he would clearly be guilty of *destroying* [Tek's] watch, i.e., terminating its existence. So it *did* exist even in its dismantled condition.”

In short, our intuitions clash with the idea that an object cannot survive disassembly and reassembly and we tend to regard even the undisturbed disassembled object as still being the pre-disassembled object.

6.2 IT NEVER STOPPED EXISTING

Several philosophers appear to have held the view that (1) objects can survive disassembly and (2) that objects retain their identity in a disassembled form. Some examples.

Laycock contends that an object “can survive, can retain their identities through, being dismantled and reassembled. Continuity, in the identity, criterion,

¹An Essay concerning Human Understanding, Bk. II, Ch. XXVII, sect. 1.

cannot be construed as requiring the continued *intact* existence of the objection in question." Laycock[68, p. 28] notes that some objects have "lapses" in their intactness (e.g., cars, bridges, and bicycles) while others (e.g., trombones, tents, guns) may exist in their unassembled form most of their lives.

Price[103, p. 211] writes

"we would not say that a watch or stereo in need of repair ceases to exist when a jeweler or repairman takes it apart. If we did say this, repairing them would be logically impossible. So if watches and stereos are material objects, a material thing can persist as a scattered object."

For Price then, being a material object through time either does not require being spatiotemporally continuous or we have to understand 'spatiotemporally continuous' in a way that permits the parts of the object being discontinuous.

Smart[125, pp. 25–26] points out that we can apply when an object is disassembled, it is disassembled into parts. These are parts of a *whole*, a whole that exists but in a disassembled state.² Further supporting this idea is that we can rightly talk of the watch as *being disassembled*. The predicate 'is dismantled' or 'is disassembled' rightly belongs to the watch that exists in a disassembled state.³

What is wrong with the claim that objects survive in their disassembled state?

First, one might claim that if an object exists it not only must exist as object of some *kind* but its continued existence requires *sameness of kind*. Let's call this the diachronic sortal identity criterion:

Definition 6.2: diachronic sortal identity

An object *A* at t_1 is one and the same object as object *B* at t_2 only if (1) *A* and *B* are spatiotemporally continuous with each other and (2) *A* and *B* belong to the same dominant kind *K*.

²Smart writes that objects can be "dismantled and reassembled elsewhere, still retaining their identity, and that identity of parts [is] a defeasible condition of identity."

³For Smart,[125, p. 26][124, pp. 147–8] the disassembled object stops existing as the collection of disassembled parts when (1) the object replaces its disassembled part with a new part (for Smart the newly added part becomes a part of the object while the old object *had* been a part but is no longer), (2) when the disassembled part becomes the part of a new object, or (3) when it becomes causally impossible to reassemble the object.

The idea then is that you (a person) can only continue to exist as a person (your dominant kind). And so a watch can only continue to exist *as a watch*. It cannot continue to exist as a chair.

If that is the case, then it would be wrong to say that the watch exists at t_2 because it does not exist *as a watch*. As Lowe[74, p. 222] puts it with respect to ships, if a ship is disassembled and then later reassembled, there is a period of time where ``there was in fact no ship at all in existence consisting of those parts, i.e., there isn't a continuous "ship-history".'' Or, using our watch example, the dismantled watch does not exist because it is not a watch (dismantled Ks are not Ks).

In reply, someone might assert that the disassembled clock *is* a clock. It is a clock in a disassembled state. If this were true, then the principle diachronic sortal identity would not be violated. In response to this objection, let's consider another objection.

Second, consider the following scenario from Michael Burke.[14]

1. At t_1 , there is a table made from thirty pieces of wood
2. The table is then disassembled and made into a chair.
3. At t_2 , there is a chair.
4. The chair is then dismantled and parts are made into a stool and the remaining parts are made into a birdhouse.
5. At t_3 , there is a stool and a birdhouse.
6. The stool and birdhouse are dismantled, and all thirty pieces are made into a table of the same size, shape, style, etc. as the table at t_1 .
7. At t_4 , there is a table.

Burke then asks a pair of questions:

First, if the table survives as a disassembled object, where is it in space at t_2 ? It seems that a supporter of the above account must say that the table is where the chair is.

Second, if the table is where the chair is, then is that object a chair? Presumably, it is a chair. If you asked someone on the street if it was a chair, they'd say yet. You would sit in it like a chair. And, it fits the standard description of a chair.

Finally, let's state the ``one-object-to-a-place principle."

Definition 6.3: one-object-to-a-place principle

For any given spatial region R at an instant of time t , there can be at most one material object wholly occupying R .

If the table is located in the exact location of the chair, and that object is a chair, then there are *two* different material objects (a chair and a table) wholly located in the same spatial location at the same time. This violates the metaphysical principle that different material objects *cannot* wholly occupy the same space at the same time.

In short, our intuitions clash with the idea that objects like watches or ships can survive as non-watches or non-ships. And, the theory violates the principle that two distinct material objects cannot wholly occupy the same spatial location at the same time.

6.3 THERE ARE INTERMITTENT OBJECTS

Our discussion of problems with the prior theories appear to count in favor of the existence of intermittent objects.

First, intuitions appear to support that the watch at t_1 and the reassembled watch at t_2 are one and the same watch. Tek and the watchmaker behave as though it is the same watch.

It is also supported by the following two points:

1. The watches look the same (qualitatively identical)
2. The parts of the watch are the same⁴

Second, it was argued that when arguments are disassembled, they no longer exist.

1. If the disassembled watch is *not a watch* and the diachronic sortal identity principle is true, then the watch no longer exists when it is disassembled.
2. If the disassembled table is a table, then it can also be a chair, or stool, or birdhouse at the same time. But this violates the one-object-to-a-place principle.

To reconcile these two results, it seems as though we must say that there are intermittent objects. The watch exists in assembled form at t_1 , does not exist in disassembled form at t_2 , and then exists again in assembled form t_3 .

⁴As Burke[14, p. 403] notes, another consideration is that “The case is not of the ‘ship of Theseus’ type. No third object has competing claims to identity with the original table.”

What is wrong with this theory?

Burke raises two potential criticisms to the above. First, imagine your phone is sitting on your desk and it pops out of existence. Five minutes later it pops back into existence on your desk (or five miles away in your apartment). The object is qualitatively identical to your phone, but how do you know it is one and the same phone? Couldn't the phone just be a copy?

Second, again imagine your phone goes out of existence in location L , and, just like before, two qualitatively identical phones come into existence in two different locations. What criteria would decide which of the two phones is the original?

6.4 IMPLICATIONS OF INTERMITTENT EXISTENCE

Who cares if some objects have intermittent existence? In this section, we'll consider two implications.

6.4.1 *Ownership*

Suppose⁵ there are three people: Tek, Liz, and Sam. Tek and Liz are sailors. Sam is a shipbuilder and shiprepairer. Tek and Liz separately order ship services from Sam.

1. Tek wants a “new” ship made from used (weathered) boards.
2. Liz wants to replace the boards on her old ship (call it ship X) with new boards.

Sam has two warehouses where ship-building services are performed: warehouse A and warehouse B . There is a team of workers in each warehouse.

After taking payment from both Liz and Tek, Sam places Liz's boat in warehouse A . As boards are removed from Liz's ship, they are taken to warehouse B . At the end of a month, there are two ships:

1. In warehouse A , there is ship Y made of new boards.
2. In warehouse B , there is ship Z made of the old boards from ship X .

One night, however, there is a fire in warehouse A and the new ship is burned as well. This leaves only ship Z in warehouse B . Sam was operating his ship business uninsured and so he left town with Liz and Tek's money.

The morning after the fire, both Tek and Liz arrive at warehouse B , each hoping to take possession of the ship. Tek's argument was as follows. First,

⁵This practical example is from [124]. I've modified it by changing which dock burns down.

he paid Sam for a ship made of used boards and ship *Z* in warehouse *B* fits that description. Second, after looking over Liz's receipt, Tek also points out that Liz asked for a ship composed of new boards and ship *Z* does not fit that description. Ship *Z* must be Tek's. Third, a former employee of Sam testifies that the board was removed from Liz's ship, moved to a new warehouse, and was used to build ship *Z*.

In response, Liz acknowledges that ship *Z* is not the new ship she paid for. Nevertheless, she argues that ship *Z* belongs to her because it is the ship she gave to Sam to repair. She claims that ship *Z* just is ship *X*. Her ship was disassembled and then reassembled and now sits in warehouse *B*. To further her argument, she points to one particular board on ship *Z* with her initials carved into it.

Exercise 6.47

Who is the rightful owner of ship *Z*?

6.4.2 Resurrection of the Body

A core part of Christian faith involves the resurrection of the dead. Paul argued that if there is no resurrection of the dead, then Christ was never resurrected. And, if Christ was never resurrected, Christian faith is useless (1 Corinthians 15:12-14).

Christians also believe that human beings who have the spirit of Jesus within them will be resurrected when Jesus returns to earth. This is affirmed in several places throughout the Bible.⁶

What is less clear is in what form and method our resurrection will take. Some Christians (e.g., Catholics) believe that we will be provided with the physical bodies we had when we were living, although the bodies will now be imperishable.⁷ Romans 8:11 reads as follows:

“ And if the Spirit of him who raised Jesus from the dead is living in you, he who raised Christ from the dead will also give life to your mortal bodies because of his Spirit who lives in you.”

⁶“So will it be with the resurrection of the dead. The body that is sown is perishable, it is raised imperishable” (1 Corinthians 15:35-44). “He [Jesus Christ] will change our lowly body to conform with his glorified body by the power that enables him also to bring all things into subjection to himself” (Philippians 3:21).

⁷cf. 1 Corinthians 15:51-53

But, how will God accomplish this feat? There are many different proposals for how God will resurrect persons.⁸ In this section, we will consider the possibility that (1) we are intermittent objects and (2) death is analogous to disassembly, and (3) God's resurrection is analogous to reassembly. To resurrect us, God would collect our parts (atoms) at the moment we die (or right before we die) and then put them in the positions they had when we were alive. If God were to do this, the resulting person would be us. Since the resulting person would be us, God's final judgment would be applied to us rather than some clone of us.

Let's call this method of resurrection, "resurrection by reassembly".⁹

Definition 6.4: resurrection by reassembly

God is able to resurrect our bodies after they have died by reassembling us: taking our parts (atoms) and putting them in a spatial-chemical configuration that we had when we were alive.

There are four main problems with the resurrection by reassembly.

6.4.2.1 Double resurrection problem

Suppose Paul dies.^[143] What particles does God use to resurrect Paul? Presumably, God can use whatever particles God wants. Two examples:

1. God uses the particles of ten-year-old Paul.
2. God uses the particles of an elderly Paul.

God could even do both (1) and (2). But, this would result in two people who were both Paul existing at the same time. This is impossible. Therefore, God cannot resurrect by reassembly.¹⁰

⁸See [57] for an overview. One alternative is van Inwagen's "body-snatching model". [143, p. 121] van Inwagen writes "Perhaps at the moment of each man's death, God removes his corpse and replaces it with a simulacrum which is what is burned or rots. Or perhaps God is not quite so wholesale as this: perhaps He removes for "safekeeping" only the "core person"---the brain and the nervous system---or even some special part of it."

⁹As Hudson[57, p. 265] defines it "[t]he parts (mereological simples, if you have them) that composed the person at his or her death are tracked down, recalled, and reconfigured so that the life of the person they then composed may resume." Yang and Davis[155, p. 213] put it this way: "God resurrects the same body (or the same person for some materialists) by gathering the matter that formerly composed the earthly body and reassembling it into its previous shape and configuration (or at least organizing it in a sufficiently close enough way)."

¹⁰For a response to this problem, see [86]

6.4.2.2 Cannibal problem

Next, let's consider the cannibal problem.^[143] Suppose a wicked person (Tek) wishes to avoid God's judgment by engaging in cannibalism. Tek eats Paul. In doing so, Tek's body is composed of particles of Tek and particles of Paul. Which particles go to which person in the reassembly?

Let's consider some less violent cases. Suppose animals are resurrected as well. Carnivores are partly composed of animals. Which particles go to which animal in the reassembly? Or suppose Tek donates his kidney to Liz. He dies on the operating table and Liz dies one year later. In the resurrection, who receives the kidney? Or suppose instead that Tek dies in a crash and his organs (kidneys, heart, corneas, liver, lungs, pancreas, intestines, skin, bone, etc.) are divvied up among a group of people. Again, which organs go to which person in the reassembly?

One solution to the above problem is that of *staggered resurrection*.¹¹ Suppose two persons *A* and *B* share particles *P* at the time of their death. Rather than resurrecting them simultaneously, *A* is resurrected entirely. God then gradually replaces *P* with new particles *P**. Once replacement is complete, God then resurrects person *B* entirely.

6.4.2.3 Annihilation problem

God is said to gather the particles that compose us and reassemble us from these particles. But, a wicked person might escape hell by destroying these particles. He could have his particles annihilated, e.g., electrons are collided with positrons producing pure energy. The question asked by the annihilation problem is this: how could God resurrect a person from their particles if there are no particles to be used to reassemble them?^[155]^[143]

6.4.2.4 Gap Problem

Suppose a manuscript authored by Augustine is destroyed in 457 by being burned. Now suppose a manuscript (qualitatively identical) to the one burned is said to have been recreated by God. The story goes as follows:^[143, p. 118]

“Augustine's manuscript consisted of a certain ‘parcel’ of matter upon which a certain form had been impressed. It ceased to exist when this parcel of matter was radically deformed. To recreate it, God needed only to collect the matter (in modern terms,

¹¹Mooney^[86, 278n10] notes that this solution was via Cody Gilmore and Joshua Spencer

the atoms) that once composed it and reimpres that form upon it (in modern terms, cause these atoms to stand to one another in the same spatial and chemical relationships they previously stood in)."

In short, we have the existence of an *intermittent object*: there is Augustine's manuscript in 457, there is no manuscript in 458 (but there are the parts), and there is one and the same manuscript from 457 in 459 (with the help of God). The manuscript from 459 also exists because of *reassembly*.

van Inwagen claims that the manuscript is not Augustine's manuscript but a clever forgery. The reassembled manuscript is qualitatively the same as Augustine's but it is not Augustine's. It is *God's manuscript* for it is he who put the ink to paper. Similarly, suppose a child builds a tower of blocks, the tower is knocked down, and the father rebuilds the tower with the same blocks. The new tower is not the child's but the father's. Analogously, the reassembled human being is not the being that once lived and then died. It is a clone or intrinsic duplicate.

Why is the person resurrected by reassembly not the original person?

According to van Inwagen, in order to be a human being, you must be a living thing. Part of being a living thing is that (1) you are "naturally-produced" and (2) the stages of your life must be the result the "operation of natural processes". The being that is created by God that is qualitatively identical to you (1) is divinely produced and (2) the stages of its life are not the result of the natural processes of the parts that compose it.¹²

More stridently, van Inwagen claims that the reassembled being "would not be a member of our species". They would not "have memories of any sort" even though "he---or it---would appear to have these features." [143, p. 119] In other words, according to van Inwagen, our reassembly would violate the principle of diachronic sortal identity because the recreated being would not be human.

¹²As Mooney[86, p. 281] clarifies this objection: "reassembly fails to meet certain necessary causal conditions on survival: a reassembled person is not causally connected in the right ways to the premortem person""

Part III

Freedom

I find myself in a world full of objects. I am no God, but unlike other objects in this world I am a *person*. There are other persons like me: (1) they are self-conscious, (2) they can reason, and (3) they can survive change.

As persons, we believe we have another distinguishing power. *We are free*. At least to some degree, we have some control over our future. Our choices and actions matter.

This power distinguishes us from the other objects in the world whose destinies are entirely out of their control. The fate of these objects are due to external factors or internal factors outside of their control.

The rocks of the mountain skip down the mountain not by any power in the rock itself but by some force or law outside of the rock. Whether the external laws are the will of some god or the operations of the physical universe, it is clear that the rock's falling was not due to any choice or will or decision found in the rock itself. The rock could not have acted in any other way than it did! The internal biological processes of plants and animals are automatic. They cannot control their beating hearts, and their actions are due to desires and instincts that are unalterable by their own hands.

These external and internal forces guide my behavior as well. My choices are constrained by gravity, social-conditioning, and physics. When I am tired, I sleep. When I am hungry, I eat. When I am sad, I'm less lively. Nevertheless, the situation in me and my compatriots in this cosmos is different. Our choices are not entirely constrained. My cosmic compatriots and I appear to have an ability not found in nearly any other object in the world that we know. We appear to be free. Me and my fellow humans are distinct because we are *not totally controlled*. We believe that some of our actions come from *within*, that we are the originators of some of our actions, and that while we did act in one way (made one choice), we could have acted differently. That is, we take ourselves to be beings whose actions are not wholly determined by some outside force (a puppet master, some god, or brute physical laws).

However, how far this glimmer of power extends is unclear and perhaps it is an illusory light. For perhaps all of our action is ruled by internal and external forces. I watch people react in an instinctively rage upon hearing bad news. I observe politicians lure masses of people away from their personal and collective self-interest with misrepresentation and false promises. I see

people duped into quack cures, get-rich-quick schemes, and buying things they don't want because some slick marketeer has made the cure or product so appetizing that people cannot help but hand over their money. In these and countless other cases, I see beings potentially capable of free will failing to (or being unable to) exercise that power.

In this unit, we consider the claim that we are free more thoroughly by exploring two questions:

1. Do persons have free will?
2. If persons have free will, then what is this ability?

However, before we address these questions, let's consider a preliminary question: is free will important?

Exercise 7.48

Do you think you have the power of free will?

7.1 IS FREE WILL IMPORTANT?

People have various abilities. Liz can play piano. Renna can draw mountain scenes. Tek can speak French. None of these abilities are likely to be of a *general* interest. A large group of people may find Liz's ability to play piano interesting as they may want to learn themselves or wish to hear her play. However, for a larger group of people, this ability is of no concern. What is of interest to them is survival, with developing other abilities, or with other activities relating to human flourishing and titillation.

What about free will? Liz, Renna, and Tek may all claim to be free. They believe they can act upon their desires, control what they do, or act in a way other than how they did. Does this have any general interest? Let's consider two reasons why free will is of general interest by considering two ways it is important broadly speaking.

identity First, free will is important for understanding who we are (our identity).

We believe our actions are not fully driven by our instincts (biology), the laws of nature (physics), and not fully manipulated by others (e.g., advertisements). Our belief that we have a power that allows us to *control* our future is a core belief, essential to how we understand ourselves and distinguish ourselves from other people and objects.

Rocks and buildings are mere objects: they cannot even act. The actions of machines and animals are preprogrammed or determined by their instincts.

They cannot exert any control over how they act nor could they have acted in any other way than how they did act. One way we are different from mere objects, machines, or animals is that we can some control over how we act and how the world will ultimately look (at least in a limited way). As Kane[65, p. 285] puts it “we believe we have free will when we view ourselves as agents capable of influencing the world in various ways.”

The ability to act freely can even be used to define what it means to be a person. In addition to the physical (having a brain) and cognitive (psychological) approaches to defining personhood, another approach is functional in nature. These approaches contend that an object is a person if and only if they have the capacity to exercise their free will.¹

Definition 7.1: free will theory of personhood

Object O is a person if and only if O is capable of free will.

In sum, insofar as we care about understanding who we are, we have reason to care about (1) whether we are free, and, if we are, (2) what it means to be free.

Exercise 7.49

One reason free will is taken to be important is because human beings are the only beings that have the power of free will. Is this true? That is, are there other beings that have the power of free will? If so, how (if at all) is our relationship with these beings different than with beings that are not free?

Second, freedom appears important for determining whether you deserve responsibility praise or blame. Suppose we wish to blame Liz for hurting Tek. This assignment of blame assumes that not only that (1) Liz is the person who hurt Tek but also (2) that the hurting of Tek by Liz was done freely (under her control) or Liz freely did some action that foreseeably resulted in A . Let's draw out this consequence in the following argument:

- P1: If S is responsible for action A (deserving of praise or blame), then S 's action A is a free action.
- P2: S is responsible for action A .
- C: Therefore, S 's action A is a free action.

In addition, if we assert that Tek did not freely do A , then we can conclude that Tek is not responsible for A .

- P1: If S is responsible for action A , then S 's action A is a free action.

- P2: *S*'s action *A* is not a free action.
- C: Therefore, *S* is *not* responsible for action *A*,

Let's consider some examples. Suppose Tek and Jon are brothers. While Tek is sleeping, Jon picks up Tek's hand and hits Jon in the face. Jon then tells Tek's parents "Tek hit me in the face." In considering whether or not to scold Tek for hitting Jon, Tek's parents will consider whether Tek's action was free. If Tek is responsible for hitting Jon, then he did so freely. But, since Tek's action was not free, his parents should conclude that Tek is not responsible.

Here is a second example. Suppose Tek is secretly given an extremely potent drug that causes him to go berserk. Suppose also that it would be unreasonable to expect Tek (at least in this case) to take the usual precautions to prevent himself being drugged. For example, suppose Tek works in an accounting office and the drug was slipped into the coffee pot in the office. Now suppose that Tek, in his rampage, destroys a mailbox. While we might require Tek to pay for the damages, it is unlikely that anyone would think Tek should be blamed for the rampage. Why isn't Tek responsible? Again, what would explain our intuitions is that Tek cannot be said to be reasonable for his actions unless he was free. But, Tek's actions were not free and so he is not responsible.

One final example. Let's compare two cases, both involving a situation where Tek kills Liz. In the first case, suppose that Tek is driving his new car. As Tek approaches a red light, he sees Liz and the red light. Tek knows he is supposed to stop at red lights and knows that if he does not hit the brakes, his car will collide with Liz. However, rather than stopping at the light, Tek chooses not to hit the brakes. The car collides with Liz and she dies instantly. When the police arrive, Tek says that the brakes did not work. Secretly, he says "yes, I could have stopped but I did not want to. Instead, I wanted to hit Liz." The police do not believe him, he is carted off to jail, found guilty of vehicular homicide, and is sentenced to 7 years in prison (the maximum punishment allowed).

In the second case, suppose that Tek is again driving a new car. Again, as Tek approaches a red light, sees Liz and the red light, and knows he is supposed to stop at red lights and knows that if he does not hit the brakes, his car will collide with Liz. In this case, Tek presses the brakes but the car does not stop. Instead, there is a once-in-a-trillion malfunction with the car's brakes. Similar to the first case, Tek's car collides with Liz and Liz dies instantly. When the police arrive, Tek contends that his brakes did not work. Again, the police do not believe him, he is carted off to jail, found guilty of vehicular homicide, and is sentenced to 7 years in prison (the maximum punishment

allowed).

With the above two cases in mind, let's consider whether Tek is *responsible* for Liz's death. Specifically, let's consider whether Tek is *deserving of blame*. People believe tend to believe that Tek is not responsible for Liz's death in the second case. Why is that the case? Again, for the same reason as the prior two examples. In the second case, Tek is not responsible for Liz's death in the second case since he could not have stopped the car. He was not free to act in a way other than the way in which he did.

If these types of examples generalize, then, insofar as we care about assigning responsibility (blame and praise), we have a clear reason why caring about free will is important. The reason is that any assignment of responsibility to a person for an act entails that the action by that person was a free act. And, if we can determine that a person's action is not free, then we can also determine that that person is not responsible for the action.

Exercise 7.50

1. In this section, we considered a pair of cases involving Tek and Liz. These cases illustrated that the assignment of responsibility to a person (in particular blame and punishment) requires some level of free action on the part of that person. Responsibility, however, need not only involve blame but also praise. Create a similar example (using a pair of cases) where it only makes sense to say that person is responsible for an act and deserving of praise for that act if that action was freely performed by that person.
2. In this section, it was said that there are at least two reasons why freedom is important. The first is that concept of freedom is important for understanding who we are as persons. The second is that it is important for assigning responsibility (blame and praise). van Inwagen[139, p. 19] contends that “anyone who rejected free will could not consistently deliberate about future courses of action.” The reason is that because whenever we deliberate about a future course of action, we believe that these are actions that we are able to do. Using an example, articulate how it would be irrational for someone with no free will to deliberate about future courses of action.

7.2 DEFINING FREE WILL

People believe they have, unlike many of the objects that exist in this world, a special ability. This special ability is free will. But what is this ability? In

this section, three different conceptions of free will are considered:

1. the ability to act on your desires without external influence
2. the ability to control your desires
3. the ability to act in a way other than the way you did

Let's consider each of these abilities in the following subsections.

7.2.1 *Acting on your desires*

Suppose you are in a store and you see many different types of cereal. There is Cookie-Crisp, Banana Frosted Flakes, and Boo-Berry.



You ultimately decide you want Banana Frosted Flakes (BFFs) and you say to yourself "I want to put BFFs" in my cart.



You at t_1

Next, you put a box of BFFs in your cart. Notice that no one prevented you from putting BFFs in your cart. No one leaped in your way, pushing you to the ground, and taking all of the boxes of BFFs. There was no electrical force field that zapped you as you reached for the box. The almighty will of some god did not stop your arm from grabbing the box. In short, you desired BFFs and then you were able to act on this desire without any external interference.

In contrast, suppose you are standing at the base of Mt. Everest and you want to jump over it. You try to jump over Mt. Everest, but you fail. In fact, no matter how hard you try, you cannot leap over Everest because there is something keeping you from satisfying your desire: gravity, limitations of the human body, etc. So, while you may be free to *try* to jump over Everest, you are not free to jump over Everest since you cannot satisfy this desire.

What we have done thus far is to define and illustrate a certain kind of ability. Perhaps this ability is the same ability as having free will.²

Definition 7.2: Free action - acting on desires

Object O is free if and only if O can satisfy O 's desires without interference.

In other words, an object O has the power to freely act provided the following is true at least one point in their existence: O desires to do action X and there is nothing preventing O from doing X .

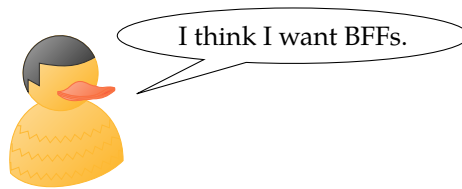
Exercise 7.51

Create two scenarios. First, create a scenario where a person has the ability to act based on their desires. Second, create a scenario where a person does not have the ability to act based on their desires. Is it right to say that the person is free in the first scenario but not free in the second scenario?

7.2.2 Controlling your desires

Let's reconsider your earlier decision to purchase Banana Frosted Flakes (BFFs). In that scenario, you were able to act on your desire to put BFFs in your cart. But why did you desire BFFs rather than Cookie-Crisp or Boo-Berry?

Suppose that in the weeks leading up to choosing BFFs, you see several advertisements for BFFs on social media. These advertisements show people eating BFFs and smiling happily. They make the contents of the BFFs box look very attractive. In addition, you talk to a friend who said that BFFs are the best cereal ever. They say that not only was the cereal delicious but they guarantee you will love it. Third, the company that makes BFFs found that by releasing the smell of bananas through a grocery's ventilation system, it made people more likely to purchase the cereal (and you are currently in a store where the banana scent is being passed through the ventilation system). Finally, as you entered the store, a shady-looking person said that if you don't come out of the store with BFFs, they will harm both you and your family.



With all of these conditions, we can say (1) you desire to put BFFs in your cart, (2) you are able to act on this desire, but (3) your desire for BFFs seems determined by a collection of external forces. In short, your desire for BFFs does not seem freely formed. Your action of selecting BFFs was not due to a power that you control but instead the result of external forces.

Let's consider another example. Suppose a person such that every desire they have is determined by some external force (e.g., a puppet-master, a collection of advertisers, politicians, a god). This person has the ability to desire certain things and they have, in certain circumstances, the ability to act on those desires, but they have no control over *what* they desire. All of their desires are determined from without. Intuitively, we would say that this person has *the freedom to act on their desires* but is not truly free since they cannot formulate their own desires. This individual is not the source of their desires. If life is moving through a maze, then the path they take is determined by someone or something else. Whenever they come to a crossroad, whether they choose to go left or right is determined by someone (or something) else.

With the above in mind, let's articulate another kind of ability.

(1) the ability to act on our desires but also (2) the ability to control our desires (to some degree). Let's call this type of freedom *freedom of the will* (this type of freedom is sometimes called "deep freedom" or "second-order freedom").

Definition 7.3: Free - ability to determine your desires

Object O is free if and only if O (1) can satisfy O 's desires without external interference (freedom to act on their desires) and (2) exert some measure of control over what they desire (freedom to determine their desires).

Exercise 7.52

1. Create an example of a person whose can act on their desires but whose desires are fixed by some outside force.
2. What are some examples of entities that attempt to not only control

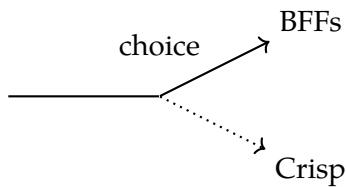
how people act but also control how what they desire?

7.2.3 Ability to do otherwise

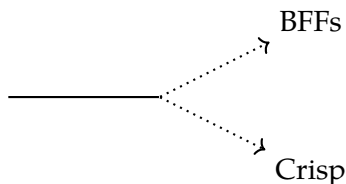
Third, let's consider one more ability. Again, consider your action of purchasing Banana Frosted Flakes (BFFs).



While you put BFFs in your cart, you might say one of two things. First, you might say that while I did choose BFFs, I could have acted in a way other than how I did act. I could have chosen otherwise. I could have chosen Cookie Crisp cereal even though I didn't.



On the other hand, you might say the following. Suppose I am in the situation where I am forced to choose between either BFFs or Cookie Crisps (only one), I have the ability (the power) to choose either one. That is, right now I can choose one or the other.



In the first case, the ability to do otherwise is clarified as an ability to have chosen differently than one did.

Definition 7.4: free - ability to have chosen otherwise

Object O is free if and only if O does A , O could have chosen otherwise (not- A).

In the second case, the ability to do otherwise is clarified as an ability to choose either option from two incompatible options.

Definition 7.5: free - ability to choose between two options

Object O is free if and only if O (now) is faced with two mutually incompatible actions $A1$ and $A2$ and O can choose either of the two options.

While we will (at least temporarily) treat both of these abilities as equivalent), let's consider an illustration of each.

Suppose that in high school, Tek breaks up with his long-term girlfriend. Tek then leaves for college and then he falls out of contact with her. Years later he sees her in a store and wonders what his life would be like if they did not break up. He says to himself, "while we broke up in high school, I could have decided to stay with her (even if only temporarily)." In making this claim, he is saying that he has the ability to have acted differently than he did.

Let's consider an example of the latter ability, where we now have the ability to choose between two, incompatible options:

"A man has free will if he is often in positions like these: he must now speak or now be silent, and he *can* now speak and *can* now remain silent; he must attempt to rescue a drowning child or else go for help, and he is *able* to attempt to rescue the child and *able* to go for help; he must now resign his chairmanship or else lie to the members, and he has it within his power to resign and he has it within his power to lie." [139, p. 8]

This ability to do otherwise suggests that if we are free, our world is, theoretically speaking, a *garden of forking paths* (see [Figure 7.1](#)). That is, while you may have chosen a path where you eat BFFs and, due to this choice and the enjoyment you reap from eating BFFs, spend a long life of eating BFFs, you could have taken a different path. That is, you could have chosen Boo-Berry cereal and decided that you would never eat cereal again.

For some philosophers, the existence of the garden of forking paths is important not only for understanding the world but also understanding what

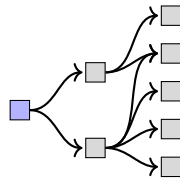


Figure 7.1: Garden of forking paths

it means to be a person. For example, Kane[65, p. 286] says that the existence of free will (and the garden of forking paths which is implied by it) is an essential part of being a person.

Exercise 7.53

Pick a famous person. Now take an important action in their life where they could have acted differently. Draw out how differently their life would have gone if they acted differently than they did.

7.2.3.1 Additional clarifications

Notice then that under both of these senses, “freedom” is defined in what a person *can* or is *able* to do. In the former sense, you are free if you *could have* acted differently. In the latter sense, you are free if you *can* (right now) do either one of two different acts. In order to get a clearer sense of what we mean when we say a person *can* do (in a way that is relevant for freedom), let's make some additional clarifications.

First, it is not the concept of moral, or legal permissibility. Suppose Tek tells his employer “you can't fire me”, meaning that they cannot legally fire him. But his employer may fire him anyway. Or suppose that after Liz hears Sal tell her that Sal has lied to his parent, she may say “you can't lie to your parents like that.” In this scenario, Liz does not mean that Sal is not free to lie to his parents. Instead, what she means is that he *shouldn't* lie to his parents. Sal is capable of lying to his parents even though he should not from a moral or prudential point of view. In short, *S* may perform free action *A* but *A* may be neither morally permitted nor legally permitted.

Second, an act may be *physically possible* but not free. van Inwagen's example is that we can imagine a person in a locked cell. While it may be physically possible for the door or the lock to open, allowing the person to exit, the person does not have the ability to exit. As van Inwagen puts it “it is physically possible that I shall leave the room. But it does not follow that in any relevant sense I *can* leave the room.”[139, p. 9] In short, *A* may be a physically possible act but *S* may not be free to do *A*.

Here some more examples illustrating how an act may be physically possible but a person is not free to do that action. With performance-enhancing drugs and genetic-engineering, it may be physically possible for someone to run a 3-minute mile. Unfortunately, Tek is not free to run a 3-minute mile right now. If we take him to a track and have him run a mile, he will finish short of the mark every time.

Third, free action is not simply a causal capacity. It would be wrong to say “because I can do X, I am free to do X.” Some causal capacities are invariant dispositions: reactions over which we have no control. For example, if Tek eats a hamburger, his body will digest the food. While it is correct to say “Tek can digest meat”, it would be wrong to say “Tek is free to digest meat”. The latter is mistaken because Tek exerts no control over the digestion. Free will then is a power, but not a reactive causal power.

Free action is better understood as a causal *ability*. Suppose Tek is running toward a hurdle. Tek has the power to leap over the hurdle. When saying that “Tek can leap over the hurdle”, we do not mean to say that this is a reactive process outside of Tek's control. For example, Tek may decide to stop short.

Here is an example from van Inwagen[139, pp. 10–11]. Consider the difference between Tek's being able to *understand French* and Tek's being able to *speak French*. For van Inwagen, the former is a causal capacity, the latter is an ability.

“For a man to have the capacity to understand French is for him to be such that if he were placed in certain circumstances, which wouldn't be very hard to delimit, and if he were to hear French spoken, then willy-nilly, he would understand what was being said. But if a man can *speak* French, it certainly does not follow that there are any circumstances in which he would, willy-nilly, speak French.”[139, pp. 10–11]

Fourth, there is a distinction between *having an ability* and *having the power to exercise that ability*. Suppose I learned to ride a bicycle. If you ask me if I can ride a bicycle, I say “yes, I can ride a bicycle because I have acquired the ability to ride a bicycle.” But, remembering that my leg is broken, I might also add “even though I can ride a bike, I can't ride a bike.” In saying this, I am saying that I have the ability to do *X* but lack the power to exercise the ability to do *X*. When talking about whether or not we are free to do act *A*, we are talking about whether you have the power to exercise one's ability to *A* not whether you have the mere ability to do *A*. Another example from van Inwagen: suppose a person can speak French (they have the ability)

but they are gagged (they lack the power to exercise that ability). In such a scenario, we would say that they are not free to speak French since they cannot exercise their power.

Exercise 7.54

Let's illustrate the above clarifications by thinking of examples of our own. Create an example for each scenario. Note that in each example, you are creating a scenario where a person is not free.

1. An action X is physically possible, but S cannot do X .
2. S has the power to do X but this power is a purely reactive power.
3. S can do X (has the ability) but S cannot exercise the ability to do X .

Exercise 7.55

Some people say 'anyone *can* become rich. You only need to work hard.' Others contend that the sense in which people *can* become rich is the same sense in which a person *can* speak French when they are gagged. Do you agree?

7.2.4 *Free Will Thesis*

Let's conclude this section by simply stating a thesis that will be relevant for later discussions concerning free will. This is the "free will thesis".

Definition 7.6: free will thesis

There is at least one humanlike creature that is free: for at least one action A , they can do A or not- A .

In stating this thesis, we will be ambiguous about whether "free will" refers to surface freedom or deep freedom.

Exercise 7.56

Is the free will thesis true or false? Explain your reasoning.

7.3 BASIC POSITIONS ABOUT FREE WILL

Now that we have stated the importance of freedom and several different conceptions of freedom, let's clarify some philosophical problems involving freedom. The problem of freedom can be phrased in several different ways.

The way we will present it here starts from the definitions of two claims about the world: determinism and indeterminism.

7.3.1 Determinism and indeterminism

Let's start with a definition of determinism.

Definition 7.7: determinism

Determinism is the contingent and empirical claim that for every event e at a time t_2 for which there is an earlier state of the world, the total state of the world w at that earlier time t_1 along with the all the laws L entail every future event at a later time t_2 . [146, p. 304].

In short, determinism is the view that the world and its laws “determine” future events.

$$(w, t_1) + \text{Laws} \longrightarrow e$$

Let's elaborate on a few parts of the above definition. First, determinism is a contingent claim. A proposition P is *contingent* if and only if it is possible that P is true and it is possible that P is false. In other words, a proposition is contingent iff it is not impossible (since it is possible that it is true) and it is not necessary (since it is possible that it is false).

Definition 7.8: contingency

A proposition P is *contingent* if and only if it is possible that P is true and it is possible that P is false.

Most of the propositions we utter on a daily basis are contingent. Suppose Tek says 'I ate a sandwich'. This proposition is contingent since it is true (so it is possible that it is true) and it is possible that it is false (since Tek might have had a soup instead). Propositions that are not contingent include propositions that express contradictions, e.g., Tek is ten-feet tall and one-foot tall, and those that express necessary truths, e.g., $2+2=4$ or either the restaurant is closed or it is not.

If someone asserts that a proposition is *contingently true*, they are asserting that it is (1) contingent but also (2) that it is true in the actual world. So, when someone claims that determinism is true, they are saying that there are some possible worlds where it is true, there are some possible worlds where it is false, but that determinism is true in our world (the actual world). To illustrate, suppose there are two possible worlds (total ways that things can

be). Let's call them w_1 and w_2 and suppose we are in w_1 . Now suppose in w_1 that the total state of affairs at a time t_1 along with all the laws L entail any event at a time later than t_1 . Such a world is determined. In contrast, suppose it is possible that there is a world w_2 where this is not true: there are some events (call one e) for which there is an earlier state of the world but the laws along with the state of the world do not entail that event e . Such a world is not determined.



In saying determinism is contingent, we are saying that (1) there are some determined worlds, (2) there are some worlds that are not determined, and (3) the world we live in is determined.

Next, in calling determinism an *empirical claim*, what is meant is that the truth (or falsity) of determinism is discovered by experience. That is, the best or ultimate science (physics, biology, psychology, etc.) will tell us whether events are entailed by the total state of the world along with the laws of nature.

Third, it is common to say that in a deterministic world, all events are determined. This is a shorthand way of saying events are entailed by the prior total state of affairs and all of the laws. There are at least two ways to think about what it means for an event to be a determined event

Definition 7.9: physically determined event

An event e is determined iff the prior state of the world plus the laws of nature are entail e .

On this definition of a determined event, an event is determined provided the prior state of affairs and the physical laws of nature entail a single event e . An intuitive way of thinking about determined events is that they are predictable with 100% accuracy by an individual with perfect reasoning powers and enough information about past configuration of the world and the way the world works. Thus, assuming that an event e is determined, a being like God would be able to predict that e would occur using only information about the world before e and the laws of nature.

However, we might understand what it means for an event to be determined in a broader way. Intuitively, we think that an event e could come about, not merely because of the laws of nature, but also because it is determined by psychological laws, or societal conditions, or even God.

Definition 7.10: causal determinism

The position that every event e (except for perhaps the first event) is a determined event.

In contrast to the garden of forking paths (see [Figure 7.1](#)), determinism gives us a different model (see [Figure 7.2](#)). If every event is determined, at least at first glance, the agent could not have acted otherwise.

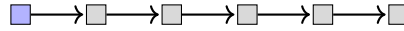


Figure 7.2: Deterministic picture of the universe

Exercise 7.57

What sorts of things limit our freedom? What types of limitations are there? How do things, people, groups determine how we think, how we act, what we desire?

In contrast to determinism is indeterminism. This is the negation of determinism.

Definition 7.11: indeterminism

Indeterminism is the contingent and empirical claim that determinism is not true. There exists an event e for which there is an earlier state of the world such that the total state of the world w at that earlier time t_1 along with the all the laws L do *not* entail every future event at a later time t_2 . [[146](#), p. 304].

There are at least two different ways of thinking about indeterminism. The first is that all of the laws of the world have limited scope (they do not have total coverage). For example, we can imagine that all of the laws of physics, chemistry, biology, etc. do not cover certain aspects of reality like the workings of the human mind, the activity of an immaterial soul, or even certain kinds of physical events that occur in the world (hauntings, ghosts). More typically, some claim that physical laws are probabilistic. On such an account, it is not the case that the state of the world along with the laws *entail* an event but rather than the state of the world and the laws *on average* will lead to the event. In other words, there is an element of randomness (or chance) in the world.

Exercise 7.58

Based on your experience, do you believe our world is deterministic or indeterministic?

7.3.2 Incompatibilism, Compatibilism, and Impossibilism

Now that we have formulated the free will thesis and two ways of viewing the world, let's articulate the problem of free will. The problem of free will emerges when we consider that our view of the world (determinism or indeterminism) appears to be inconsistent with our view of whether there are any humanlike creature that have free will (the free will thesis). To begin, let's consider two different sets (collections) of possible worlds: deterministic worlds and indeterministic worlds.

deterministic

indeterministic

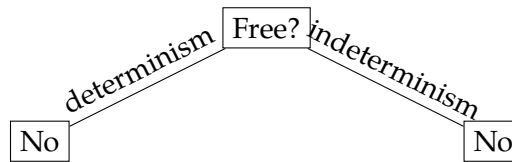
Next, let's say something is "metaphysically possible" if and only if it can be true in at least one of the worlds in these sets. For example, if it is metaphysically possible for Tek to jump five feet in the air, then Tek jumps five feet in the air in at least one possible world (either a deterministic world or an indeterministic world, or both). In contrast, something is "metaphysically impossible" if and only if it is not true in any of the worlds in these sets. Take the claim that "Tek is 6'0 and 5'0". While it may be metaphysically possible for Tek to be 6'0 and it may be metaphysically possible for Tek to be 5'0, there is no world (either deterministic or indeterministic) where Tek is simultaneously these two different heights. And so, "Tek is 6'0 and 5'0" is metaphysically impossible.

Next, let's consider the theory called "impossibilism".

Definition 7.12: impossibilism

Free will is metaphysically impossible for human-like creatures.

According to the impossibilist, the free will thesis is necessarily false. If the world is deterministic, there are no humanlike creatures that are free. If the world is indeterministic, there are no humanlike creatures that are free. Regardless of what the world is like (deterministic or indeterministic), no humanlike creatures are free.



It is important to note that the impossibilism is not a claim about whether determinism or indeterminism is true. The ultimate science may find our world is deterministic or indeterministic. Either way, there will be no humanlike creatures who have free will. We can represent this using the following table:

	Determinism	Indeterminism
Free will	X	X

Table 7.1: Impossibilism

Second, in contrast to impossibilism is incompatibilism.

Definition 7.13: incompatibilism

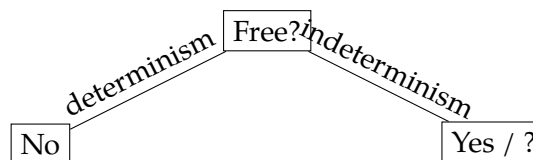
The theory that determinism and the free will thesis are inconsistent.

The incompatibilist contends that there are no deterministic possible worlds where humanlike creatures are free (have the ability to do anything other than what they do). Alternatively put, the incompatibilist says “it is necessarily the case that if determinism is true, then the free will thesis is false.”

The incompatibilist can say at least one of two things about indeterministic worlds:

1. the incompatibilist asserts that the free will thesis is true in indeterministic worlds
2. the incompatibilist is agnostic about whether the free will thesis is true in indeterministic worlds.

While typically incompatibilists tend to assert that free will is only possible in an indeterministic world, they may be agnostic about whether this is the case.



It is important to note that incompatibilism is a thesis about what is possible. An incompatibilist may believe any of the following:

1. our world is determined and so we are not free,
2. our world is indeterministic and we are free
3. our world is indeterministic but in this indeterministic world, we are not free (although in some other possible world we are free).

	Determinism	Indeterminism
Free will	X	✓

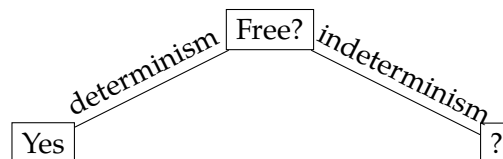
Table 7.2: Incompatibilism

A third option is compatibilism.

Definition 7.14: compatibilism

Free will is metaphysically possible for human-like creatures even if the world is deterministic.

Compatibilists believe that the free will thesis and determinism can both be true. That is, there are some possible worlds where the world is deterministic and there are humanlike creatures with free will. However, the compatibilist is agnostic about whether there are possible worlds that are indeterministic and humanlike creatures have free will.



To illustrate the fact that compatibilists believe that the world may be deterministic and there is free will, we can use a ✓ and to indicate that they are agnostic about whether there are indeterministic worlds where we have free will, we can use a question mark.

	Determinism	Indeterminism
Free will	✓	?

Table 7.3: Incompatibilism

Exercise 7.59

1. Can someone who accepts impossibilism also accept that determinism is true?
2. Can someone who accepts incompatibilism accept that determinism is true? Can they alternatively believe that the world is indeterministic?
3. Rank each theory from best to worst: impossibilism, incompatibil-

ism, or compatibilism.

4. Pick one of the three theories (impossibilism, incompatibilism, or compatibilism) and give an argument in support of it.

In the previous chapter, the problem of free will was introduced. This problem is that it is not immediately obvious whether the following two claims are consistent: (1) there is at least one humanlike creature that has free will (free will thesis) and (2) the world is deterministic or indeterministic.

In the previous chapter, three views on free will were outlined: (1) impossibilism, (2) incompatibilism, and (3) compatibilism. In this chapter, we focus on arguments in support of impossibilism.

8.1 IMPOSSIBILISM

There are three different types of impossibilism:

1. Strong inevitabilists
2. Weak inevitabilists
3. Contemporary fatalists

Now let's articulate each of these views by (1) defining the theory and (2) illustrating the theory with a story. We will start with *strong inevitabilism*.

8.1.1 *Strong inevitabilism*

In this section, we will state the first type of impossibilism: strong inevitabilism. We will start by stating what it means for an action to be strongly inevitable, then we'll state the theory of strong inevitabilism, and finally, we'll consider some objections to strong inevitabilism.

8.1.1.1 *Strongly inevitable actions*

Let's begin with a definition of strong inevitabilism and what it means for an event E to be strongly inevitable.

Definition 8.1: strongly inevitable

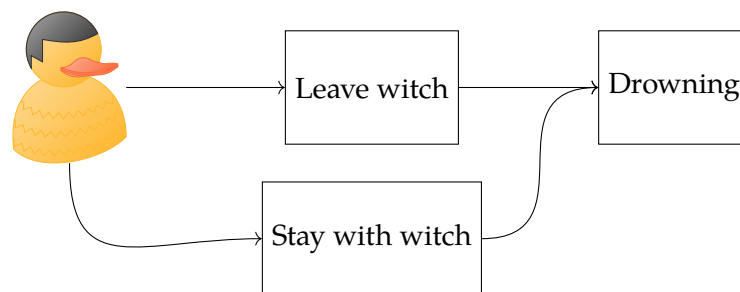
An event E is strongly inevitable for S if and only if E will happen no matter what. That is, if an event E is strongly inevitable for S , then there is nothing anyone *can* do to prevent E from occurring.

In broad terms, surely some events are strongly inevitable. Whether I will ultimately die is strongly inevitable *for me* since while I may do a variety of things to prolong my life, there is seemingly nothing I can do to prevent my

death. If I eat healthy, I will die. If I eat poorly, I will die. If I sometimes eat healthy and sometimes eat poorly, I will die. Since I will die no matter what I do, the event of me dying is strongly inevitable for me.

In addition, there are some events that are likely strongly inevitable *for everyone*. Everyone will die, so a person's individual death is strongly inevitable for them. *Remote cosmic events* are also strongly inevitable. Suppose two space rocks will collide in some far away part of the universe. There is nothing that you or I or anyone can do to prevent this event. Therefore, that event is strongly inevitable.

Another somewhat fantastic example.[139, p. 23] Suppose you enter a witch's home and the witch says you will drown within the next twenty-four hours. You decide to leave the witch's home to go to the highest mountain. But, as you are leaving, bandits grab you and put you to death by drowning you. If you decided to stay in the witch's home, they would have entered and drown you. If you tried to hide in the witch's home, they would have found you and drown you. No matter what you did, it was strongly inevitable that you would drown.



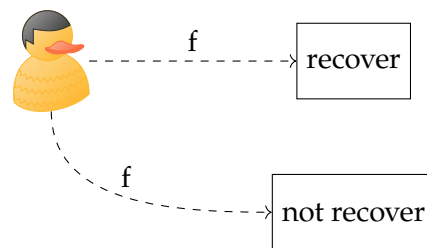
Exercise 8.60

Create a story where a person is “fated” (in the strong inevitable sense) to do some action.

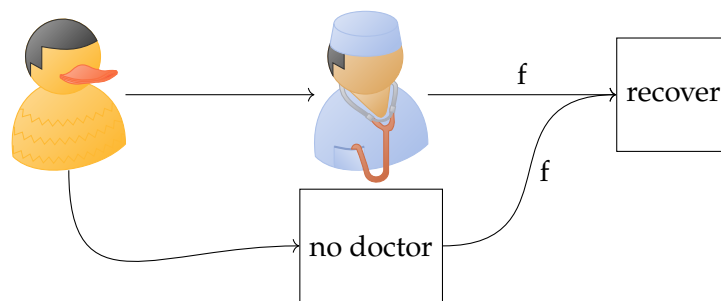
Let's consider another classic example. Consider the so-called “lazy argument” (Aristotle's *De Interpretatione*, chapter 9, Cicero, *On Fate* 28-9, Origen, *Against Celsus* II.20, Pseudo-Plutarch, *On Fate* 574e, Seneca *Natural Questions* II.38.3). In this argument, it is assumed that it is either *strongly inevitable* that some event *E* will occur or *strongly inevitable* that not-*E* will occur. The argument then contends that since either *E* or not-*E* are strongly inevitable, *what you do does not matter*. And so, since what you do does not matter to what happens, you can be idle or lazy. The argument goes as follows:

- P1: You are either fated to recover from the illness or fated to not recover from the illness (“fated” = “strong inevitability”).
- P2: If you are fated to recover from this illness, then it does not matter whether you consult a doctor. You will recover either way.
- P3: If you are not fated to recover from this illness, then it does not matter whether you do or do not consult a doctor. You will not recover either way.
- C: Therefore, it is pointless to contact a doctor.

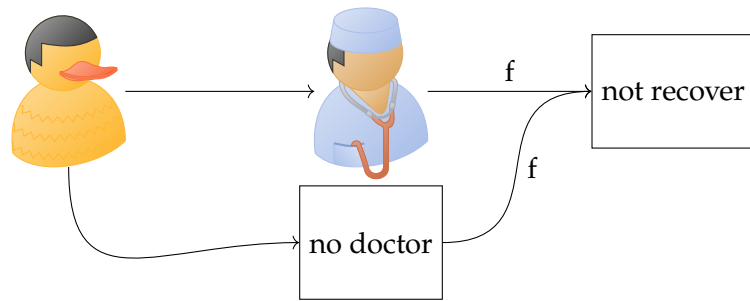
Let's unpack this argument with a diagram. P1 says that it is either fated (strongly inevitable) that you will recover or it is fated (strongly inevitable) that you will not recover. This premise is designed to cover the two possible options. It does not assert that anyone knows which of the two options is fated.



Let's consider the two possibilities. First, suppose you are fated to recover. If that is the case, then it does not matter whether you consult a doctor since you will recover either way. That is, under the recovery option, (1) if you contact a doctor, you will recover and (2) if you do not contact a doctor, you will not recover. This is because, if your recovery is strongly inevitable, your actions make no difference to the outcome.



A similar diagram can be created if it is fated that you will not recover. Since it is strongly inevitable that you will not recover, *your actions do not matter* (it does not matter what you do).



Given that you will either recover or not recover, and that in both scenarios it makes no difference what you do, it follows (according to the argument) that it is pointless to contact the doctor.

Exercise 8.61

The idle argument is supposed to show that you can be idle or lazy about any given action. Take a task that you don't want to do and construct your own version of the idle argument to justify your inaction around it.

8.1.1.2 Strong inevitabilism

The theory of strong inevitabilism asserts that all events are strongly inevitable. That is, take any future event E you wish. There is nothing that you or I or anyone can do to change that E is the case.

Definition 8.2: strong inevitabilism

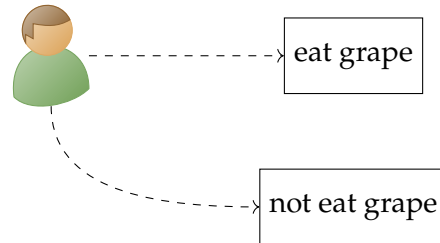
For every person S and for all future events E , E is strongly inevitable for S .

Let's illustrate strong inevitabilism. To do this, let's select some future event E : the future event of *Tek dying tomorrow by choking on a grape*. If strong inevitabilism is true for a subject S , then Tek will die tomorrow by choking on a grape and there is nothing anyone can do to prevent this event from occurring.

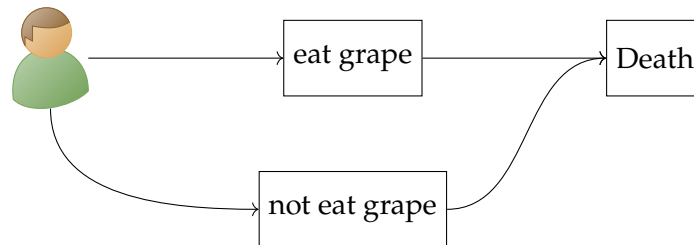


But isn't this absurd? Surely there is *something* that someone could do to prevent Tek from choking on a grape and dying. Surely, if Tek closed himself up in a room and avoided all food for the entire day, he would not die from

eating a grape. But consider that strong inevitabilism asserts that *all* actions are strongly inevitable. So, let's consider Tek's choice between eating a grape and not eating a grape.



If all actions are strongly inevitable, then Tek's decision to eat a grape is strongly inevitable. And, if that is the case, then Tek will decide to eat a grape, no matter what. There is nothing he or anyone can do to prevent this actions. There is no way that Tek could make another choice.



8.1.1.3 Problems with strong inevitabilism

There are several problems with strong inevitabilism.

First, some of the evidence used to support strong inevitabilism comes from examples of events that are strongly inevitable. People reason that because such-and-such events are strongly inevitable events, it follows that all events are strongly inevitable.

- P1: Events $E_1, E_2, \dots E_n$ are strongly inevitable.
- C: Therefore, all events are strongly inevitable.

This is a hasty generalization. Just because some events are strongly inevitable does not mean that all events are strongly inevitable.

Second, many of the examples of events that are strongly inevitable are only thought to be strongly inevitable because they are *vague*. To illustrate, consider two different statements:

1. Tek will die.

2. Tek will die tomorrow in a car crash (driving a blue Honda Civic) at precisely 5:02 PM EST.

Statement (1) is vague. It strikes many as strongly inevitable. Statement (2) is more precise. It is not intuitively true that it is strongly inevitable. We would need to be supplied significant evidence to believe there is nothing Tek (or anyone else) could do to prevent this event.

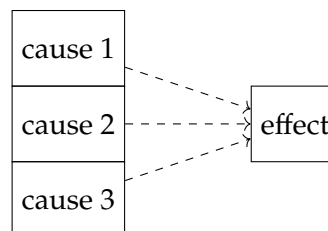
Let's consider another example. Suppose a witch tells me I will drown. If the event is strongly inevitable, it will happen no matter what I do. Consider two cases. First, I run to the top of a mountain and fall down a well and drown. Second, if I were to run to the sea, I would fall into the sea and drown. But now suppose that we ask the witch to specify what event is strongly inevitable with much more specificity. We ask her to state where, when, and how I will drown. Our illustration of strong inevitabilism now looks more implausible because if the witch specifies that I will drown in the sea at such-and-such time but I immediately proceed up away from that sea, it becomes increasingly implausible about how this event will occur.

Third, strong inevitabilism asserts that certain events will happen *no matter what*. It asserts that there is nothing anyone *can* do to prevent events from occurring. But this is *too strong* of a statement. It is too strong because the theory entails the following: *even if things were different, event E would result*. This is wrong because it severs the connection between cause and effect.

Let's illustrate this with an example. Suppose it is claimed that it is strongly inevitable that I will break a dish. My breaking of the dish can only come about given a certain set of causal conditions. Let's specify the minimum requirement as follows:

The minimum causal requirement for the effect of my breaking a dish is that I come into contact with the dish either directly or through a direct causal chain.

But, if strong inevitabilism is true, then this dish will break *no matter what the causal conditions*. It is saying, take any causal situation you wish, this causal situation will lead to the breaking of the dish.



That is, even if things were different, I would still break the dish. But let's consider how things might be different:

1. The day before I get rid of all my dishes and sit in silence in my room.
2. I am locked in a padded room with no dishes.
3. I don't exist.

But this violates the minimal causal requirement since in the above scenarios I do not come into contact with the dish (or set off a causal chain that leads to the dish breaking). Therefore, it is false that the dish will break *no matter what I do*.

8.1.2 Weak inevitabilism

Let's turn to weak inevitabilism. First, let's define what it means for an event E to be weakly inevitable.

Definition 8.3: weak inevitability

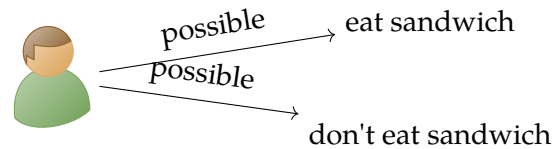
A future event E (an event that is going to happen) is weak inevitable if and only if

1. E is not strongly inevitable (whether E will happen *does* depend upon what S does). There are some events that if S acted differently, E would not occur.
2. in any scenario where S tries to prevent E from occurring, then S will always choose the wrong method for preventing E because S is ignorant of the right method for preventing E .
3. it is strongly inevitable that S is ignorant of the right way to prevent E .^a

^aVan Inwagen[139, p. 25] defines "weak inevitability" as follows: "a future event, or state of affairs, is *weakly inevitable for me* if both of the following conditions hold: (i) it is not strongly inevitable, and (ii) if I tried to take measures to prevent it, then I should the wrong measures out of ignorance, and it is strongly inevitable for me that I should be ignorant of the right measures."

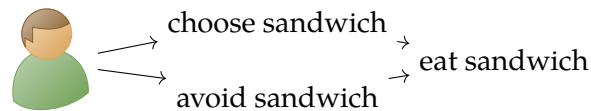
Let's clarify what it means for an event to be weakly inevitable.

First, if an event E is weakly inevitable, it is not strongly inevitable. This means that it is *possible* for not- E to occur. Suppose it is weakly inevitable that you will eat a sandwich. If this is the case, it is possible for you to not eat a sandwich.



This is an improvement over strong inevitabilism since it means that weakly inevitable events won't occur no matter what but only when the right causal conditions are in place.

Second, even though it is possible for you not to eat the sandwich, in any scenario where you would try to prevent E from occurring, you would fail to prevent E from occurring. So, if eating a sandwich is weakly inevitable and you were to try to avoid eating the sandwich, you would fail to avoid eating the sandwich.



One important clarification is that the weak inevitabilist is not saying that *only when you try to prevent some event does that event occur*. Rather, they are saying

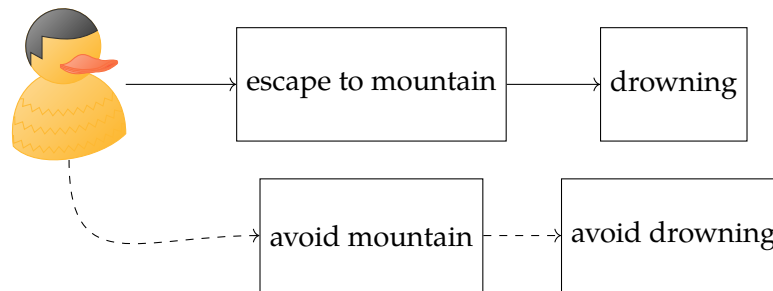
1. it is possible to prevent E from occurring
2. *if* you were to try to prevent E , then whatever method you would choose would be a method where you fail to prevent E .

What this means is that while it is possible that E could fail to occur, there is no possible scenario where *you* could prevent E from occurring.

This takes us to the third component of a weakly inevitable event. This component answers *why* you always fail to prevent E . If E is weakly inevitable, whenever you try to prevent E , you fail to prevent E because you are always ignorant of *how* to prevent E . That is, it is strongly inevitable that you are ignorant of the right way to prevent E . That is, you will never know how to prevent E no matter what.

Let's consider some examples of events that are weakly inevitable. First, let's consider a variation on an earlier example.[\[139, p. 22\]](#) Suppose a witch tells you that you will drown within twenty-four hours. The witch also writes exactly how this will occur on a piece of paper and then drops it into a boiling pot. The future event E then is *you drowning*. According to the weak inevitabilist, this event is not strongly inevitable. This means it is preventable.

After hearing the witch, you travel to the highest mountain. As you are ascending the mountain, you stumble into a well and drown. You could have avoided drowning if you had pursued a different action. For example, if you stayed with the witch and had lunch, you would have avoided drowning. But, you were ignorant of the right way to prevent your drowning. And, there is no way you could ever know how to prevent your drowning.

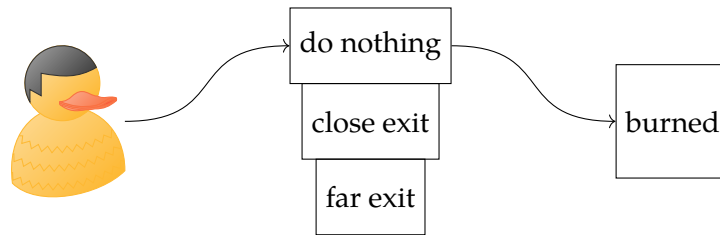


Notice that the drowning is not strongly inevitable since whether you drown *does depend* on your action. You could avoid drowning. The problem is that you never would choose an action that would result in your non-drowning because you were ignorant that you would drown on the mountain. Plus, since the act is weakly inevitable, there is no way you could have known that you would drown on the mountain.

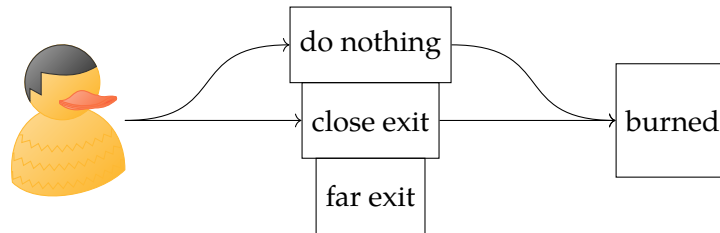
Here is another example from van Inwagen. Suppose you are in a building that is on fire and it is weakly inevitable that you will be burned in the fire. Since the event is not strongly inevitable, it is possible for you to avoid being burned. As you are standing in the burning building, you know that there are two exits: (1) One exit is close to you (the near exit) and (2) another exit on the other side of the building (the far exit). As you feel the heat from the flames and breathe in the smoke, you consider your options:

1. Do nothing: allow yourself to be burned to death.
2. Try to leave the building via the near exit.
3. Try to leave the building via the far exist.

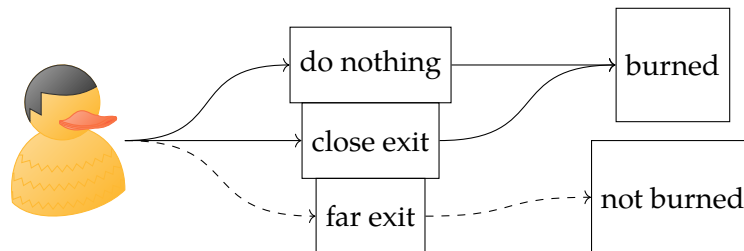
Since it is weakly inevitable that you will be burned, you could do nothing and let the event occur. If that is the case, then if you were to do nothing to prevent being burned, then you would be burned.



But, of course, you do not want this to occur. You want to prevent this weakly inevitable event. Since you want to avoid being burned, you decide that you either need to exit through the near exit or the far exit. You decide to exit through the near exit. In fact, in this scenario, it is the only option that makes sense. Unfortunately, this choice leads to you being burned as well.



At this point, it looks like the event of being burned is strongly inevitable since we were burned in the fire *no matter what choice we made*. However, you could have selected the far exit instead of the near exit. And, if you did, you would have avoided being burned. So, in this scenario, where your action was weakly inevitable, your choice *did matter*. If you would have acted differently, you would have avoided being burned.



One final point is that since the act is weakly inevitable, *you never would have selected the far exit*. You would never select this far exit because *you were ignorant that this way to avoid being burned* and it was strongly inevitable that you would be in this state of ignorant. There is no way for you to have knowledge that the far exit would prevent you from being burned.

Exercise 8.62

Create a scenario that is weakly inevitable. Make sure that you state (1) what event is weakly inevitable, (2) what would happen if the person does nothing (doesn't try to prevent the event from occurring), (3) what would happen if the person tries to prevent the event from occurring, and (4) how it would be *possible* for the person to prevent the event from occurring.

With the weak inevitability defined and illustrated, let's define *weak inevitabilism*.

Definition 8.4: weak inevitabilism

For every person S and for all future events E , E is weakly inevitable for S .

If every event in our lives is weakly inevitable, then while it is theoretically possible for us to prevent some of these events, we have no power to prevent these events from occurring. And if we have no power to prevent these actions from occurring, then freedom is impossible.

8.1.2.1 Objections to weak inevitabilism

Weak inevitabilism says that while it is possible for a weakly inevitable event E to not occur, there is no way for you to prevent E from happening because you will always be ignorant of how to prevent E . In other words, *it is impossible for me to know how to prevent E* . But this seems false. Suppose I am in a burning build with two exits. I might be naturally disposed to take the near exit (which will lead to my death), but it seems possible for me to know how to avoid being burned.

Here is just such a possibility. Suppose instead of writing poetry, I study architecture, fire safety, how to read smoke patterns, how to evaluate the location of fire given damage to a building, and whenever I enter a building, I consider the best way to escape if there is fire. Now instead of taking the close exit, I might be calm-headed enough to see that the fire appears most concentrated in front of me and so I should try to escape through the far exit.

Let's consider a more mundane scenario. Of course, the fact that events are weakly inevitable does not imply that we *know* which events are weakly inevitable. So, let's suppose I am writing my lecture notes. It is either weakly inevitable that I will do one of two actions:

1. stop writing my lecture notes
2. continue writing my lecture notes

Let's consider each case. Suppose (1) is true. If weak inevitabilism is true, then it is possible that I might continue writing my lecture notes but it is not possible *for me* to continue writing my lecture notes because I do not have the knowledge of how to continue. This is absurd. I do know how to continue: I just keep writing.

So, perhaps it is weakly inevitable that (2): that I continue writing my lecture notes. If that is the case, then it is not possible *for me* to stop writing my lecture notes because I do not have the knowledge of how to stop. This is absurd: I just get up from chair and walk outside.

Exercise 8.63

This question is inspired by the short story "The Skull" by Philip K. Dick. Suppose it is 2160 and a prisoner named "Congor" is asked by the government to go back in time (approximately 200 years) to kill The Founder of "the Movement". According to the Government, The Founder was a relatively unknown man who lived in Denver, Colorado in 1960. The Movement was a group that aimed to stop war by eliminating weapons, and eliminating weapons by stopping the production of technology. Their logic was relatively simple: "man was losing to his machinery and science", it was "getting away from him, pushing him to greater and greater wars. Down with society, they shouted. Down with factories and science [besides medicine]! A few more wars and there wouldn't be much left of the world." [37, p. 50]

It is known that during his time, the Founder only spoke once in public, and after speaking, he was arrested and executed. However, some of his disciples claimed to have seen him walking around after his death and so enthusiasm for The Movement grew on account of the belief that the Founder rose from the dead. While little was known about the identity of the Founder, the government was in possession of the Founder's bones. So, to aid in the assassination, Congor was given the skull of the Founder and told to (1) kill the Founder, (2) use the skull to confirm that he killed the Founder, and then (3) return to 2160 with both skulls so that the government could confirm the Founder was assassinated.

In exchange for assassinating the Founder, Congor would have his prison sentence commuted. With that in mind, Congor enters the time

machine and goes back in time to April 1961. As he's walking through the town, he notices several people look at him strangely. He makes his way to the local library and begins going through old newspapers to check for headlines about anyone who may have been arrested in town after spreading a message of non-violence. He discovers a headline that reads:

MAN ARRESTED FOR UNLICENSED DEMONSTRATION

An unidentified man who refused to give his name was picked up in Cooper Creek. It was said that the man was recently noticed in this area and had been watched continually. It was --

The date of the article was Dec 1st 1960. With this new information, Congor reentered the time machine and went back to Nov 1960.



After spending a few days in town to gather information, Congor had a disagreement with a local man who accused Congor trying to steal his girlfriend. The man would contact the local sheriff who would contact federal authorities. Congor would ultimately become the subject of a man-hunt and would need to flee into the woods. As authorities began to descend upon Congor's location, Congor noticed something about the skull: it matched the profile of his own head. Realizing that the man he was ordered to go back in time and assassinate was himself, he faced the authorities and gave a short speech about the importance of non-violence and the abuse of technology. Congor was then arrested and executed.

Consider the event *stopping The Movement*. Is this event strongly inevitable or weakly inevitable? Explain your reasoning.

8.1.3 Fatalism

So far, we have considered strong and weak inevitabilism. Next, let's consider fatalism.¹ Fatalism is the theory that it is a logical or conceptual truth that no one is able to act in a way other than they do. Here is how van Inwagen puts it,

Fatalism is the view that "it is a logical or conceptual truth that

¹"Do you believe in fate Doctor? Our paths carved before us, guided by an invisible hand? Not Fate. Evolution. Nature's greatest force forever in flux. No. Evolution has a destination, not to combat nature, but to supersede it. The final glorious evolution?" - Arcane, S2, Ep6.

no one is able to act otherwise than he in fact does; that the very idea of an agent to whom alternative courses of actions are open is self-contradictory.”[139, p. 23]

What is fatalism?

Definition 8.5: fatalism

For every person S each of their actions is fated.

What does it mean for an action to be fated?

Definition 8.6: fated

A future action E is fated for a person S if and only if there is an extra-worldly fact that makes it such that S cannot do not- E .

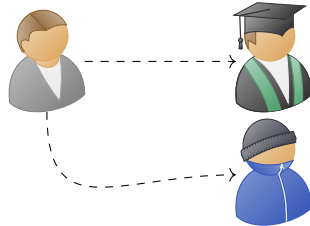
Three additional points of clarification. First, fatalism implies we do not have free will since if a person did not have the power to act differently than they did (do not- E), then they do not have free will. This conflicts with the free will thesis since this thesis says that if they are free concerning some action E or not- E , then they (1) have the power to do E and (2) have the power to do not- E .

Second, fatalism references *extra-worldly facts*. These extra-worldly facts include God, logic, or some other conceptual consideration. So, for fatalists, the reason that we cannot act otherwise is due to some fact that does not have to do with the nature of this world, e.g., physics, our biology, our desires, etc.

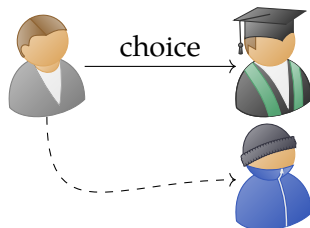
Third, the differences between strong inevitabilism, weak inevitabilism, and fatalism are not immediately obvious. To distinguish between them, it is helpful to think about what each theory says about different possible worlds.

1. If E is strongly inevitable, E occurs even in other possible worlds. Not- E is impossible.
2. If E is weakly inevitable for S , not- E occurs in other possible worlds. But, there is no possible world where you to prevent E because there are no possible worlds where you have enough knowledge to know how to prevent E .
3. If E is fated for S , it is unclear what happens in other possible worlds. But, in this world, there are extra-worldly factors why S cannot do not- E .

Let's illustrate this with an example. Suppose Tek is considering whether to continue his education or pursue a life of crime.



Tek believes himself to be free. He believes that (1) he has the power to continue his studies *and* (2) he has the power to pursue a life of crime (let's suppose he cannot do both at the same time). Now suppose that Tek continues his education.



According to the fatalist, since Tek chose to continue his studies, for some extra-worldly fact, *it is not the case that he had the power to choose a life of crime*. Thus, it is false that Tek has “has the power to continue his studies *and* (2) the power to pursue a life of crime”.

8.2 ARGUMENTS FOR FATALISM

Fatalism says that there are extra-worldly factors that make it such that we are not free. In considering arguments for fatalism, we will consider what these extra-worldly factors are that supposedly make free will impossible.

8.2.1 *From the inability to change the truth of a proposition*

What is a proposition? A proposition is something (typically expressed by a sentence) that is either true or false. I am 6'0 and so the sentence “David W. Agler is six feet tall” is true. Meanwhile, the sentence “David W. Agler is five feet tall” is false.

Next, let's consider three key claims.

First, the principle of bivalence is true.

Definition 8.7: principle of bivalence

The principle of bivalence is the principle that every proposition is either true or false (not both and not neither).

This means that for any proposition whatsoever, that proposition takes exactly one value: True or False.

It is worth pointing out that not knowing whether a proposition is true or false does not show the principle of bivalence to be false. Suppose a shop sells marbles by the scoop. You are given a brown paper bag and you absent-mindedly scoop some marbles into the bag. After you pay and before you open the bag, you say "There are 43 marbles in the bag." Again, you don't know whether there are 43 marbles. Neither does the clerk who was not paying attention when they scooped the marbles in your bag. However, your sentence is a proposition, and, as such, it is either true or false, even though you do not know which.

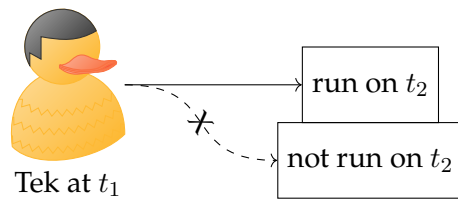
Second, not only is the proposition true (or false), but it is unchangeably true (or false). To illustrate, let's take the proposition "Tek runs 3 miles on 1 July 2040". If the principle of bivalence is true, then this proposition is true or false (one or the other). If it is true today, then it would be strange to say that the truth of this proposition can switch to false. What would change the truth value? And, if it is false, then it would also be strange that it could change to true. What would change the truth value? To illustrate, it would be strange to say the following:

1. When Tek is born, it is true that Tek runs three miles on 1 July 2040.
2. When Tek is 12, it is false that Tek runs three miles on 1 July 2040.
3. When Tek is 14, it is true that Tek runs three miles on 1 July 2040.

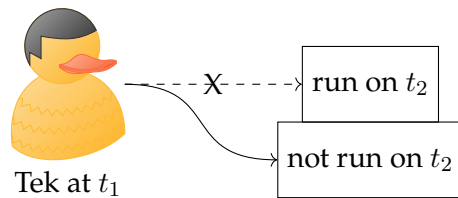
In short, propositions not only have one truth value (true or false) but that truth value is fixed, invariant, and unchangeable.

Third, if Tek do not have the power to change the truth value of a proposition "Tek runs 3 miles on 1 July 2040", then either (1) he has no power to run on 1 July 2040 or (2) he has no power to not run on 1 July 2040. In either case, he has no power to do the other activity. Therefore, he has no power to otherwise, and so has no free will.

To illustrate, suppose that it is true that "Tek runs on 1 July 2040". If that is the case, then he has no power to not run on that date.

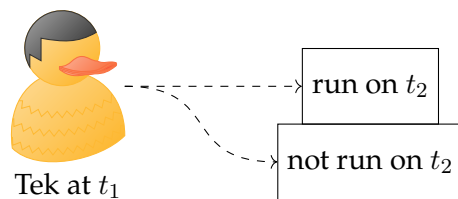


Alternatively, if it is false that “Tek runs on 1 July 2040”. If that is the case, then he has no power to run on that date.



So, while Tek has the power to run or not run, in virtue of the fact that he does one of these, it follows he lacks the power to do the other. So, Tek either he has (1) he has no power to run on 1 July 2040 or (2) no power to not run on 1 July 2040.

With the above points in mind, consider that if Tek is free, then it follows that he has (1) the power to run on 1 July 2040 *and* (2) the power to *not* run on 1 July 2040.



This however is inconsistent since both of these propositions cannot be true:

1. Tek has (1) no power to run on 1 July 2040 *or* (2) no power to not run on 1 July 2040.
2. Tek has (1) the power to run on 1 July 2040 *and* (2) the power to *not* run on 1 July 2040.

In sum, the fact that every proposition is either true or false, that the truth value of a proposition is unchangeable, and what the unchangeableness of a truth value implies for what we can and can't do shows that we are not free.¹

Exercise 8.64

Consider the proposition expressed by the sentence “you will meet the love of your life on spring break”. Using this proposition, recreate the argument from the inability to change the truth of a proposition.

8.2.2 †*Taylor's logical fatalism*

In this section, we will examine an abbreviated argument for fatalism from Richard Taylor.[\[132\]](#) To begin, consider whether a naval battle occurred yesterday. Presumably, the battle either occurred or it did not.

- P1: A naval battle occurred yesterday or it did not.

Now let's suppose that there is a perfectly reliable news organization that always correctly reports the prior day's events with perfect accuracy. Since this paper is always accurate and comprehensive, every morning you read the headlines from their web page. Let's suppose that you have no power to avoid ignoring the headlines and you read every single headline correctly. From this it follows:

- P2: If you see a headline about the sea battle, then there was a sea battle.
- P3: If you do *not* see a headline about the sea battle, then there was no sea battle.

Now, of course, this is not saying that the headline (today) caused the sea battle (yesterday). Rather, it is saying that an essential condition of there being the headline is that the sea battle occurred. Similarly, an essential condition of there not being the headline is that the sea battle did not occur.

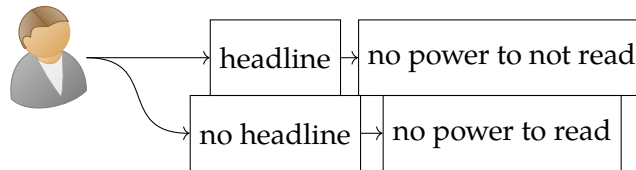
Next, consider that if the battle occurred, then you have *no power* to avoid seeing the article (for recall you are examining all of the headlines with perfect concentration). Similarly, if the battle did *not* occur, then you have no power to see the article.

- P4: If the sea battle occurred, then you have no power to not see the article.
- P5: If the sea battle did not occur, then you have no power to see the article.

If P1-P5 are true, then IC follows:

- IC: You either have (1) no power to *read* the headline or (2) no power to *not read* the headline.

But IC conflicts with your having free will about whether or not you can read the headline. For assume that you are free. If that is the case, then you have (1) the power to read the headline *and* (2) the power to not read the headline. But this is in conflict with IC.



And so,

- IC2: The free will thesis conflicts with IC.
- C: Therefore, free will is false.

Here is the argument in its entirety:

- P1: A naval battle occurred yesterday or it did not.
- P2: If you see a headline about the sea battle, then there was a sea battle.
- P3: If you do not see a headline about the sea battle, then there was no sea battle.
- P4: If the sea battle occurred, then you have no power to not see the article.
- P5: If the sea battle did not occur, then you have no power to see the article.
- IC: You either do not have the power to *read* the headline or do not have the power to *not read* the headline.
- IC2: The free will thesis conflicts with IC.
- C: Therefore, free will is false.

In short, I can't avoid seeing the headline if the battle occurred. On the other hand, I can't avoid not seeing the headline if the battle did occur. Given that the battle occurs or it doesn't and that the paper records the prior day's events with perfect accuracy, I lack the power to do otherwise.

One criticism to the above argument is that while it is true that we have no free will concerning events that happened *in the past*, we do have free will over events in the present or future. In response, Taylor asks us to consider a different scenario:

“Let us now imagine that I am a naval commander, about to issue my order of the day to the fleet. We assume, further, that, within the totality of other conditions prevailing, my issuing of a certain

kind of order will ensure that a naval battle will occur tomorrow, whereas if I issue another kind of order, this will ensure that no naval battle occurs.”[132, p. 60]

With the above scenario in place, anyone who accepts the free will thesis will think that whether you issue the order or not is up to you. That is, you have the power to issue the order and you have the power to not issue the order. However, Taylor argues that we can apply the form of the above argument to this scenario to show that you do not have both powers.

- P1: A naval battle occurs tomorrow or it does not.
- P2: If you issue the order, then there was a sea battle.
- P3: If you do not issue the order, then there was no sea battle.
- P4: If it is true that the sea battle occurs tomorrow, then you have no power to not issue the order.
- P5: If it is false that the sea battle occurs tomorrow, then then you have no power to issue the order.
- IC: You either do not have the power to issue the order or do not have the power to *not* issue the order.
- IC2: Being free conflicts with IC.
- C: Therefore, free will is impossible.

It worth emphasizing that Taylor's argument is for a fatalistic (impossibilist) conclusion since it does not depend upon whether the world is deterministic or indeterministic.

Exercise 8.65

1. Recreate Taylor's argument using your own scenario.
2. How convincing is Taylor's argument for fatalism?

8.2.3 *Theological Fatalism: Pike's Argument*

Recall that fatalism asserts that in this world, we cannot act otherwise for purely extra-worldly reasons. In a previous section, fatalism was argued for on the grounds of a logical (or metalogical) reason: that every proposition is either true or false. In this section, we'll consider how the existence of an all-knowing and eternal God supports fatalism.

Let's² consider an argument for fatalism from Nelson Pike.[101] Pike claims that “there is a selection from among the various doctrines and principles

²“In this world, is the destiny of mankind controlled by some transcendental entity, or law? Is it like the hand of God hovering above?” - *Berserk*

clustering about the notions of knowledge, omniscience, and God which, when brought together, demand the conclusion that if God exists, no human action is voluntary.”[101, p. 27]

Before stating the argument, it might be useful to state a few assumptions about God that are needed to support fatalism rather than some other view that asserts that there is no free will. Let's assume the following about God:

1. God is omniscient: knows everything
2. God is eternal: is everywhere
3. God knows everything by seeing everything (not by reasoning): As Calvin puts it “all things have ever been and perpetually remain before, his eyes, so that to his knowledge nothing is future or past, but all things are present.”³

The third point is particularly helpful for supporting fatalism because if the world were indeterministic and God's knowledge is based on his ability to reason, then even God could not predict the future.

Next, consider the following argument:

- P1: If God is omniscient, then God knows whether I will do X or not-X.
- P2: If God knows whether I will do X or not-X *and I will do X*, then it is necessarily the case that I will do X.
- P3: If God knows whether I will do X or not-X *and I will do not-X*, then it is necessarily the case that I will do not X.
- IC: Therefore, it is either *necessary that I will do X* or *necessary that I will do not-X*.
- P4: If it necessary that I will do X, then it is not possible for me to do not-X.
- P5: If it necessary that I will do not-X, then it is not possible for me to do X.
- IC2: Therefore, it is either not possible for me to do not-X or not possible for me to do X.
- P6: But, if I am free to do X (or not-X), then it is possible for me to do X and possible for me to do not-X.
- C: Therefore, if God is omniscient and eternal, then freedom is impossible (fatalism is true).

Let's consider a concrete example. Let's abbreviate the event of Tek lifting his hand on 1/1/2040 as *E*. From Tek's perspective, he is free. And so, Tek believes he has (1) the power to lift his hand *E* and (2) the power to not lift his hand not-*E*.

³Institutes of the Christian Religion, Bk. III, ch. xxi

However, recall that God is both eternal and omniscient. God not only “sees” all of time (past, present, and future) at once but also knows everything there is to know. Suppose God knows that Tek will lift his hand on 1/1/2040. If God knows this, then it is impossible for Tek to not lift his hand on 1/1/2040. Therefore, Tek has no freedom.

Why is it impossible for Tek to keep his hand down? The short answer is: because it would conflict with God's omniscience. Why is this the case? This is the case because Tek's freedom would imply at least one of the following:

1. Tek could make God believe something that is false. (impossible)
2. Tek could make God refrain from believing something that is true. If Tek could keep God from believing something about whether he would lift his hand, then God would not form a false belief (impossible).
3. Tek could make God not exist. If God didn't exist before Tek's action, then he would not “see” the event of Tek lifting his hand (impossible).

So, if we assume that God is eternal and omniscient, then Tek is only free if he has certain powers that are impossible to have (the power to fool God, to keep God from doing something, or to destroy God). But, since it is impossible that Tek has such powers, Tek must not be free.

Exercise 8.66

One obvious response to the theological argument is to say that none of this is problematic since the conclusion is only “if God exists and is omniscient, then God we are not free.” But, is there another way to (1) accept that God exists and (2) is omniscient, but (3) claim that we are free?

In the previous chapter, we considered impossibilism: the view that *regardless* of whether the world is deterministic or indeterministic, the free will thesis is false. In this chapter, let's assume that the world is deterministic. If the world is deterministic, is free will possible? The structure of this chapter is as follows. First, the theory of incompatibilism is stated. This theory asserts that if the world is deterministic, then the free will thesis is false. Second, one important argument ("the Consequent Argument") for incompatibilism is presented and evaluated. Third, we then look at the view that free will is possible in a deterministic world, and evaluate arguments for and against this view.

9.1 INCOMPATIBILISM

Incompatibilism is defined as follows:

Definition 9.1: incompatibilism

Determinism and the free will thesis are inconsistent.

The incompatibilist contends that there are no deterministic possible worlds where humanlike creatures are free. Alternatively put, the incompatibilist says "it is necessarily the case that if determinism is true, then the free will thesis is false." If incompatibilism is false, then it follows that there is at least one possible world where (1) every event (besides the first one) is determined by the prior state of affairs and the laws of nature (the world is determined) and (2) at least one event is free.

It is important to stress a key difference between the incompatibilist and the impossibilist. The impossibilist makes two claims:

1. if the world is deterministic, then there is no free will.
2. if the world is indeterministic, then there is no free will.

In contrast, the incompatibilist only asserts the first claim: if the world is deterministic, then there is no free will. This leaves open two possibilities open to the incompatibilist:

1. Positive Claim: the free will thesis is true in indeterministic worlds
2. Agnostic Claim: it is unknown whether the free will thesis is true in indeterministic worlds.

The negation of incompatibilism is compatibilism. This theory asserts that there is at least one possible *deterministic* world where the free will thesis is true. That is, free will and determinism are consistent.

The incompatibilist faces two challenges. First, they need to show how free will and determinism are not compatible with each other. They bear the burden of proof because most philosophers reject incompatibilism. In a 2020 survey of approximately 1758 philosophers, 59.16% “Accept or lean towards: compatibilism”.^[11] Since the dominant view is compatibilism, the onus is primarily on incompatibilists to show that the free will thesis is false in every deterministic world.

Second, if the incompatibilist aims to offer an answer to “whether we are free”, they also need to specify whether we are free in indeterministic worlds. That is, they either need to (1) deny that we are free in indeterministic worlds or (2) accept that we are free by showing that free will is possible in an indeterministic world (the positive side of incompatibilism). In this chapter, our focus will primarily be on the first challenge: showing how there is no free will if the world is deterministic.

9.2 ARGUMENTS FOR INCOMPATIBILISM

9.2.1 *The Consequence Arguments*

One of the primary arguments in support of incompatibilism is known as “the consequence argument”. While there are a variety of different statements of this argument, we will consider two versions. The first version we will consider is what I will call the “Simplified Consequence Argument”. The second version is a slightly more formalized statement of this simplified version. Let’s call it the “Semi-Formal Consequence Argument”.

9.2.2 *Simplified consequence argument*

The consequence argument aims to show that if determinism is the case, then we do not have the capacity to do otherwise. And, if we do not have the capacity to do otherwise (or it is impossible for us to change what we do), then we are not free.

The consequence argument is often credited to Peter van Inwagen who writes:

“If determinism is true, then our acts are the consequences of the laws of nature and events in the remote past. But it is not up to us what went on before we were born, and neither is it up to us what the laws of nature are. Therefore, the consequences of these things (including our present acts) are not up to us.”^[139, p. 16]

Let's present this argument in a more step-by-step fashion.

First, since we are trying to determine whether free will is consistent with determinism, let's assume that determinism is true. If determinism is true, then it is necessarily the case that the laws of nature and some past state of the world entail the present (and future) state of the world. Let's express this as follows:

1. Necessarily, $((\text{Laws} + \text{Past world}) \rightarrow \text{Present})$

Next, there is nothing we can do about (1). That is, there is nothing we can do about the fact that the laws of nature along with the past state of the world entail the present (future) state of the world. Let's express this as follows:

2. No one has (or ever had) any choice or control about the fact that the $(\text{Laws} + \text{Past world}) \rightarrow \text{Present}$

Third, there is nothing we can do about the laws of nature. Newton's second law is that the acceleration of an object depends directly on the net force acting on it and inversely on its mass ($F = ma$). Neither you or I or anyone has (or ever had) any choice or control over this law. While gods may be able to manipulate the laws governing the universe, humans do not. Let's express this as follows:

3. No one has (or ever had) any choice or control about the fact that the Laws

Fourth, there is nothing we can do about the past state of the world. The past is fixed, unchangeable, gone. If we have any control or choice about how things go, it is over the present or future. Let's express this as follows:

4. No one has (or ever had) any choice or control about the past state of the world.

What can we infer about the above argument. The basic inference is that if we have no control over the fact that the laws of nature and the past state of the world entail the present / future. And, we have no control over the laws of nature and no control over the past, then we have no control over the present / future. Here then is our conclusion:

5. No one has (or ever had) any choice or control about the present / future state of the world.

In other words, if the world is deterministic, then we are not free.

9.2.3 *Semi-formal consequence argument*

Skipped.

9.2.4 Criticisms of the consequence argument

There are two main criticisms of the Consequence Argument. First, there is a criticism of the claim that “no one has, or ever had, any choice about the distant past.” Second, there is a criticism of the rule that determinism implies that “no one has, or ever had, any choice about whether the laws of nature and the past state of the world entail the present / future state of the world.” The focus of this text is on the first of these criticisms.

9.2.4.1 Control over the past

A key premise of the Consequence argument is that “no one has, or ever had, any choice about whether P ”, where P is proposition about the past. Intuitively, this premise is true for two reasons. First, if an event occurs in the past and it is intuitively true that no one can change the past. Second, if P is about some time before humans existed, it is obviously true that no one has any control over times in which they do not exist. Let's consider both of these criticisms in turn.

9.2.4.2 Past-to-Non-Control Argument

For simplicity, let's abbreviate “no one has, or ever had, any choice about whether P ” as follows: $N(P_0)$ What justifies $N(P_0)$? Let's call one argument that justifies this wff, the *Past-to-Non-Control Argument* (PNCP):

- P1: P_0 is a true proposition about a time in the past: 1 July 1900.
- P2: No one has, or ever had, any choice about *any* past event *because no one can change the past*.
- C: Therefore, $N(P_0)$: No one has, or ever had, any choice about P_0 .

In short, no one has (or ever had) any choice about the past because no one can change the past.

Campbell argues that the fact that no one can change the past is “irrelevant to whether anyone ever *had* a choice about whether some true proposition about the past is true.”[20, p. 107] Just because we cannot change the past does not mean we never had a choice about the past. There are two points to tease out. Let's begin with the simpler point.

First, just because we cannot control some past event *today* does not entail we did not have a choice about that present event *in the past*. Suppose that yesterday I ate a sandwich with some spoiled meat. Today I might utter, I never had any choice about whether I would eat that sandwich because I cannot (right this second) change the fact that I ate the sandwich. The

obvious reply is: “you don't have any control now but you did have a choice in the past.”

Second, just because we could not control the fact that some past event E_1 entails another past event E_2 does not entail that we did not have any choice over E_2 . As this is a more subtle point, let's consider an example drawn from van Inwagen himself.[144, p. 236]

“*Drunk Driver*. Smith begins drinking at t_1 and a few hours later he gets in his car and drives home down a dark road in the middle of the night. At t_2 he is driving well over the speed limit and he is so drunk that both his vision and his reaction time are greatly impaired. A moment later, at t_3 , he strikes and kills Jones who is walking in a crosswalk.”[20, p. 107]

Let's list the events in the scenario above:

- t_1 : Smith drinks heavily.
- t_2 : Smith is drinking heavily and driving fast.
- t_3 : Smith crashes into Jones (Jones dies).

Now imagine how Smith might defend himself at t_4 (at his trial). He might say that he had no choice about whether he killed Jones at t_3 because he had no control at t_3 . Smith has no control over t_2 leading to t_3 and so he did not have any choice about crashing into Jones. Smith was already driving too fast, his vision was too impaired, and his reaction time was too slow given his drunken state. Because Smith has no control over this past chain of events, he never had any choice about the past.¹

- P1: $N(t_2)$
- P2: $N(t_2 \rightarrow t_3)$. Intuitively true.
- C: $N(t_3)$. Rule β .

What is the defense of P1? Here Smith might argue as follows:

- P1: $P(t_2)$ is a true proposition about the past.
- C: Therefore, $N(t_2)$. Rule χ (If an event is in the past, then no one has, or ever had, any control over events in the past).

The typical response to Smith however is to say that while Smith being intoxicated and driving recklessly is in the past ($P(t_2)$ is true), and Smith *has* no choice or control about whether this occurred now, it does not follow that Smith ever *had* any choice about the events at t_2 . In the end, Jones could have prevented the death of Smith at t_3 since he did have control over events at t_1 : he could have abstained from drinking.

So, since the mere fact that an event is in the past does not entail that we had no choice about it. In short, just because we can't change the past now does not mean that we couldn't at an earlier date. This means that rule χ is false. And, if rule χ is false, then the PNCP argument is invalid. And, if the PNCP argument is invalid, then the Consequence Argument is unsound (it contains a false premise).

Exercise 9.67

Suppose at t_2 Tek has a deeply-rooted character trait that, if asked a question by Liz, would invariably lead him to lie to Liz. Call the state of the world at this moment P_2 . Now suppose at t_3 , Liz asks him a question and Tek lies to Liz (call this P_3). Now it seems true to say that $N(P_2 \rightarrow P_3)$. Tek will then argue that he is not responsible for the lie since (1) he does not have, nor had, any choice about whether his character trait would lead him to lie: $N(P_2 \rightarrow P_3)$ and (2) since he does not have, nor had, any choice about the past, he has no choice about having the lying character trait: $N(P_2)$

1. Why might it be questionable to say that $N(P_2)$: No one has, or ever had, any choice as to whether Tek has the liar character trait?
2. Create another example of a scenario where people claim that they are not responsible for actions that flow from their characters.

9.2.4.3 *Remoteness-to-Non-Control Argument*

Let's consider another attempt to defend one of the key premises of the Consequence Argument: $N(P_0)$.

- P1: P_0 is a true proposition about the state of the world 5 million years ago.
- P2: No one can change events in the remote past because no human exists to change anything.
- C: Therefore, $N(P_0)$: No one has, or ever had, any choice about P_0 .

As Campbell puts it, someone who supports this argument might say

“since P_0 describes the state of the world at a time prior to the existence of any human beings, no one has, or ever had, any choice about whether P_0 . No one was around back then to have had such a choice.”[20, p. 108]

Campbell rightly points out that this sort of argument only supports incompatibilism in a limited way. Incompatibilism is the claim that there are *no*

possible worlds where (1) the world is deterministic and (2) human beings are free. But, the incompatibilist's argument for $N(P_0)$ only applies to *some* deterministic worlds, ones where there are events before human life begins. It would not apply to deterministic worlds where humans are around from the outset. Campbell writes,

“Consider, for instance, the possible world W . Suppose that W is a determined world such that some adult person exists at every instant. Thus, W has no remote past. At its first moment of existence lived Adam, an adult person with all the knowledge, powers, and abilities necessary for moral responsibility. Shortly after Adam comes Eve, and the rest is history.”[20, p. 109]

The supporter then of the Consequence Argument says that the argument proves that the free thesis is inconsistent with a subset of deterministic worlds (ones where there is a time before any person existing).

Exercise 9.68

1. Consider Campbell's example where Adam exists at the beginning of time and Eve emerges not too long after. Suppose that the Consequence Argument depends upon the Remoteness-to-Non-Control Argument (RNCA), and the RNCA shows that “if there are events before a person exists, then they have no power to control over those exists”. Does this mean Adam is free in the deterministic world? Does this mean that Eve is free?
2. The above criticism of incompatibilism technically shows it is (technically) false, but is this a serious blow to incompatibilism?

9.3 COMPATIBILISM

Incompatibilism contends that free will and determinism are incompatible. If incompatibilism is true, then there is no possible world where determinism is true and humanlike creatures are free. The important implication for us is that if scientists discover that *our world* is deterministic, then *we* are not free.

In contrast, compatibilism is the view that the free will thesis and determinism are consistent. That is, there is some possible world where (1) the world is deterministic (every event is necessarily entailed by the prior state of affairs and the laws of nature) and (2) the free will thesis is true.

Definition 9.2: compatibilism

Determinism and the free will thesis are consistent.

There are a few points worth clarifying about compatibilism.

First, compatibilism does not say that if the world is deterministic, then we are free. Nor does it say that if the world is deterministic, then we are *not* free. Compatibilism only says that there is *at least one deterministic world* where humans are free. It does not hold that there is free will in *every deterministic* nor does it normally aim to prove that *our world* is one of the deterministic worlds where free will exists.

Second, compatibilism does not carry any implications about whether we are free in indeterministic worlds. One might accept compatibilism but also believe (1) if the world is indeterministic, then we are free or (2) if the world is indeterministic, then we are not free.

At first glance, compatibilism looks like an absurd doctrine. How can it be the case that every single action is determined by prior events and the law of nature *yet* humanlike creatures are free? Let's begin our discussion of compatibilism by considering a few arguments to support compatibilism.

Exercise 9.69

Create a diagram that illustrates compatibilism. Your diagram should include at least one possible world where determinism is true and persons are free.

9.4 ARGUMENTS FOR COMPATIBILISM

9.4.1 *The Paradigm Case Argument*

Let's begin with the so-called "Paradigm-Case Argument" (PCA). [139, pp. 106–114] [45] [46] [149] The PCA for compatibilism has two parts. The first part begins with a claim about how words get their meaning and it uses this claim to argue that free will exists. Let's begin a simple reconstruction of this first part:

- P1: There are cases (paradigm cases) where anyone who understands the word would apply that word to those cases without any hesitation.
- P2: Many words obtain their meaning from (1) its repeated application to paradigm cases and (2) the teaching of the word to others by reference to those paradigm cases.

- IC1: Therefore, the meaningfulness of a word, name, or other descriptive expression depends upon the existence of these paradigm cases.
- P3: The expression "free will" is meaningful. It means "the ability of an entity to act in a way other than the way they acted."
- IC2: Therefore, "free will" gets its meaning from paradigm cases of humanlike creatures acting freely, e.g., cases where a person takes some action but could have acted differently.
- C: Therefore, if there are paradigm cases of free will, then humanlike creatures are free.

A few clarifications about the above argument. First, let's consider an illustration of P2. Suppose I ask you to think of a "bird", the word likely brings to your mind an example, e.g., a robin. In addition, if I were to present to you an image of a robin, and ask you what type of animal it is, you would unhesitatingly reply "a bird" or "a robin". In short, a robin or some bird in your mind is an unquestionable exemplification of the meaning of the word.

Second, words get their meaning from how they are used. Why does the word "cat" mean something like "a small domesticated carnivorous mammal, typically with short fur and retractable claws"? Why doesn't it mean "a domesticated carnivorous mammal, typically with an excellent sense of smell and a barking voice"? The reason "cat" refers to cats rather than dogs is because we use the word "cat" to refer to cats rather than dogs.

With respect to P2, if you were to asked to teach the meaning of the word "bird" to a child, you could resort to several different strategies. One way we learn new words is through definitions or descriptions. For example, someone might tell me that the word "ether" is "an invisible, massless, continuous fluid that is millions of times more rigid than steel (since light waves have high frequencies)." However, another way that words get their meaning is by reference to *actual* cases. For example, we might teach a child the meaning of "bird" by pointing to *actual* birds. And, to give the child a sense of the range of beings to which the word applies, you might point to a several different birds that exist in the wild.

Next, IC1 says that if a word is meaningful, then this meaningfulness entails the existence of certain objects. For example, if the word's meaning is a descriptive expression, then there is some object that corresponds to that object's description. Or, more concretely, the word meaningfulness of "cat" entails paradigm cases of cats.

P3 points out that "free will" is meaningful and so, by the same line of reasoning, there must be paradigm cases of free acts (IC2). The PCA ends with

the conclusion that the free will thesis is true: there are humanlike creatures with free will.

As noted, the first part of PCA aims to establish the existence of free will. Of course, this argument does not explicitly state that compatibilism is true. This takes us to the second part of the PCA, which shows that even if the world is deterministic, free will exists. There is some debate about how to interpret the second portion of the PCA, but, for our purposes, we will take the most understandable route (even if oversimplified or inaccurate). [139, pp. 107–108]

- P1: Scientific investigation could never show that we are not free.
- P2: Scientific investigation could show that the world is deterministic.
- P3: If scientific investigation did show that determinism is true, *and* incompatibilism were true, then free will would be impossible.
- P4: The impossibility of free will would be in conflict with the existence of free will (and the meaningfulness of “free will”).
- IC: Therefore, incompatibilism is false.
- C: Therefore, free will is compatible with determinism (compatibilism is true).

Let's consider this argument. P1 is supported by the first part of PCA and the following argument by analogy.

- P1.1: There are paradigm cases of color (1st PCA argument)
- IC: Therefore, science may be able to tell us what color really consists in (e.g., perception of wavelengths of light being refracted through our eyes, and stimulating photoreceptor cells in the back of the eye), but it cannot tell us that colors do not exist.
- P1.2: Analogously, there are paradigm cases of free will.
- C: Therefore, while science can show what free will *really consists in*, but it cannot show that there is no free will.

The remainder of the argument is self-explanatory. If science shows that determinism is the case then, rather than reject the scientific conclusion, the claim is that free will and determinism are compatible.

9.4.1.1 Problems with PCA

Let's consider some objections to PCA.

First, the argument appears *explosive*. That is, the argument seems to require that there exists something that corresponds to every meaningful word that has a so-called *actual event* that corresponds to it (some object that could

serve as its paradigmatic case). But, it might be argued that many words are meaningful but nothing corresponds to these words.

To illustrate, we will consider an ad hominem attack proposed by Watkins[149] against the Flew (the source of the PCA). Let's say a supporter (Flew) accepts that there is free will but that there are no miracles. Watkins argues that if the PCA proves the existence of free will, then it also proves the existence of miracles. And so, any consistent adopter of the PCA must accept that there is an object that corresponds to any meaningful descriptive or referential expression in English. In other words, Flew cannot deny the existence of miracles.

Why can't a supporter of PCA accept free will and deny miracles? Well, (1) "miracles" is meaningful, (2) we might learn the word by pointing to a variety of so-called actual cases of miracles (e.g., Jesus turning water to wine), and so (3) miracles must exist. If one wishes to accept the PCA for free will, one must also accept miracles and many other objects (e.g., ghosts).

One important point of clarification. In calling the argument *explosive*, this is not meant to imply that there exists an object that corresponds to every possible word. The supporter of PCA contends that for at least some words ("free will" being one of them), we are able to learn what a word means by first looking at what that word *denotes*, viz., actual cases of that word. Our situation is not one where we must first learn the definition of that word and then go searching to see whether or not anything in the word corresponds to its meaning.[149, p. 29] What this means then is that a supporter of PCA does not claim that there exists something that corresponds to every theoretical definition, e.g., "ether".

A second objection to PCA involves compound words. Supporters of PCA contend that the PCA does not apply to compound expressions that are built up from simple expressions, whose simple terms are definable through paradigmatic cases. For example, while the simple terms "gold" and "mountain" may get their meaning via ostension (pointing to actual cases), the compound term "golden mountain" does not. That is, if "golden mountain" is meaningful, it does not get its meaning by pointing to paradigmatic cases.

This clarification, however, raises the following question: is "free will" or "free action" a simple or compound term? The answer is not clear. Surely, we cannot take the "size" of the word to tell us whether the word is simple or compound because there are single words whose meaning is constructed through compounding simple terms. For example, Watkins[149, p. 29] points out that the word "vacuum" is defined by combining words like "space", "emptiness", and "matter" in the definition "space of empty

matter". The question then is "free will" more like "vacuum" and "golden mountain" (a compound term) or like "bird" or "mountain" (so-called simple terms).

A third objection is whether there are any *actual* paradigm cases of "free action". While there may be hypothetical cases of free action, these hypothetical cases would not warrant the conclusion that there is free will in this world. "Free will" then is analogous to "ghost". Proponents of ghosts may define the word by reference to what they call paradigm cases of ghosts, but such cases are highly debatable. We might be willing to grant them that there are hypothetical cases, but these hypothetical cases have almost no bearing on whether ghosts exist in this world. The same is true for actual cases of free action.

Suppose Tek believes his marriage to Liz was a free action. He explains the meaning of "free action" to their youngest son Jon by pointing to it as a paradigm case. However, their oldest child Renna (a student of evolutionary biology) says that the choice was not clearly free. She contends that Tek's decision was the result of the world prior to Tek's decision. His decision was forced by his brain chemistry at the time, societal norms, evolution, etc., etc.

Exercise 9.70

1. The supporter of the PCA contends that some words get their meanings by ostension (by referring to paradigm cases: things that unquestionable fit the descriptive content of the word or define its use). Is that true? Is that true for "free action"?
2. The supporter of the PCA contends that there are actual paradigm cases of "free action". Is this true? Can you think of an event in this world that is clearly an instance of free action?

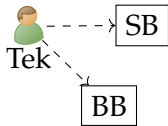
9.4.2 Argument from conditional analysis

The incompatibilist contends that if determinism is true, then even if (1) a person desires to do *E* and (2) there is nothing external preventing them from doing *E*, that person is not free if that person is not the ultimate source of their desires. In short, if a person's desired are determined, then the action that flows from that desire is not free.

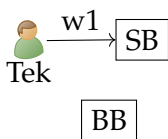
One way of thinking about whether or not a desire is determined is to consider whether or not we have *the ability to do otherwise*. If Tek acts on his strongest desire and has no ability to change that this strongest desire is to use drugs, then he has no ability to do otherwise. His desire is determined

and so he is not free.

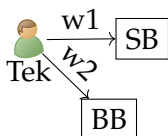
One response to this criticism is to offer a conditional analysis of freedom. Let's build up this response in steps. First, take a deterministic world where there are several cereals in front of you. You are considering which cereal to purchase. You have two choices: Smurfberry (SB) and Booberry (BB).



Suppose that this world $w1$ is deterministic and you choose Smurfberry.



Why did you make this choice? Since the world is deterministic, you made this choice because of the prior state of the world (which include your desires, biology, conditioning, etc.) and the laws of nature. However, let's consider another deterministic world $w2$ (where the physical laws are the same) where you made a different selection. Let's suppose in world $w2$, you choose BooBerry.



Why did you make this choice? Again, since the world is deterministic and the laws of nature are the same, you made this choice because the *prior state of the world* was different. For example, maybe you were born with a slightly different biological makeup that made you desire Booberry more.

What we can see from these two cases then is this: in some cases, if your past is different (e.g., you had a slight alteration to your brain), then your present and future might have been different.

Next, the incompatibilist says that free will requires the "ability to do otherwise". And, they claim, if the world is deterministic, then our actions are determined by the past along with the laws of nature. Since they are determined in this way, we lack the ability to do otherwise. In response, the compatibilist can give what is called a "conditional analysis" of the "ability to do otherwise".

Definition 9.3: conditional analysis of free action

An action A from S is a free action if and only if had S desired to do B rather than A , then S would have done B rather than A .^a

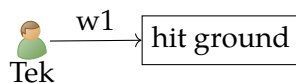
^aPhrased in terms of choice: A choice A from S is a free choice if and only if had S desired to choose B rather than A , S would have selected B rather than A .

Let's consider Tek in our world: w_1 . Tek is determined to choose Smurfberry. But, does Tek have the "ability to do otherwise". Using the conditional analysis of "the ability to do otherwise", we only need to consider what Tek *would have done if he desired differently*.

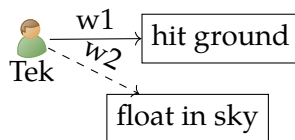
So, again, let's consider a world w_2 where Tek *desires* Booberry instead of Smurfberry. What would Tek have done in this world? We have not built anything into our story such that Tek would not act based on his desires. So, all other things being equal, Tek would have purchased Booberry. And so, Tek in w_1 does have the ability to do otherwise; therefore, Tek is free in w_1 even though his actions are determined.

It is important to point out that when considering whether Tek has the ability to do otherwise, we only consider a specific set of possible worlds. Namely, we can imagine scenarios (possible worlds) where the only facts that are altered are those necessary to change Tek's desire to make a different choice. In the actual world, Tek desired Smurfberry and purchased Smurfberry. But, when we consider whether Tek is free, we consider whether Tek would be able to choose a different cereal by altering features concerning Tek to the point that he would desire a different cereal.

To illustrate this point. Suppose Tek is in an airplane. Criminals take the passengers hostage and toss Tek out of the airplane. As Tek falls to the ground, he says "I am free to not fall to the ground."



To see if that is true, consider another world w_2 where Tek desires not to fall to the ground. In that world, would Tek not fall to the ground? No. The action of falling to the ground will occur even though Tek desires differently. Therefore, Tek's choice of whether to fall to the ground or not in this world is not free.



One more example drawn from a former student of mine. Imagine that Liz goes to buy a car. She buys a modestly-priced car within her budget. As she leaves the car lot, she considers whether she could have acted differently. That is, was she free to purchase a different car? The conditional compatibilist determines whether this is the case by considering a world (or scenario) where she desired differently. Let's consider two such scenarios:

- Scenario 1: There are other cars within her budget. In this scenario, if she had desired differently, she would have acted differently. In this scenario, she is free.
- Scenario 2: The car dealer only has luxury cars that Liz could never afford. In this scenario, if she had desired differently, she would not have acted differently (since she would not be able to purchase the other car). In this scenario, she is not free.

In sum, the incompatibilist says that we are not free because our actions are determined by our desires so we cannot act in any other way than we do. In responds, the compatibilist says that our actions (choices) are free provided if we desired differently, we would have acted differently.

Exercise 9.71

Create an example where a person does an action X . Now consider (1) a counterfactual scenario where had they desired differently, they would have acted differently (done Y rather than X) and (2) a counterfactual scenario where had they desired differently, they would not be able to act differently (unable to do Y rather than X).

Let's unpack the argument for compatibilism from a conditional analysis of freedom:

1. P1: S is free if and only if S did X, then S could have done Y.
2. P2: S could have done Y if there is a counterfactual scenario in a deterministic world where S wanted to do Y and did Y.
3. P3: There are other counterfactual scenarios in deterministic worlds where S wanted to do Y and did Y.
4. C: Therefore, even if determinism is true, then S is free.

Exercise 9.72

Are you convinced by this argument? If it is flawed, which premise is false?

9.4.3 *Objections to the conditional analysis*

Let's consider some objections to compatibilism. First, consider that we are often torn between two desires. To illustrate, suppose Tek has two primary desires:

1. Tek desires Smurfberry
2. Tek desires Banana Frosted Flakes (BFFs)

Now suppose that Tek chooses Smurfberry. Intuitively, Tek had the ability to choose BFFs. But, on the conditional analysis this is false since "if Tek had desired to choose BFFs, then he would have chosen BFFs" is false. It is false because Tek did desire BFFs but he did not choose them. The conditional analysis then needs some type of revision. One popular revision is to say that we *choose based on our strongest desire*.

Definition 9.4: conditional analysis (with strongest desire)

Assuming *S* did *A* rather than *B*, *S* was free to do *B* rather than *A* if and only if had *S* more strongly desired to choose *B* over *A*, then *S* would have chosen *B*.

At this point then, the freedom to choose or do otherwise does not require that you are actually able to different than you are determined to act. Instead, it only requires that *had you a different strongest desire* at the time you originally acted one way that you would have acted differently in that circumstance.

Exercise 9.73

How convincing is the conditional analysis of the "ability to do otherwise"?

Let's review. The world is either deterministic or indeterministic. So far, we have considered the prospects of our being free in a deterministic world. Incompatibilists contend that there is no freedom in any deterministic world. Compatibilists contend that there is at least one possible deterministic world where humanlike creatures are free.

But what about if the world is indeterministic? Indeterminism is the view that not every event in the world is entailed by the prior state of affairs and the laws of nature.

Definition 10.1: indeterminism

The view that not every event is determined by the prior state of affairs and the laws of nature.

If the world is indeterministic, then there appears to be two main theories about whether we are free:

1. No, there are no indeterministic worlds where humanlike creatures are free.
2. Yes, there is at least one indeterministic world where humanlike creatures are free.

One version of the second theory is known as *metaphysical libertarianism*. This theory asserts (1) free will and determinism are incompatible and (2) free will is only possible in an indeterministic world.

Definition 10.2: metaphysical libertarianism

Metaphysical libertarianism is the theory that (1) free will and determinism are incompatible, and (2) that free will is only possible if the world is indeterministic.

The metaphysical libertarian faces two main problems.

First, since libertarians accept the reality of free will in an indeterministic world, they must show that free will is incompatible with determinism. This is known as the **ascent problem** for libertarians. To do this, they can either reject determinism as an empirical thesis or they can accept *incompatibilism*. One way they might accept incompatibilism is through putting forward some version of the **Consequence argument**.

Second, given that they reject determinism, and thereby accept indeterminism, they must show how free will is compatible with indeterminism. This part of the problem is known as the **descent problem** for libertarians.

Exercise 10.74

State the general problem of free will. That is, suppose you are extremely skeptical about there being free will, how would you state the general problem to a complete stranger?

10.0.1 Three types of libertarianism

Let's just suppose that some form of the Consequence Argument is true. This solves the ascent problem. There are three different ways of solving the descent problem. These correspond to three different types of libertarianism:

1. non-causal libertarianism
2. event-causal libertarianism
3. agent-causal libertarianism

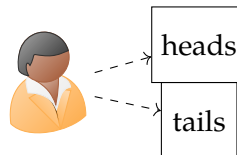
10.0.2 Non-causal libertarians

The first species of metaphysical libertarianism we will consider is non-causal libertarianism (NCL).

Definition 10.3: non-causal libertarianism (NCL)

Non-causal libertarianism (NCL) contends that a action A by S is free iff (1) S performed the action and (2) the action is not determined by prior causes and the laws of nature, viz., the action is indeterministically caused.

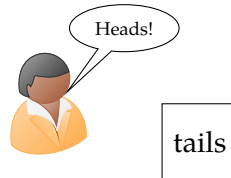
To illustrate, suppose a coin is to be flipped. Alice is asked to choose *heads* or *tails*. Assume that choice is indeterministic. There is nothing about the prior state of affairs and the laws of physics that entail which choice Alice will make.



Since Alice's choice is indeterministic. If we were to rewind time and run the scenario again (where the world is exactly the same way), Alice might

pick differently. She might pick differently because there is nothing about the state of the world and the laws of nature that entail which choice she will make.

Next, let's assume that Alice makes a choice. She selects *heads*.



According to NCL, since Alice is the person making the choice and the action of choosing is indeterministic, Alice's action is a free action.

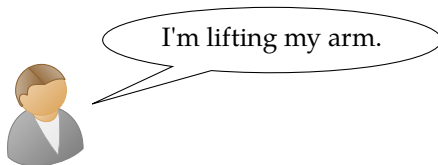
Exercise 10.75

Think of some action that you made but you feel that if you were in exactly the same situation, you would act differently.

10.0.2.1 Criticisms of NCL

The most glaring problem with NCL is that the theory makes *every* act performed by a subject a free action. The problem with this is that there are indeterministically-caused actions performed by agents that we would not consider free. Let's consider some examples.

First, consider some action by Tek that fits the NCL definition of free but was *spontaneous or random* rather than connected with the desires or intentions of Tek. To illustrate, suppose the teacher asks a class a question and Tek spontaneously lifts his arm.



Tek made no conscious decision to raise his arm but he raised it anyway. In other words, the act was not under Tek's control but was a spontaneous event performed by him. This act seems to fit the NCL definition of free but intuitively we would say that Tek's act was not free at all since it was simply a random event performed by him and connected in any way to his desires or

intentions. It would be more accurate to say that the act was due to *cosmic chance* than Tek exercising his power of free will.

Second, suppose Liz desires to not eat a sandwich she made earlier. She wants to save it for later. However, Liz is busy at work and unconsciously begins eating the sandwich. Let's say that the act is not determined. Surely the act is performed by Liz. It is therefore, according to NCL, a free act. However, intuitively we might say that acts that are *not under our conscious control are not free*. In fact, eating the sandwich was against Liz's will rather than unconsciously in line with it.

Third, suppose two subjects Tek and Jon. Suppose they are identical in every way or suppose they are the identical but exist at different possible worlds. Now suppose they both face the same scenario:

1. kill David for evil gain
2. save David at personal cost

If Tek and Jon are the same, indeterminism is true, and their action results from chance, then if Tek chooses to kill David and Jon chooses to save David, then there doesn't look like there is any grounds for condemning Tek and praising Jon. The fact that one chose to kill David and the other chose to save him was due to chance, luck, or the randomness that exists in the world. The non-causal libertarian is faced with explaining how a subject can be said to (i) act otherwise given the same past and same laws of nature yet (ii) act in a way where they can be said to be responsible for that act.

Exercise 10.76

Create your own scenario where *S* does *X*, *X* is not determined, but (intuitively) *S* cannot meaningfully be said to be responsible for *X*. Try to make it as realistic as possible.

10.0.2.2 *Extra-factor strategies*

We saw that NCL's definition of freedom is too inclusive: it makes every non-deterministic action performed by a subject a free act. To fix this problem, we need to be *more restrictive* about the number of acts that count as free. To achieve this, metaphysical libertarians employ what are called "extra-factor" approaches. They agree with the two major conditions of a free act but say that a free act requires something in addition to (1) being indeterministic and (2) being performed by us. That is, the definition of a free action needs to look something like this:

Definition 10.4: extra-factor definition

An action A by S is free iff (1) S performed the action, (2) the action is non-deterministic, and (3) and *this extra factor*.

There are at least two theories that try to specify this extra-factor. These are the following:

1. event-causal libertarianism
2. agent-causal libertarianism

Both of these types of libertarianism are called **extra-factor** libertarianism since they will posit some *extra-factor* that is present when an agent makes a free choice.

10.0.3 Event-causation Libertarianism (ECL)

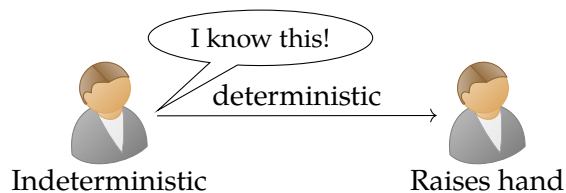
Event-causation libertarianism (ECL) asserts the following:¹

Definition 10.5: event-causal libertarian (ECL)

Event-causal libertarianism (ECL) says that an action A by S is a free action iff:

1. A is performed by the subject S ,
2. an indeterministic event E causes A ,
3. E is an “apt” mental events (desires, reasons, beliefs, intentions) that are nondeterministically caused and found in S (the “extra-factor”)

Let's illustrate. Suppose Tek is class and his teacher asks the class a question. An indeterministic event occurs in Tek's mind: I want to raise my hand and answer this question. This event causes Tek to raise his hand.

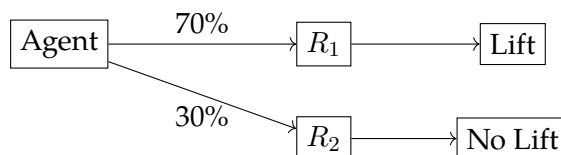


ECL asserts that this action is free because (1) Tek performed the action, (2) there is an indeterministic event that caused him to raise his hand, and (3) the indeterministic event E that caused him to raise his hand is “apt”, meaning that it is relevantly related to the action he performed.

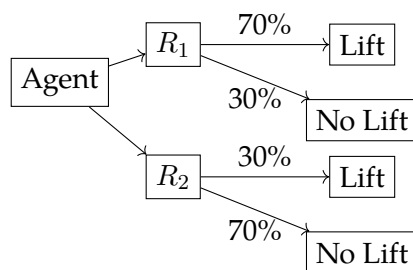
Let's make several clarifications about this view.

First, the connection between the free act and the mental event that caused the act must be relevantly related (“apt”). To illustrate, suppose again the teacher asks the class a question. Tek is thinking about his favorite video game and this thought causes him to raise his hand. Such an action would not be considered free because the act is not aptly related to the random mental event that occurred in Tek's mind.

Second, free action involves a non-deterministically-caused mental event that deterministically causes an act by the subject. If the act itself by the subject were indeterministic, then ECL would be no different than NCL. Let's make this explicit with an example. Suppose that if Tek considers reasons R_1 for lifting his hand, then he will raise his hand. And, if he considers reasons R_2 , then he won't raise his hand. Next, let's suppose that which reasons he will consider is not determined by the laws of nature and the world. We can assign each a probability: there is a 70% chance he will consider R_1 and a 30% chance he will consider R_2 .



In the above example, a person's reasons do not nondeterministically cause outcomes. If that were the case, then the following would be true: if Tek considered R_1 , then there would be a 70% chance of him lifting and 30% chance of him not lifting his hand. While if he considered R_2 , there would be a 30% chance of him lifting and 70% chance of him not lifting.



It is worth stressing that one of advantage of ECL over NCL is that if an act is freely done by the subject S , then while the act has a nondeterministic cause, that cause is not wholly unconnected to the subject who performed them. That is, in order for the act to be free, it must be not only performed by the subject S but caused by the reasons or motivations that the person S considered when they acted. As Balaguer puts it,

“If an ordinary human torn decision is wholly undetermined, then it is L-free---that is it is not just undetermined but also appropriately nonrandom, and the indeterminacy increases or procures the appropriate nonrandomness.”

When discussing NCL, it was stated that when a person performs a free act, if we were to rewind time, they might act differently. The same is true with ECL. The only difference is that the act would be different because they would be considering different reasons.

Exercise 10.77

Create an example where a person chooses to do something based on one set of reasons but if we were to rewind time, they might choose to do something else based on another set of reasons.

10.0.3.1 Criticisms of ECL

The primary objection to the event-causation theory is that it is the *reasons* (mental states) that determine what a person does rather than the *person*. Thus, the event-causation picture is insufficient for any real self-determination or choice.

To illustrate this objection. Suppose that Tek is offered a piece of candy and there is a 50% probability that he will accept the candy and a 50% probability that he will reject the candy. Now suppose we rewind the causal state of affairs 100 times and examine Tek's choice. In 55 of the scenarios, Tek chooses to accept based on reasons *R1*. In 45 of the scenarios Tek chooses to reject based on reasons *R2*. There does not seem to be any account of how Tek himself is deciding on whether to accept or reject the candy. Rather, the event-causation picture merely says that Tek's reasons result in him choosing one way rather than another. To put it bluntly, it is a random selection of Tek's reasons, not Tek, who is choosing.²

10.0.4 Agent-causation Libertarianism

Recall that event-causal libertarianism (ECL) attempted to solve the problem of non-causal libertarianism by adding an *extra condition* necessary for a free act. However, in the previous section, we saw that ECL faces the problem that it makes free acts due to random events that happen in the brain (or mind) rather than choices made by agents themselves.

Agent-causal libertarianism (ACL) tries to fix this problem by asserting that

free acts are not due to random mental events in a subject's brain but are non-deterministic events that are caused by *the agent themselves*.

Definition 10.6: agent-causal libertarian (ECL)

Agent-causal libertarianism (ACL) says that an action *A* by *S* is a free action iff:

1. *A* is performed by the subject *S*,
2. *A* is indeterministically caused
3. *A* is due to "agent causation".

There are two key questions that ACL needs to answer:

1. what is agent causation?
2. what is an agent?

Let's consider these questions below.

10.0.4.1 Agent causation

The agent-causation theory posits a unique type of causal relation alongside the normal cases of causation. First, let's consider an example of *standard causation*, which can be called "event-event causation." In event-event causation, the cause of an event is another event.

To illustrate, imagine Liz pushing Tek and then Tek falling. The event of Liz pushing Tek *caused* Tek to fall. Here there is a relation between the *event* of Liz pushing Tek and the *event* of Tek falling.

Here is another example of event causation. Again, take the event of Liz pushing Tek that leads to Tek falling. We can characterize the events as follows:

1. Liz has the desire to push Tek
2. Liz raises her bent arms and moves toward Tek
3. Liz's hands come in contact with Tek and she extends her arms
4. Tek loses his balance and begins to fall toward the grounds
5. Tek hits the ground

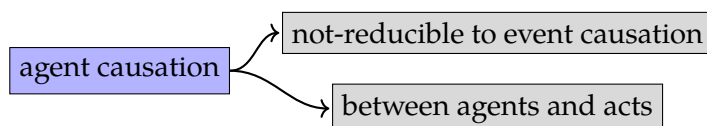
Again, in this example, we have the *event* of Liz desiring to push Tek, a series of intermediate events, and the ultimate *event* of Tek hitting the ground.

The agent causation theory contends that not only are there instances of causation where one event is related to another event (*event-causation*), but there is also **agent causation**.

Definition 10.7: agent-causation

A type of causal relation between an *agent* and an *act* that is not reducible to a type of causal relation between events (whether these events be mental/physical)

What distinguishes agent-causation from event causation is (i) the relation of the relation (it is between agents and acts, rather than events and other events) but also (ii) that the relation is said to be **non-reducible** to event causation that is, it cannot be explained in terms of relations between events (whether these be mental or physical events).



In other words, in agent causation, free choices are made by the whole agent, not the reasons in their mind or brain events in their brain or some part of their personality.

10.0.4.2 Agent causation view of personhood

The next key component of the agent-causal theory is its conception of a person (agent).

First, an agent is a type of thing that can make choices without that choice being caused by prior events. They are the type of thing that decide between A and B but their decision to do A rather than B is not settled by the prior state of the world. That is, they can choose without having their choice be caused by their brain states or reasons.

To illustrate, suppose a chain of events e_{-2}, e_{-1}, e_0 leading up to event e_1 . Let's suppose the causal scenario at e_0 is such that it does not settle whether e_1 or some other event will occur. Let's call this e_{+1} . The agent-causation theory seems to say that an *agent* settles whether e_1 or e_{+1} will occur by acting. Now if the agent's action was caused by the *events* at e_0 , either indeterministically or deterministically, then the decision would not be by the agent but by the prior events. So, the agent-theory posits the existence of a subject that is capable of acting on the world without being acted upon by it. Philosophers often say that agent theorists take the agent to be a *substance*.

Second, it is important to stress that the free choice of an agent is not made by some *event that is inside me*. To say that my brain states or my reasons are re-

sponsible for the action is to say that some *event inside of me* (or some part of me) is the cause of the choice. Even if that event inside of me is nondeterministic, then it is not me who is performing the act (it is the nondeterministic event). As Taylor (p.217) puts it,

“When I believe that I have done something, I do believe that it was I who caused it to be done, I who made something happen, and not merely something within me, such as one of my own subjective states, which is not identical with myself. If I believe that something not identical myself was the cause of my behavior---some event wholly external to myself, for instance, or even one internal to myself, such as a nerve impulse, volition, or whatnot---then I cannot regard that behavior as being an act of mine, unless I further believe that I was the cause of that external or internal event.”

Taylor also notes (in contrast to the event-causal theory) that my reasons cannot be the cause of my action since if an action:

“is free, it must not only be such that it is caused by the agent who performs it, but also such that no antecedent conditions were sufficient for his performing that action. In the case of an action that is both free and rational, it must be such that the agent who performed it did so for some reason, but this reason cannot have been the cause of it.”

The agent-causal person's actions are not reducible to some part or event in the person. That is, the person is not their brain states or their reasons or their desires. So, in the case of a free action *X*, it would not be correct to say that your brain is the indeterministic cause of *X* or your reasons or motives are the indeterministic cause of *X*.

A couple additional points about the agent-causation theory:

1. the libertarian thinks that free acts cannot be completely uncaused (otherwise they are just due to chance),
2. nor do they think that events are completely caused by the prior state of affairs and the laws of nature
3. the agent-causal libertarian posits a unique type of cause between an agent/subject/self (that has continued existence) and an act/event
4. the libertarian is committed to the idea that the agent / subject / self is not a thing, event, state of affairs, etc., but a unique type of thing capable of acting on the world.

One argument put forward by agent-causation is that in order to explain any act by a subject, we ultimately need to posit some sort of agent-causation. One argument runs something like this:

- P1: Every act needs a cause.
- P2: Some causes can be explained by prior events (event causation)
- P3: Some causes / events cannot be explained by prior events (not event causation)
- P4: The type of cause that could explain these causes is agent-causation
- C: Therefore, there is reason to believe in agent-causation

One of the key premises of this argument is **P3**. One type of event that is cited as being unexplained by event-causation is the *first event* in a sequence of events. That is, since every chain of action needs a beginning (a starting point), it is claimed that the starting point of the chain of action is not another event, but an act by the agent.

The general idea is that while we might be able to explain some act in terms of prior events, *ultimately* some acts need to be explained not in terms of events (states of affairs) but in terms of an agent directly causing something to happen.

For example, suppose Liz pushes Tek. Suppose that Liz pushes Tek and we ask ourselves what caused this event. We could explain her pushing Tek in terms of prior states of affairs and events:

1. Liz's pushing of Tek was caused by Liz raising her arms and exerting force toward Tek
2. Liz's raising her arms and exerting force toward Tek was caused by her desire to push him
3. Liz's desire to push Tek was caused by an event in her brain

But ultimately, we need to explain what caused the event in her brain to occur. And, proponents of agent causation will contend that is due to the agent.

Exercise 10.78

Try to explain agent-causal libertarianism with an example.

10.0.5 Criticisms of ACL

Let's consider some objections to agent-causal libertarianism. The first objection is that agent-causation is mysterious.³ Agent causation sounds like

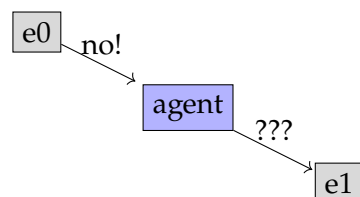
an immaterial thing is blasting into the physical world and altering its trajectory, viz., something outside the laws of nature that nevertheless acts in this world. This is entirely fantastic. As Kane[64, p. 33] (p.33) puts it “libertarians have posited transempirical power centers, nonmaterial egos, noumenal selves outside space and time, unmoved movers, uncaused causes, and other unusual forms of agency or causation--thereby inviting charges of obscurity or mystery against their view.” Kane[64, p. 34] characterizes the descent problem for libertarians metaphorically when he writes:

“The air is thin and cold up there on Incompatibilist Mountain; and if you stay up for any length of time, say critics of libertarianism, your mind gets foggy. You start having fantastical ideas, such as transempirical power centers, noumenal selves, and unmoved movers, which libertarians have often invoked to explain their view.”

A second objection. Even if we accept that an event can be (mysteriously) caused by an agent, what is the argument for saying that an agent's act or decision is entirely undetermined by prior events and the laws of nature? The libertarian contends that agents are not the types of thing that is determined by prior events and laws of nature (either indeterministically or deterministically). But this is not argued for and there is no reason to believe it.

As Kane[64, p. 48] puts this objection: “[t]he first problem [for the libertarian agent-causation theory] is this: how can we know from the immediate experience of our own agency alone that our actions are not determined by events (some of which may be hidden from us)?”

Third, let's put aside both of the previous objections and suppose that (1) agent-causation is possible, (2) an agent's actions are not determined by prior events, including reasons, beliefs, intentions, etc. If agent-causation is not determined by the agent's biology, their reasons, their values, their intentions, then the agent's decision to do *X* rather than *Y* seems no different than mere chance. The reason it is seems no different from chance is because no part of what makes up who the agent is, e.g., their beliefs, reasons, intentions, psychology, etc., plays a role in determining what action the agent will make.



In other words, on what grounds would the agent be choosing one option over another if not on the basis of their desires, thoughts, deliberations, prior experience, prior conditioning, etc. The same is true for cases involving assigning moral praise / blame for if there is nothing about the agent's powers, capacities, beliefs, prior experience, etc. that is responsible for the choice for one act over another, then the fact that an agent acted in one way rather than another seems completely spontaneous.

The agent-theorist can respond by saying that there is a difference between events that are mere chance vs. those that are the result of agent causation.

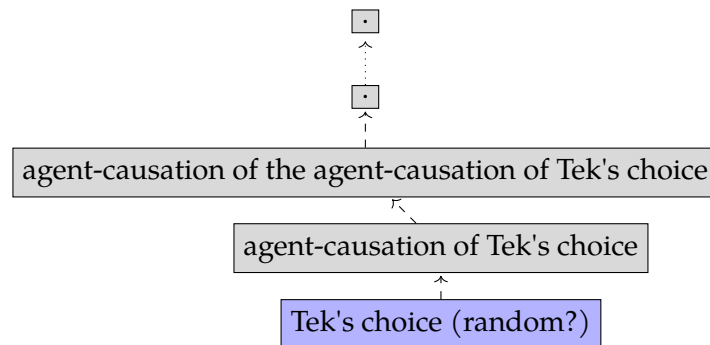
1. in the case of *chance*, the event could be caused by anything. That is, if we say that event *e* is not determined by the prior state of affairs or the laws of nature, then *e* could be determined by *anything*!
2. in the case of *agent-causation*, the event was not be caused by anything. It is instead caused by the *agent*, and this is a type of event is not explainable using the laws of nature nor the prior state of affairs.

Thus, in the case of moral praise and blame, it isn't mere luck that an individual chooses to do good rather than evil. If Tek has the choice between selfish gain or self-sacrificing charity, Tek's decision to choose one over the other is the result of *agent-causation*. That is, Tek's choice is caused by agent-causation, and since it is caused by the agent, it is not caused by anything (random).

No: choice is result of agent-causation

Tek's choice (random?)

But consider the following objection. While Tek's choice may not be random because it is due to agent-causation, if agent-causation is not determined by anything, then agent-causation is random. And were the advocate of agent-causation to say that that instance of agent-causation is itself caused by agent-causation, then again there would be the question of whether the agent-causation of the agent-causation is random, ad infinitum.



A final response to this is to say that agent-causation is a special kind of causation that doesn't need to be caused or influenced by prior events in order to be non-random. As Kane[64, p. 51] characterizes the position "[i]n response to the objection that for all we know immanent agent-causation might be determined by hidden causes, they insist that agent-causation is not the sort of thing that could in principle be caused or determined by prior events or circumstances." In addition, Kane[64, p. 51] notes "[i]n response to the randomness and luck objections, they add that the agent-causal relation is not the sort of thing that could in principle occur randomly or by chance either, since it is the agent's consciously controlling something."

Exercise 10.79

Do you think a libertarian view that accepts agent causation works? Explain your reasoning.

We can't keep ourselves from deliberating, planning, weighing the pros and cons of our actions. Even if we do not have many long-term plans in life, we at least deliberate over better or worse outcomes in the short-term.

But the act of deliberation seems to imply a belief in free will. So, it seems that anyone who says that there is *no free will* is contradicting themselves because:

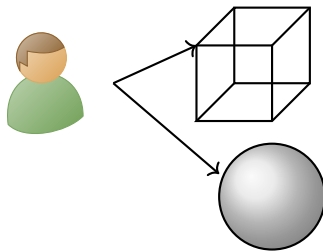
1. They say there is no free will.
2. They deliberate which implies they believe in free will.

This chapter explores this seeming contradiction.

11.1 DELIBERATION AND FREE WILL

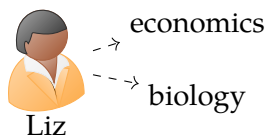
Deliberation does not entail that *we are free*, but it does seem to entail that *we believe in free will*. Let's see why this is the case.

When we deliberate, we consider at least two actions: *A* and *B*. In doing so, we believe that both *A* and *B* are possible courses of action. For example, suppose a cube or sphere is placed in front of me.



If I believed that it was impossible for me to pick up the cube (e.g., because it weighed a million pounds), I would not say to myself "I'm not sure which object I will pick up. Let me weigh the pros and cons." It thus appears that the activity of deliberation implies the belief that given two actions *A* and *B*, that one at least *believes* has the power to do either *A* or *B*.

Let's consider a more practical example. Suppose Liz, a very capable and smart student, is trying to decide whether to major in economics or biology.



Her university offers both majors and she has the prerequisites for both. Before making her decision, she carefully evaluates the pros and cons of each major. Even if Liz is destined to choose biology over economics, Liz's act of deliberating seems to imply that she believes she could take economics.

Now it may be the case that our belief in free will is false. This would entail that deliberation is a foolish activity since while one believes one can do either *A* or *B*, one of these choices is either inevitable, determined, or due to chance operations that are outside the control of the agent. Nevertheless, the activity of deliberation indicates that the agent believes that they are free.

11.2 THE INCONSISTENCY

In the previous section, it was argued that deliberation entails a belief in free will. But let's consider someone who *denies* that there is free will. This approach seems contradictory. You cannot believe in free will and not believe in free will. Let's spell out this inconsistency:

1. Tek says there is no free will.
2. Tek deliberates.
3. Deliberation implies a belief in free will.
4. (1) and (3) are inconsistent.

Generally, when a person has two inconsistent beliefs, the diagnosis is that the person is *irrational*. This is because both beliefs cannot be true. The typical prescription for this irrationality is to give up at least one of these beliefs.

In the following subsections, let's consider (1) whether this is the correct diagnosis and (2) whether this is the correct prescription given the diagnosis.

11.3 NO INCONSISTENCY: ACTIONS AND WORDS

Let's consider the possibility that the person that we have described (who denies free will and deliberates) does not have inconsistent beliefs. How is this possible? Consider that the argument showing the inconsistency does two things:

1. extracts a person's beliefs from what they say ("there is no free will")
2. extracts a person's beliefs from what their actions ("there is free will")

Now consider that people often say that "actions speak louder than words". What people mean by this is that we should assign more weight to a person's actions than their words: their actions serve as better evidence for what they truly believe.

If it were the case that a person's real beliefs are derived from their actions rather than their words, then we might say that the person we have described does not contradict themselves in the sense of having *two conflicting beliefs*. Here is our new diagnosis and prescription:

- Diagnosis: The person's actions indicate that they believe in free will but they are saying that they don't. The person does not really believe what they are saying.
- Prescription: The person should change what they are saying so that it better reflects what they really believe.

Let's consider a concrete application of this. Consider the following example adapted from van Inwagen[139, pp. 153–161]:

Suppose that a man is asked whether he believes his wife is faithful. He says yes. He is asked to swear to it, and he responds "I solemnly swear that I believe my wife is faithful." He is then taken to court, sworn in, and then firmly reminded that if he lies under oath, he will commit perjury, facing a serious punishment. The man again insists that he believes and always has believed that his wife is faithful. After all of this attesting, a court official brings a stack of evidence into the courtroom. There are logs of the man checking his wife's phone messages and emails when she is asleep, video of him following her when she's out on errands, flurries of text messages asking about her whereabouts and questioning her when she is out with her friends.

What are we to say about what the man believes concerning the fidelity of his wife? According to our account thus far, we would diagnosis the man's situation as follows: (1) he *says* he believes she is faithful, but (2) he *does not* really believe she is faithful because his actions indicate he thinks she is cheating on him. His actions speak louder than his words. According to our prescription, the man should change what he is saying.

Exercise 11.80

Create an example where someone says they believe *A* but their actions indicate they believe not-*A*.

11.4 PROBLEM 1: ACTIONS AND WORDS

Our approach thus far is to say that a person's actions always better indicate their real beliefs. This would be false in two cases:

1. The person's actions and words equally indicate their real beliefs (neither holds greater value).
2. The person's words better indicate their real beliefs.

Let's consider each case.

First, suppose that someone believes what they say and what they do. If that were the case, then it is possible that when the words and actions of a person contradict, it is not *always* the case that their actions indicate their real beliefs. Instead, the words and actions both indicate their real, albeit contradictory beliefs.

In the earlier case involving the man who says he believes his wife is faithful but his actions indicate he believes his wife is unfaithful. It is possible that the man simply has inconsistent beliefs. And, perhaps this is true with respect to free will and deliberation. Someone might say they believe *there is no free will* but the activity of deliberation indicates they believe *there is no free will*.

Second, while van Inwagen[139, p. 157] considers the possibility that a person can have contradictory beliefs (one belief stemming from what they say, the other from what they do), there is another possibility: perhaps my words trump my actions. That is, perhaps there are cases where what a person really believes is what they say.

Let's consider an example. Suppose that I am riddled with pessimism and self-doubt. I have an exam tomorrow and if I fail, then I will fail the course. You agree to study with me, but while we are studying you repeatedly hear me say "I will surely fail. There is no point in taking it." I proceed to study for the exam, but all the while you hear me saying "this is pointless. Barring some unforeseen event, there is no way I can achieve a passing grade." On the day of the exam, you see me present, ready to take the exam. Do I believe I can pass the exam? My actions (the studying, the attending the exam) seem to indicate that I believe I can pass. But, my words indicate that I do *not* believe I can pass.

It is not clear to me that my actions in this scenario are a better indicator of my real beliefs than my words. In fact, it appears that I don't believe I can pass, but I'm taking the exam for no good reason. Perhaps I am taking it out of habit, or as a way to pass the time, or find great amusement in realizing the inevitable.

In addition, were you my study partner, your actions would likely indicate that you believe my words better indicate my real beliefs than my actions. Consider what you might say to me after I utter "there is no way I can pass." Being the nice person that you are, you might try to instill some hope in me

by saying, ``Don't say that. You can pass. Let's go over the exam study guide one more time.''

Exercise 11.81

Try to create a scenario where a person's words (what they say they believe) is a better indicator of what they believe than their actions.

11.5 PROBLEMS 2: THE PRESCRIPTION

Let's consider the other part of our story. We said that when a person's actions conflict with their words, they ought to change their words so that the two align. This seems to assume the following:

Definition 11.1: no contradictions principle

Whenever a person's beliefs or actions are inconsistent (or seemingly inconsistent), one ought to take action to remove the inconsistency.

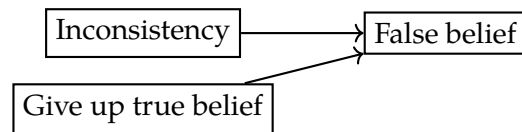
Is this true? The idea is that any rational person who, when realizing that their beliefs are inconsistent, is obliged to give up at least one of their beliefs so as to ensure they are consistent. Why? Well, since the man's beliefs are inconsistent, one of them is false. And, having false beliefs is a bad thing, rationally speaking. Even further, some philosophers point out that having an inconsistent set of beliefs is even worse than having a merely false belief because not only do you have a false belief but the *inconsistency* of your beliefs *logically ensures* one of your beliefs is false.[139, p. 158]

Let's consider the possibility that there are some cases where a person might have genuinely have inconsistent beliefs but they *ought not* to remove the inconsistency. van Inwagen[139, p. 159] points to a case of a Japanese astronomer who (1) cannot give up on his belief that the sun is a divine being but (2) intellectually shares the belief held by his fellow astronomers that the sun is an astronomical body. If consistency is so important and the Japanese astronomer *cannot* give up on belief (1), then the astronomer should give up on belief (2).

But there appears to be no practical benefit in doing this. And, it might be argued, that there are situations where it would be of practical detriment to abandon an inconsistent set of beliefs. Let's consider this claim in two steps. First, we will show that there are scenarios where there is *no* practical benefit in giving up a belief to obtain a consistent set of beliefs. Second, we will

reconsider this same scenario, pointing out how abandoning inconsistency for consistency comes is practically unwise.

Consider the options of the Japanese astronomer. First, the astronomer might continue to hold both beliefs: the sun is divine and the sun is an astronomical body. His beliefs are inconsistent and so he has a false belief. Second, the astronomer might give up the only belief he can (his Western astronomical belief) so that his beliefs are consistent, retaining the belief that he simply cannot give up: the sun is a divinity. His beliefs are no longer inconsistent, but he still has a false belief.



The Japanese astronomer will either have (1) an inconsistent set of beliefs (where one is false) or (2) a consistent set of beliefs (where one is false). If what makes having an inconsistent set of beliefs bad is that it entails having a false belief, then there is no reason why the consistent set is preferable to the inconsistent set. The Japanese astronomer will have a false belief no matter what.

Second, a case might be made for claiming that the inconsistent set of beliefs is preferable. For reconsider the options of the astronomer. In giving up the Western astronomical beliefs for a consistent set of beliefs, the astronomer gives up a true belief so that his beliefs may be consistent. This may be practically disadvantageous since the astronomer gives up all of the advantages associated with having a true belief, e.g., perhaps he treats the sun as an astronomical body when he is acting as a professional astronomer.

It is claimed then that the Japanese astronomer is in the same situation as someone who says there is no free will but consistently deliberates.

Sun is divine	Deliberates
Sun is astronomical body	There is no free will

The astronomer believes that the sun is an astronomical body, but also believes (and could never stop believing) that the sun is divine. Analogously, people claim that there is no free will, but also can't act in a way that indicates they believe in free will (since they deliberate). The astronomer could give up on their belief that the sun is an astronomical body, but this appears to have no practical benefit and / or be practically disadvantageous. Analogously, the denier of free will could give up on the belief that there is no free

will, but, it is claimed, this would be to give up on a true belief for the sake of consistency.

Exercise 11.82

Create a scenario where (1) a person has two inconsistent beliefs *A* and *B*, (2) they cannot give up on belief *A*, and (3) giving up on belief *B* so that their beliefs would be consistent would either (a) offer no practical benefit and / or (b) be practically disadvantageous.

11.6 SUMMARY: DELIBERATION AND FREE WILL

At first glance, it appears that a person who says that there is no free will and who also deliberates has inconsistent beliefs. However, we saw that it is possible that no such inconsistency is present since the person does not really believe what they say. It is their actions that convey what they truly believe. However, the argument for this view is not without its problems for it is possible that their actions and their words both convey what they truly believe (and so their views are inconsistent).

SHOULD WE PUNISH CRIMINALS?

In a previous chapter, we saw that the denial of free will potentially conflicts with the act of deliberation. In this chapter, let's consider whether the denial of free will conflicts with the belief that certain people *deserve* punishment or praise.

12.1 RETRIBUTION

One intuition that people have is that bad people deserve punishment and good people deserve praise. Focusing on the former, let's define the retributive theory of punishment as follows:

Definition 12.1: retributive punishment

The theory of retributive punishment contends that (1) it is morally permissible to punish criminals because they *deserve* to be punished (even if punishing them will do no good), (2) it is morally impermissible to punish innocent people, and (3) the punishment of criminals should be proportionate to their wrongdoing.^a

^aThe retributive theory of punishment is typically accepted because (i) it is intuitive and (ii) problems with other theories of punishment that usually entail (a) punishing individuals who have not committed any wrong doing and/or (b) inflicting disproportionate punishments on criminals.

Let's clarify the retributive theory with a few points.

First, the retributive theory is ancient idea: "eye for an eye" (lex talionis).

Second, it is largely interpreted as the principle that the punishment needs to be *proportionate*, not necessarily equivalent. For example, if you kill my daughter, then it isn't the case that I need to kill your daughter. Rather, I am entitled to some appropriate response to the evil you have caused.

Third, retributive punishment is not about revenge. It is often taken to be a requirement of justice. It can be thought of as enacting a punishment that you personally deserve (what is owed to you). That is, you deserve to be punished because YOU did an action that DESERVES to be punished (we are not punishing you to achieve some goal).

The reason that the theory of retributive punishment needs to be given up is because if criminals are not *ultimately responsible* for their actions then they

we cannot say it is permissible to punish them because they deserve to be punished.

Related to this, the justification for retribution can be obtained by applying the criminal's own maxim for life to themselves. In other words, if the criminal lives a certain way (causing evil to others), then we apply that way of living to them. For example, suppose you kill Tek. You must be living by the maxim: it is morally permissible to kill Tek when it serves an individual's desires. We thus apply this maxim to you (we universalize it): it is morally permissible to execute anyone when it serves society's desires

The retributive theory of punishment is typically accepted for a few reasons:

1. it is intuitive: bad people deserve bad things to happen to them.
2. other theories of punishment are more problematic for they allow (1) punishing individuals who have not committed any wrong doing and/or (2) inflicting disproportionate punishments on criminals.
3. they are rooted in the concept of fairness: we are applying the rules of the evil-doer on themselves. For example, if the evil doer kills, then it is acceptable to kill the evil doer.

Exercise 12.83

The retributive theory of punishment requires that punishments be proportionate. State a bad action and state (1) a punishment that is not proportionate and (2) a punishment that is proportionate.

12.2 FREE WILL AND PUNISHMENT

Suppose someone does not believe in free will. At first glance, it appears that someone who does not believe in free will cannot consistently believe that bad actors *deserve* to be punishment. They do not deserve to be punished because they did not choose to act badly. Rather, there bad action was the result of cosmic chance (in an indeterministic world) or the laws of nature and the prior state of the world (in a deterministic world). Let's state this argument in argument standard form:

- P1: There is no free will.
- P2: If there is no free will, then persons are not responsible for their actions.
- P3: We punish / incarcerate criminals (bad actors deserve punishment).
- P4: A person only deserves to be punished iff they are responsible.
- C: Therefore, criminals do not deserve to be punished.

As we have already considered several theories that argue that there is free will, let's consider two responses that accept that there is no free will.

1. Accept the argument: There is no free will and therefore it is never morally acceptable to punish criminals.
2. Reject the argument: there is no free will but it is still morally permissible to punish criminals.

Let's focus our attention on the second of these two responses.

Exercise 12.84

Suppose it is not morally acceptable to punish criminals. They do not deserve to be punished. Their imprisonment is the same as punishing a person completely innocent of a crime (a crime they did not commit). Further suppose that all prisoners were set free. How would society need to be different in order to prevent total collapse?

12.2.1 Rejecting retribution

What we are looking for is a way to say (1) there is no free will and (2) it is still morally permissible to punish criminals. This is possible in two scenarios:

1. A person did not freely do action *A* but is still responsible for *A*
2. A person is not responsible for *A* but it is still morally permissible to punish them.



Let's focus on (2). We might argue that it is morally permissible to punish a person because it results in better social outcomes. That is, punishing people (even though they did not freely do their bad actions) is morally acceptable because this punishment increases the overall amount of good in the world. Let's consider three different approaches.

12.2.2 Deterrence

We might try to justify the punishment of criminals because it **deters** wrongdoers from committing similar wrongdoing in the future. Let's call this theory "deterrence theory".

Definition 12.2: deterrence theory

It is permissible to punish people because it deters them (or others) from committing wrongdoing in the future.

The justification for this form of punishment is found in the fact that penalties deter (condition/scare) future bad behavior. On this account, you are not being punished because you are responsible for the action. Instead, you are being punished to set an example and deter future behavior.

For example, suppose David sees a sign that reads "Pets Must Be On a Leash" (Fine up to \$300). Ignoring the sign, he defiantly takes his dog off a leash and has his dog run through the woods. David receives a ticket. In punishing David, the aim is to deter him from the action in the future.

Here is another example. Suppose Liz is thinking about fighting her neighbors. But discovers that doing so would come with incarceration. The punishment deters her from the action.

Deterrence theory has application outside of the criminal justice system. For example, parents use punishments to deter bad behavior (e.g. go to your room! Or you are grounded!). Schools use punishments to deter bad behavior (detention, school suspension, removal from sporting activity). In addition, employers use punishments to ensure good behavior (e.g. firing, docking pay, limiting hours).

In the case of deterrence, justification for punishment does not depend upon you being ultimately responsible for an act. Rather, punishment is designed to influence people to not commit the act again (what justifies the punishment is that IT DETERS!). Again, if David receives a fine of \$300, it does not matter if he wasn't free to act otherwise. The punishment is designed to scare / condition him from doing the act again.

Exercise 12.85

Create an example of a punishment that would deter people from doing an action in the first place or from doing an action again (it does not need to be drawn from the criminal justice system).

12.2.3 Rehabilitation

We might contend that it is not morally permissible to simply *harm* criminals by putting them to death or incarcerating them indefinitely. However, it is morally permissible to *incarcerate* criminals in the effort to **reform** or **rehabilitate** them so as to prevent future crime.

The next theory we will consider is the rehabilitation theory.

Definition 12.3: Rehabilitation theory

It is not morally permissible to punish criminals but it is morally permissible to “rehabilitate them” or “treat them” so they don’t commit future crimes

On the rehabilitation model, criminal behavior is treated as a type of sickness. Just as we treat sick people, we are justified in treating the disease of criminality. Just as we give people physical therapy when they have injured their knee or ankle, prison and various other forms of “punishment” are just methods for rehabilitating people so they act lawfully in the future.

Rehabilitation may involve locking them in a “room” (not cell) in a “treatment center” (not prison). We hospitalize seriously ill individuals (e.g. people with serious diseases or injuries). Again, on this model, it is not assumed that people are responsible for their illnesses (criminality).

For example, suppose David kills his neighbor. On the rehabilitation model, David is sick / diseased. It is necessary to incarcerate (Oops! I mean “hospitalize”) him in order to “treat” David’s disease.

Exercise 12.86

While the rehabilitation model may make sense in theory, can you give some examples of how the theory does not live up to its goals or aims? In other words, suppose our society is operating with the rehabilitation approach. How are we failing to meet these goals?

12.2.4 Quarantine

We might contend that it is morally permissible to *incarcerate* them because they are **diseased**. That is, rather than viewing a prison sentence as a form of punishment, it is viewed as a form of **quarantine**. The rationale here is that just as we quarantine individuals with deadly viruses or contagious diseases, if we view criminal behavior as a kind of disease, then incarcerating

criminals is an effort to protect criminals from from hurting others.

The next non-retributive theory we will consider is the quarantine theory.

Definition 12.4: Quarantine theory

it is not morally permissible to punish criminals but it is morally permissible to quarantine criminals in the effort to protect society from them

Definition 12.5: Quarantine

to restrict the movement of an individual, group of individuals, or some item in order to prevent the spread of a disease, virus, or pest.

Let's illustrate this theory with an example. Suppose Tek contracts a deadly virus. This virus can harm other individuals. To prevent the harmful effects of this virus, we might quarantine Tek (remove him from society until we are confident that the harmful effects of the virus have passed). The justification for quarantining Tek does not depend upon him being ultimately responsible for contracting the virus. Rather, the justification for quarantining is rooted in protecting others from harm or increasing social good.

Here is another example. Suppose Jon kills his neighbor. He is likely to strike again. It is necessary to incarcerate him (a form of quarantine) to prevent him from harming other individuals.

Quarantine theory is similar to rehabilitation theory in that it can be understood as treating criminality as a type of disease that needs to be managed either by healing the patient (rehabilitation) or by isolating them until it is safe (quarantine)

Recall that the argument from punishment says if we accept HD then there is no moral justification for the incarceration of criminals (we have to give up on retributive theory of punishment). But we have shown that there are other forms of punishment that do not require assigning responsibility (non-retributive theories).

Exercise 12.87

1. Which one of the non-retributive theories do you find most compelling and why?
2. Which one of the non-retributive theories do you find the most worst and why?

12.2.5 *Problems with non-retributive theories*

In previous sections, we saw that someone who denies the existence of free will might argue that we can still punish criminals for bad actions. They might argue that punishment is justified on non-retributive grounds: Deterrence theory, Rehabilitation theory, or Quarantine theory. In this section, let's consider some general problem with these approaches.

Non-retributive theories of punishment are said to entail the following:

1. Non-retributive-punishment principle (NRPP): it is morally permissible to punish someone regardless of whether the individual is responsible for the crime (HD: not free / responsible; non-retributive: ok to punish)
2. Utility principle: morally justification for the punishment is due to it increasing overall good in the world (not because we need to punish the individual)

Critics contend that these non-retributive theories of punishment are unacceptable. There are two key problems:

- Problem 1: they allow for punishing people who do not deserve to be punished (e.g., innocent people).
- Problem 2: allow for disproportionate forms of punishment in either the form of too extreme a form of punishment or no punishment at all.

Let's consider both of these problems. The first problem is that they permit punishing innocent people. NRPP says it is permissible to punish people regardless of whether they were responsible for the act. Therefore, it is permissible to punish people to increase social good

Let's consider an example involving deterrence theory. Suppose Jon kills Vic but the murder is unsolved. Next, Suppose David has nothing to do with the murder. While David is not responsible, punishing David would be morally permissible as it would deter others from committing such crimes (increase in social good). Deterring others from murder makes society better off.

Next, let's consider an example involving the rehabilitation theory. Suppose David hasn't committed any crime but he is likely to commit a crime in the future. He has "the disease" but isn't showing full-blown "symptoms" yet. Therefore, we can incarcerate him before he actually commits a crime.

Punishing a person from a crime that they will commit (or are likely to commit) but have not committed is sometimes called "precrime". The notion of was coined by Philip K. Dick in his short story "The Minority Report" (later

made into a movie) where individuals are arrested before they commit a crime. This idea has its roots in the 19th century idea that some people are “born criminals” and that you could predict that they would commit crimes (rehabilitate them before they do commit those crimes). This could be determined using “tests” for delinquency or crime prediction algorithms.

Finally, let's consider an example involving quarantine theory. Again, suppose Jon kills Vic, the murder is unsolved, and the murder flees the country. Society is enraged because they think David is guilty of the crime (he isn't) and would be terrified if he was free. While David is not responsible, punishing David by removing him from society (quarantine) might benefit society: quell their rage.

This sort of example is sometimes seen in “cop” shows where there is pressure from the public to “solve the crime” aka “put someone behind bars” even if they didn't commit the crime. Typically, the public (the mob) is acting crazy and so the police reason that it is better to lock someone up to calm the public than to let the public believe that a murderer is still on the loose.

Objection 2 In addition to punishing the innocent, there is a second problem with non-rehabilitative theories. The second problem is that they allow for disproportionate forms of punishment in either the form of too extreme a form of punishment or no punishment at all. NRPP says it is permissible to punish people regardless of whether they were responsible for the act. Therefore, it is permissible to punish people to increase social good.

For example, consider amputation as a form of punishment. It is reported that in Sharia-controlled parts of Nigeria and Saudi Arabia (and in Iran) individuals have hands / legs amputated for theft. Or, consider execution. Roughly 53 countries out of 195 countries use the death penalty. Finally, while most countries are “against torture” some nevertheless use it to punish or extract information (regardless of its efficacy).

If amputations, death penalty, or torture deter people from certain crimes then deterrence theory would say they are morally permissible.

It is often thought that drug possession laws are too strict. Why are those so disproportionate? One possible answer is that the punishment is not designed to punish people for doing bad behavior but instead to deter people from taking up bad health habits.

Let's consider an example involving the rehabilitation theory. Suppose Socrates kills his neighbor but we are 100 percent sure that he will not commit any future crimes. The rehabilitation theory says that we are morally permitted to incarcerate Socrates in order to rehabilitate him. But if we know Socrates

won't commit future crimes then he is already fully rehabilitated. Therefore, there are cases where a person can murder and not spend a day in prison! Again, punishment seems disproportionate.

In this example, Socrates seems to deserve punishment but the non-retributive (rehabilitation) theory entails he does not need to be punished at all!

Exercise 12.88

If hard determinism is true, then it seems like we need a non-retributive theory of punishment. But, non-retributive theories have all sorts of problems. What can we do in this situation?

Part IV

Time

WHAT DO WE BELIEVE ABOUT TIME?

13.1 INTUITIONS ABOUT TIME

In the previous section, we considered how to do metaphysics. In this section, we will start our investigation into metaphysics by looking at our first topic: time and persistence. Our approach will be to start with our intuitions and then formulate theories of time.

In this section, we will consider a long list of intuitions that people have about time. Since this list is rather long, we won't consider all of these intuitions in great detail. Instead, in the following section, we will isolate some intuitions that will be relevant for formulating our theories of time.

1. **What time is:** Time either just is clock time (what a clock says) or what a clock is supposed to measure (e.g., "What is the time? Ah! It is 11 o'clock.")
2. **Reality of Time:** Time is real. It is not an illusion.
3. **Value:** Time is valuable and important to human life, the economy, and society (e.g., "time is money")
4. **Movement:** Time moves, passes, or is dynamic. The event that is the present is constantly changing. Time is not frozen.
5. **Duration of the Present:** It is not clear how long the present moment is, but it is not very long.
6. **Experience of the Present:** We only experience the present moment. No one is experiencing the future or the past.
7. **Rate of the movement of time:** Time flows uniformly although our experience of the flow can vary (e.g., clocks tick at regular increments but time can fly when you are having fun).
8. **Directionality:** Time flows in one direction, toward the future ("arrow of time", "time is a river"). Events go from being future, to being, present, to being past.
9. **Losing Time:** Time is something you lose and you cannot get it back ("time is a thief you cannot banish")
10. **Event Importance:** Certain times are more important than other to humans (e.g., birthdays, anniversaries, holidays, the start of a new year)
11. **Time and Causation:** Time is important for causation. This is because causes always occur before their effects.
12. **Time and Free Will:** If we are free, then our free actions take place in some time where we can exert some control.
13. **One Dimension to Time:** In contrast to space (L, W, H), there is only

one dimension to time (only one coordinate to represent its position: it is like a line)

14. **Past, Present, Future:** Conceptually and linguistically, there is a ``past'', ``present'', ``future'', and these concepts are exclusive and exhaustive.
15. **Rationality of past, present, future:** People believe that the *distinction* between past, present, future is a rational distinction. There is evidence / justification to support this distinction. The point here is that we have some intuitions that we might take to be irrational, but this distinction is rational. For example, I may have the intuition that I'm not an expert in philosophy (imposter syndrome), but also know that this intuition is irrational. Another example, I may have the intuition that I'll fall or am being pulled off of a building (fear of heights), but I know this intuition is irrational.
16. **Present is Metaphysically Special:** The present is metaphysically special in some way. It is somehow more real than the past or future. The present simply exists (exists simpliciter) while the past and future have qualified existence (the past *did exist*, while the future *will exist*).
17. **Temporal Bias:** The importance of an event can be impacted solely based on when it happens. For example, for human life, events that happen in the present are more important than those that happen in the future or past, and events that happen in the future are more important than those that happen in the past (Present > Future > Past).
18. **Time Travel:** Time travel is metaphysically possible but currently not a technical reality. By implication, everyone here is in the present. No one is from the future or past.
19. **Endurance:** Objects persist through time. Essentially, I move through time like a car moves through space.
20. **Origin:** There may be time before the origin of things (before there were things).

Exercise 13.89

Pick one of the folk intuitions of time listed in this section. Second, articulate this intuition in your own words. Third, illustrate this intuition with an example.

13.2 THE KEY INTUITIONS

In this section, we will consider several

13.2.1 *Intuition 1: Past, present, and future*

The first intuition is that the past, present, and future are mutually exclusive and jointly exhaustive of events in time.

The categories of past, present, and future are mutually exclusive in that if an event falls into one of these categories, it does not fall into the other two. In other words, no event is both present and past, or past and future, or present and future. More concretely, to say that “the baseball game happened in the past” is to say that it did not occur in the present or the future.

These categories are also jointly exhaustive since every time can be characterized as either past, present, or future. There is no other temporal property had by events. , e.g. superfuture, hyperpast.

13.2.2 *Intuition 2: Our belief in past, present, future is rational*

The second intuition is an intuition about the first intuition. This intuition is that the first intuition is rational: supported by evidence. That is, with respect to time, what is being said is not only do events have exactly one temporal property (past, present, and future), but also that this belief is (unlike other beliefs we may have) rational.

To avoid confusing these two intuitions, it is helpful to consider intuitions we have that we know are irrational. For example, some people suffer from *imposter syndrome*. That is, there are people who are experts in a field but believe themselves not to be experts (imposters, frauds). In this case, their intuition (that I am not an expert in my field) is not supported by evidence and is irrational. Another example are people who suffer from the *Dunning-Kruger effect*. These individuals have, after initially learning a minor amount of some subject matter, the intuition that they are geniuses in that subject area. This intuition is irrational since their “expertise” is not supported by the evidence.

Exercise 13.90

What are some other intuitions (common-sense or widespread beliefs) people have but we know them to be irrational?

If the second intuition is that the first intuition is supported by evidence, one question we might briefly consider is “what evidence is there that events in the world correspond to our categories of past, present, and future”? We might say that things just seem that way, or that is how things look, but what other evidence is there? At this point, we will not consider all of the

evidence. Instead, let's look at one piece of evidence, which relates to what we can and cannot control.

Consider that our perceived scope of our control is limited by an object's position in time. We believe cannot change the past. Since Caesar crossed the Rubicon in 49 BC, we believe that we cannot change this fact. On the other hand, we believe we can change the future. If there is an upcoming local election, we believe we can influence who will win by voting and trying to convince others to vote in the same way we do. In short, the scope of our control differs with respect to an event's position in time. If it is in the past, we cannot control it. If it is in the future, then we can control it.

Now consider the following argument. If the past, present, and future were merely conceptual distinctions (categories in our minds) and not part of the fabric of reality, then the reason we cannot control past events is because we have *categorized* an event as *being in the past*. If this were the case, then presumably if we were to view Caesar crossing the Rubicon as *being in the future*, then we could exert some control over that event. But, this is not the case. This event has already occurred and we cannot influence it at all (unless we can time travel). So, this suggests that the past, present, and future are not mere conceptual distinctions. They are not merely in our minds but are part of reality itself.

- P1: The scope of our control differs with respect to an event's position in time: whether it is past, present or future.
- P2: If past, present, and future were mere fictions, then what we can and cannot control would be determined by our minds not reality.
- P3: Reality limits what we can and cannot control.
- C: Therefore, there is a real difference between an event being past, present, and future: events really are past, present, or future.

Exercise 13.91

What other evidence can you think of that indicates that events really are past, present, and future?

13.2.3 Intuition 3: There is something special about the present

A third intuition is that events in the present are somehow *more real* than events in the past or future. Philosophers will sometimes capture this intuition by saying that the present moment is "metaphysically special". Another way this is expressed is to say that present events exist *simpliciter*.

What this means is that the present exists without any qualification required.

This contrasts with events in the past and the future which exist with qualification. That is, we say that an event in the past *did exist* (we qualify it by using past tense).

To understand this more clearly, let's consider some examples. When it comes to objects or events in time, we qualify their existence in some way. Consider Bigfoot (Sasquatch). Bigfoot is a large, creature (slightly larger than an adult human) that is said to occupy the forests of the Pacific Northwest. Belief in Bigfoot is largely supported by (1) people who have claimed to spot him, (2) dubious forensic evidence, e.g., strange clumps of hair or casts of footprints, and (3) blurry video footage. However, it is generally believed that Bigfoot does not exist.

Nevertheless, while most people would say that Bigfoot does not exist, they might also say that Bigfoot *could* exist or *might* exist. Or, they might say that Bigfoot exists *in a possible world* or exists in the *the minds of conspiracy theorists*. In these cases, Bigfoot's existence is qualified.

In contrast, consider your own existence. If you were asked whether you exist, you would say "yes". There is no need to say how you exist or where you exist or in what mode you exist. You just exist. Or, to use the philosophical language "you exist *simpliciter*", without the need for any further qualification.

It is claimed that we have an intuition that the events in the present exist *simpliciter*, while those in the past or future exist with qualification.

1. Events in the past did exist (were real) — we qualify their existence
2. Events in the future will exist (will be real) — we qualify their existence
3. Events in the present exist *simpliciter* — they exist with no qualification

Exercise 13.92

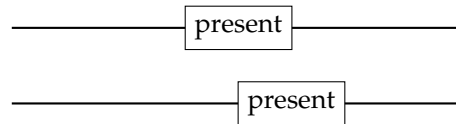
1. Do you think that there is anything metaphysical special about the present moment? That is, is the present moment's reality (or existence) somehow *more real* than events in the past or future. How so?
2. Create an example of something that exists *simpliciter* and something else that exists with qualification.

13.2.4 Intuition 4: The present moves

The fourth intuition is that time passes. This intuition is often expressed in several different ways. Here are a few of them:

1. The present moment is moving
2. The present is constantly changing
3. There is real change in the world that relates to the passage of time
4. Time is not fixed: we are not frozen in time.

The fact that time passes is often expressed by how we talk about the world. For example, consider a typical day:



To illustrate, consider your day:

1. You wake up - This is the present.
2. You brush your teeth - This is now the present.
3. You get dressed - This is now the present.
4. You go to metaphysics class - This is now the present.

What changes in the above? It seems as though what changes is not the events themselves but the temporal property each event has. That is, what changes about *going to metaphysics class* is that it is in the future, then the present, and then the past.

Exercise 13.93

Some philosophers believe that time pass even though nothing in the world changes. For example, imagine a universe where nothing is changing but time is marching forward. Other argue that time and change are intricately linked together. That is, if the universe were completely frozen, time would not pass. Which one of these views is correct?

Let's construct a few theories of time. To do this, we will articulate some concepts necessary for stating our theories of time.

14.1 TIME AND CLOCKS

To state a theory of time, we need a few working definitions.

Let's begin with the definition of a "clock". A clock is any physical system with a repeating, periodic process that is used to mark the simultaneity of events. Examples of clocks include digital clocks, a swinging pendulum, a bouncing beam of light, or vibrating atoms.

A few points of clarification. First, a clock is not some magical device that is used to tap into the invisible flow of time that applies universally. Instead, it is a device used to mark the *simultaneity of events*.

For example, suppose you have a watch that reads "6:59". What does this tell you? It does not mark some invisible universal flow of time. It is simply a visible, physical event that occurs in the universe at a specific location. What then is "6:59" used for? It is used to mark when other events are occurring. So, when you say "the train arrived at 6:59", what you are saying is "the train arrived at the same time (simultaneous with) my watch reading 6:59".

What is "a time"? Philosophers typically define "a time" as follows:

Definition 14.1: a time

A *time* is what is described by a *complete* and *consistent* specification (snapshot, full description, picture, account) of a world.

Let's unpack this definition. First, a time is what is described by a *complete* description. For example, say we had a description of the location of every object in the universe at exactly 12:01PM EST. What is described by this description is a time. In contrast, if the description did not include where I am or how tall I am, then this would not be a time.

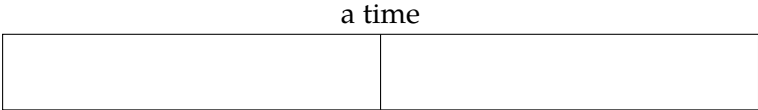
Second, a time is what is described not only by a consistent description of the world but a *consistent* description. It is inconsistent to say that "I am six-foot tall and I am not six-foot tall." Equally so, it is inconsistent to say that I am in New York City and in Paris. Since what is described is not consistent (I can't be wholly located in both places), what is described by my description of the world is not a time.

Third, if we have two complete but inconsistent descriptions of our world, then these descriptions would correspond to different times. That is, suppose there are two complete descriptions (D_1 and D_2) of our world. In addition, suppose D_1 and D_2 are inconsistent with each other. For example, suppose D_1 says Tek is in New York while D_2 says Tek is in Chicago. If this is the case, then these descriptions correspond to two different times. That is, since $D_1 \neq D_2$, it follows that $D_1 = t_1$ and $D_2 = t_2$ and $t_1 \neq t_2$.

Fourth, the definition of a time does not imply that some complete and consistent description of the world exists simpliciter. For example, suppose I give a complete and consistent description of 1776. The object (world) that corresponds to this description would be a time but there is no claim that this time exists, e.g., I might say that the time *did exist* or exists with certain qualifications. Similarly, if I gave a complete and consistent description of what *could have happened but did not*, I would be giving an account of a time that exists in some possible world.

Fifth, from the definition of time, there are no assumptions made about whether time is linear, how it is ordered, whether time exists, whether time passes, whether there is a beginning or end to time, what objects exist in time, or how we perceive time.

Sixth, since we can give multiple “complete and consistent specifications” and we take our world to consist of more than one time, it seems reasonable to assume that the entire history of our world can be thought of all of the times $t_1, t_2, t_3, \dots$ that ever did exist, do exist, or will exist. Each of these times can be thought of as a three-dimensional “slice” of the universe and the total history of the world can be thought of as a “block” composed of these different times.



Exercise 14.94

Suppose our world consists of times. The definition of a time does not imply any order or flow. Create a world where the order or flow of times is much different than our own.

14.2 THREE SUB-THEORIES OF A THEORY OF TIME

Now that we have a definition of a time and are assuming the hypothesis that there are multiple times, let's compose a theory of time. A theory of

time consists of at least three main ingredients (or subtheories):

1. **Theory of temporal ordering:** States how times (events) are ordered. This concerns the directionality of time and how we characterize the events in time (specify their position).
2. **Theory of temporal ontology:** States what times (events) exist. This concerns the relation between an object being in time and the object existing. Theory about how an event's position in time determines whether that event is real (e.g., are past events real?)
3. **Theory of temporal passage:** States whether time passes. Theory about whether there is real change in the world.

14.2.1 *Theory of temporal ordering*

A set is defined by the elements in the set, not the order or structure of the set. In a set, the *order* of the members of the set does not matter. Take the set A that has the members $\{1, 2, 3\}$. The position of members of the set do not matter for defining the set. For example, all of the following sets are the same set as A :

$$A = \{3, 2, 1\} = \{3, 1, 2\} = \{2, 1, 3\}$$

But time is said to have an **order** or **structure**, and events (times) have *position* relative to events. This order is said to be **objective** feature of the world and thus not the result of how we see the world or our subjective whims. For example, events in time are not ordered in a sequence from (1) my favorite to (2) my least favorite.

Here is another way of thinking about temporal ordering. Suppose we believe our world has multiple distinct times: t_1, t_2, t_3 . That is, these descriptions accurately describe *our* world. If there are multiple times, this means that there are multiple, complete and consistent accounts of the world. But notice that these descriptions are *inconsistent* with each other. In other words, what they are describing exists but not at the same time.

For example, suppose at t_1 , there is a penguin in my room. And, at t_2 , there is not.



Surely the penguin cannot both exist and not exist, but the penguin can exist at one time but not at another.

If multiple times exist, then we might ask about their relation to each other. That is, the entirety of the world is not a single time t_i but all of the times: $w = t_1, t_2, t_3$. Are these times arranged in any particular way? Surely, they are not arranged based on subjective considerations, e.g., from times I like the most to times I like the least. Instead, if there is an order to these times, that order is independent of my will. Let's call the objective arrangement or ordering of times, a world's "temporal order".

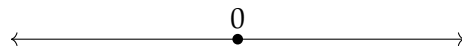
A *dimension* is the minimum number of coordinates needed to specify a point (something). As a construct, a dimension is a system used to specify the location or position of something. For example, space is often said to have three dimensions. With respect to the spatial position of an object, we can specify the position of the object using three different coordinates (x, y, z). If you have ever been asked to determine the location of a point on a piece of paper, then you made use of two dimensions: the x and y dimensions.

Time is also said to be a *dimension*. When making this claim, what is being asserted is that it has a structure that allows us to specify the position or location of events relative to other events. Unlike space, which is said to have three dimensions, time is said to have one dimension as events can be specified as occurring earlier or later than other events. However, the exact nature of this dimension is debated. That is, it is not clear how, objectively speaking, various times / events are ordered. The question then of temporal ordering is thus asking how we go about deciding how to specify the position of various events or objects in time.

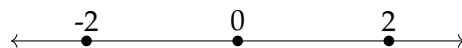
The temporal order of the world is a series. There are two main types of series: the A-series and B-series.

14.2.1.1 A-Series

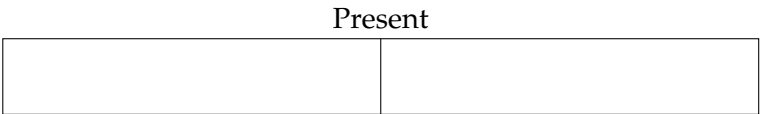
Let's start with an analogy. Take a horizontal line. Now specify a special point (0) on that line.



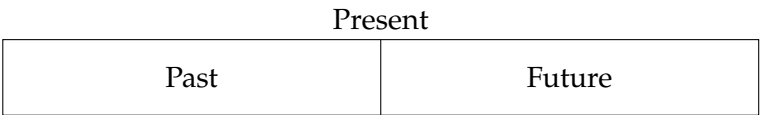
With this point, we can talk about other points on the line relative to point 0. For example, point 2 is 2 units to the right of 0, or -2 is 2 units to the left of 0.



We can do the same thing with respect to time. Take the entire history of the world. One of these times seems unique because it, unlike any other, is the only time that really exists (exists simpliciter). This is the present. Once we have the present moment, we can use it to specify the position of every other moment in time (the entire history of the world).



In broad strokes, we can say that times later than the present are in the future and times earlier than the present are in the past.



We can also specify how far times are away from the present using various other ways of measuring time.

- E0 is two days ago.
- E2 is three days, five hours, and two seconds in the future

When time is structured by picking out some absolute position, identifying it as the present moment (technically, it can be any absolute position), and then identifying the position of other times based on its distance from that absolute position, one is characterizing time in what is called the **A-series**. As Power[102, p. 32] puts it, "[t]he main feature of the A-series is that events in it are either positional in the past, present, or future, or are some determinate of those temporal positions."

Let's define the A-series of time as follows:

Definition 14.2: A-series

The A-series structures time in terms of the past, present, and future.

It is worth emphasizing that there is a *single* present rather than multiple present moments. Saying that there are "multiple presents" is, at least at first glance, inconsistent. To illustrate, suppose there are three present moments in a single A-series. If there are three distinct present moments in the A-series, then one present occurs later than another (and so is in the future). This would mean that the later "present" is both present (by supposition) and future (by the fact that it is later than some other present).

Now that we have examined the A-series, let's examine the B-series.

14.2.1.2 B-Series

In the A-series, a special time was picked out ("the present") and the other times were ordered based on their distance from (or relation to) that special time.

Another way to order time is solely through relations. If time is **relationally structured** such that some events (times) occur before other events, then we can always specify the *relative* position of two events $E1$ and $E2$. That is, we can say things like:

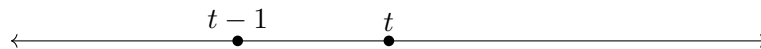
- $E1$ occurs before $E2$, or
- $E1$ is earlier than $E2$, or
- $E1$ is fifty years before $E2$.

For example, for two events $E1$ and $E2$, we might say that $E1$ is in the present, or $E2$ is in the future, or $E1$ occurs before $E2$ and so on.

Definition 14.3: B-series

The B-series structures time in terms of relations: earlier than, later than, simultaneous with

The idea here is that we can take any two points in time we wish. Call the first point t . With respect to the second time, that point will either be earlier than t and so we might label it as $t - 1$, later than t and so we might label it as $t + 1$ or simultaneous with t , in which case it is just t .



Now that we have examined the A-series and B-series, let's specify what it means for a theory to be an A-theory or B-theory.

14.2.1.3 A-theory or B-theory

Now that the A-series and B-series have been explained, let's state the two main theories of temporal ordering: the A-theory and B-theory.

Definition 14.4: A-theory

The world is temporally ordered via the A-series and B-series.

Definition 14.5: B-theory

The world is temporally ordered via the B-series.

An **A-theory** of time asserts that events are organized by the A-series and B-series. This means the world really has a present, past, and future (time is organized around a present moment) and events can be said to be earlier than or later than other events.



Let's consider a concrete example of the A-theory. Take three events:

- t1: July 16, 1945 (5:29 AM) - Testing of a nuclear weapon at the Trinity Site, New Mexico
- t2: July 21, 1969 (2:56 UTC) - Neil Armstrong steps foot on the moon
- t3: Sept 5, 2101 (8:56 UTC) - Alien life arrives on earth

The A-theorist accepts that events are ordered in terms of the B-series and the A-series. If we take a random point in time, we can specify those times that occur earlier and later than that time. However, in addition, we can also specify the distance of that time from the present moment.

A **B-theory** of time asserts that events are only organized by B-relations. This means that there is no such thing as an objective present, objective past, or objective future. Instead, there are events that are earlier than or later than or relatively simultaneous with another.



With respect to the concrete events above, the B-theorist will say that t1 is earlier than t2, and t2 is earlier than t3 but they will not say any of these events are present, past, or future.

Exercise 14.95

1. State whether you are an A-theorist or B-theorist, neither, or undecided.

14.2.2 Part 2: Theory of temporal ontology

A theory of temporal ontology refers to what times exist. I may have a complete and consistent description of the world, but whether or not something corresponds to that description depends upon the world itself. I may describe a world where there are flying cars and talking robots, but whether

there is anything that corresponds to that description is a matter for the nature of the world.

There are many possible theories of temporal ontology. In this section, let's consider the most popular ones.

1. Presentist
2. Past-presentist
3. Eternalist

In what follows, we will articulate each of these theories.

14.2.2.1 Presentist

Here is the presentist theory of temporal ontology:

Definition 14.6: Presentist temporal ontology

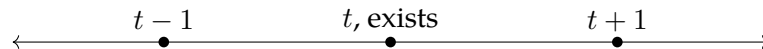
The presentist theory of temporal ontology states only present times (events) exist.

This theory is sometimes expressed as follows:

- Reality is exhausted by what exists in the present
- Only present events are objectively real
- An object exists in time if and only if it is in the present moment.

Here is an example. We know that George Washington was the 1st President of the United States. At some point, he probably ate an apple. We also know that an event is happening the moment you are reading this sentence. According to the presentist theory of ontology, only the latter moment exists. The presentist might say that the former moment *did exist* but only the present moment exists *simpliciter* (without qualification). Similarly, if in 2030, you will be sitting in jail and waiting your trial. That event *will exist* but does not exist *simpliciter*.

In other words, if t is the present moment, then events earlier $t - 1$ than the present (past events) and later $t + 1$ than the present (future events) do not exist.



Here is another way to think of the presentist theory of temporal ontology. Imagine that the universe is a rectangle block with each vertical slice being all of the events happening at a particular time. The presentist theory says that the only time that exists is one of those vertical slices.

14.2.2.2 Past presentist

The next theory of temporal ontology we will consider is the past-presentist theory of temporal ontology:

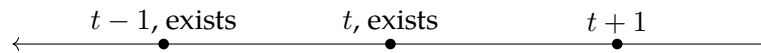
Definition 14.7: Past-presentist temporal ontology

Past-presentist theory of temporal ontology states that only past and present events exist.

Here are some more compact ways of expressing this position:

- Reality is exhausted by what exists in the past and the present
- Past and present events are objectively real
- Whether or not an event exists depends upon its position in time since only past and present events exist.

In other words, if t is the present moment, then earlier events t_1 exists but later events $t + 1$ do not exist.



Here is an example. Again, take three events:

1. George Washington eating an apple (in 1776).
2. The event that is happening right now (the present).
3. You are sitting in jail (in 2030)

According to the past-presentist, events (1) and (2) exist but event (3) does not exist. The past-presentist might say that (3) will exist, but only (1) and (2) exist *simpliciter* (without qualification).

Here is another way to think of the past-presentist theory of temporal ontology. Again, imagine that the universe is a rectangle block with each vertical slice being all of the events happening at a particular time. The past-presentist theory says that what exists is a certain chunk of that block, starting from the slice that represents "the present" and all times before that slice.

14.2.2.3 Eternalist

Here is the eternalist theory of temporal ontology (also known as *permanen-*
tism and *event everywhereism*):

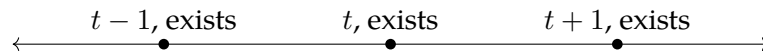
Definition 14.8: eternalist temporal ontology

The eternalist theory of ontology says that all events in time exist in the same way.

This means that temporal position does not condition an event's existence. Here are some other ways of putting the eternalist position:

- Everything that ever existed, exists, or will exist exists eternally.
- Whether something exists right now does not matter to whether it exists.
- If the world began in 0 AD and ends in 3000 AD, then each event between 0 and 3000 AD exists in the same way.

A couple points of clarification. First, the eternalist's theory goes by several definition names: **permanentism** and **event everywhereism**. These names can be helpful for understanding the position. In calling the view “permanentism”, it emphasizes that insofar as an event in time exists, it exists permanently. In calling the view “event everywhereism”, it emphasizes the connection time has to space: just as every event in space exists equally, every event in time exists equally.



Second, for the eternalist, an event's location in time does not matter for determining whether that event exists. If an event is in time, then it exists, regardless of location. Consider again these three events:

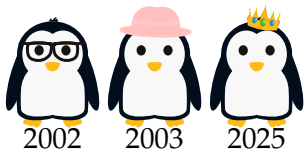
- t1: July 16, 1945 (5:29 AM) - Testing of a nuclear weapon at the Trinity Site, New Mexico
- t2: July 21, 1969 (2:56 UTC) - Neil Armstrong steps foot on the moon
- t3: Sept 5, 2101 (8:56 UTC) - Alien life arrives on earth

For the eternalist, all three of these events exist. There is still the testing of a nuclear weapon at the Trinity Site, New Mexico, Neil Armstrong is still stepping onto moon, and Alien life is arriving on earth.

Third, for the eternalist, an event exists simpliciter (without qualification). What this means is that we do not say that the *testing of a nuclear weapon at the Trinity Site* did exist or it had existed. Instead, the eternalist says this event exists (full stop, without any further qualification).

The eternalist theory of temporal ontology is likely the hardest theory for people to wrap their heads around, so let's consider another example. Suppose you were born in 2002. Let's take 2003 as your first birthday. Now let's

take 2022 as what we call “now”. For the eternalist, both events exist. Just because 2022 is “now” does not make it any more real than the event that occurred in 2003. Similarly, let’s say in 2030, you are in jail and waiting for your trial. That event exists in the same way that events happening “now” exist.



Fourth, eternalists are sometimes called “spatializers”. This is because they treat objects and events in time similar to objects that exist in space. Consider a road that stretches from New York to Los Angeles. The eternalist can be seen as saying that time is analogous to space in that just as the road can be exist as something that is spread out in space, an object can exist by being “spread out” in time.

Here is another way to think of the eternalist of temporal ontology. Again, imagine that the universe is a rectangle block with each vertical slice being all of the events happening at a particular time. The eternalist theory says that what exists is the entire block.

Three theories of temporal ontology (review)

Let’s make one final point with respect to the theories of time. Each theory attempts to answer the following question: Does an event’s position (location) in time condition whether that event is real (exists)?

The eternalist thinks that an event’s position in time is irrelevant to whether an event exists or not. An object’s location so-called past, present, or future has no bearing on whether or not that object exists. If it exists in time, then it exists just like every other object that exists in time.

In contrast, the past-presentist and the presentist contend that an object’s position in time is relevant to whether it exists. For the presentist, if an object does not exist in the objective present, then it does not exist. Similarly, for the past-presentist, if an object does not exist in either the past or the present, then it does not exist.

	eternalist	presentist	past-presentist
position in time matters?	N	Y	Y
all times exist	Y	N	N

	eternalist	presentist	past-presentist
only present exists	N	Y	N
past and present exists	N	N	Y

Exercise 14.96

1. Which theory of temporal ontology do you accept?
2. We only considered the three most plausible theories of temporal ontology. Construct another theory.

14.2.2.4 Part 3: Theory of temporal passage

As noted, a theory of time involves three components: (1) whether it is an A-theory or B-theory, (2) a theory of temporal ontology, and (3) a theory of temporal passage. The third component of a theory of time is a theory of temporal passage. A theory of temporal passage gives an account of whether "time passes". What it means for time to pass is unclear and so the question of temporal passage is usually articulated in terms of whether or not there is *change* in the world. The idea behind this is that if no change occurred in the world, then the state of the world would correspond to the same time (the same complete and consistent description of the world). As such, the question of temporal passage can be formulated as follows: Is there change in the world?

In short, a theory of temporal passage attempts to answer the following questions

1. whether there is temporal change in the world
2. whether the present moves
3. whether time passes

There are two main answers to the theory of temporal passage:

1. Dynamic answers: Yes, there is change (time does pass)
2. Static answers: No, there is change (time does not pass)

A **dynamic theory** of temporal passage asserts:

Definition 14.9: dynamic theory of temporal passage

1. Temporal passage involves real change
2. Change is due to the objective reality of a moving present

There are two dynamic theories of temporal passage:

1. Property change dynamic theory
2. Ontological change dynamic theory

One key claim made by *both* dynamic theories is that temporal passage (the passing of time, change) is a *feature of reality*. What this means is that change is not an illusion, subjective, or perspectival: not simply do to how we conceptualize or perceive the world. Change in the world is due to things *actually* happening. The difference between the two dynamic theories concerns their differing accounts of what kind of change is happening.

According to the property change dynamic theory:

Definition 14.10: property change dynamic theory of temporal passage

1. There is real change in the world
2. Change comes about by times / events changing their *temporal properties*
3. The passage of time is a change to a time's temporal properties

According to the property change dynamic theory, when we say "the present moves", what we mean is that "times" gain and lose temporal properties (being present, being past, being future). To illustrate, suppose there is a white wall, but then I paint it red. The property of the wall changed from white to red. On this view, there are no new objects being introduced or taken out of the world: the only thing that is changing are the properties that those objects have. However, in the case of temporal passage, what changes are the *temporal properties* of an event.

Let's briefly explain what it means to change an event's temporal properties. Suppose you are hungry and are going to eat a sandwich later in the day. We can say the event *you eating the sandwich* has the property of being in the future. While you are eating the sandwich that event no longer has the property of *being in the future* and instead has the property of *being present*. Notice that there is a change to the temporal properties: the event change from *being future* to *being present*. After you finish eating the sandwich, the event no longer has the property of *being present* but instead has the property of *being past*. Again, there is a real change to events. The event of you eating a sandwich went from *being future* to *being present* to *being past*.

On this account, there was a change in the world that occurred. It was a change to the event's temporal properties. The property dynamic theory of temporal passage says that "the present moves" by events/times gaining and

loses the property of A-properties (e.g., being future, being present, being past). On the property change theory, change occurs (time flows) because events are gaining or losing temporal properties.

Next, let's consider the ontological change dynamic theory. On this theory,

Definition 14.11: ontological change dynamic theory of temporal passage

1. There is real change in the world
2. Change comes about through objects coming in (or going out) of existence (ontological change) or those objects gaining or losing certain properties that are not purely temporal properties.
3. The passage of time is this coming in (or going out) of existence (ontological change)

On this theory, when we say "the present moves", what we mean is that "new things (or properties) are added to or subtracted from what exists". The passing of time is nothing more than the creation of (or destruction of) events.

According to the ontological change dynamic theory, temporal passage involves real change to what exists in the world in that what exists changes with the moving present moment. Let's consider two examples:

- t_1 : Baby Davey does not exist.
- t_2 : Baby Davey is born and now exists.
- t_3 : David (previously 'Baby Davey') is now old (still exists).
- t_4 : David dies. David no longer exists.

A similar example can be created for artefacts:

- t_1 : A lump of clay exists, but a statue made of the lump does not.
- t_2 : Liz uses the clay and creates a statue (the clay and the statue exists).
- t_3 : David criticizes the statue and Liz feels bad (the clay and statue still exists).
- t_4 : Liz smashes her statue (the clay exists but the statue does not).

Notice that in the above examples, objects go in and out of exists. In addition, new properties of things are going in and out of exists. For example, David had the property of *being a baby* and then that property went out of existence when David took on the property of *being an adult*.

The Ontological change dynamic theory says that change occurs when e2 becomes present since when e2 comes into existence, there is a change to the number or type of events that exist in the world. We account for the

passage of time by events not by events gaining and losing the property of being present but by new events coming into existence.

So far, we have considered two theories that assert that change is real. Insofar as the reality of change accounts for the movement of time, both of these theories assert that time really does pass. However, one theory of temporal passage still remains. This is the *static theory of temporal passage*.

Definition 14.12: static theory of temporal passage

1. There is no change in the world (this is a perspectival fact)
2. Real change does **not** occur either by times gaining or losing temporal properties or by new times coming into or going out of existence (no property or ontological change)
3. Time does not pass! There is no moving present.

Static theorists generally say that temporal passage (the passing of time) is a *perspectival fact*:

- If we do have the perception of time passing or change, this is only a matter of perception.
- The world does not actually change. There is not actually a passage of time.
- Time does not flow

This sounds strange at first since we believe that our perception of change (the passage of time) corresponds to real change in the world, but keep in mind that (1) there is a difference between appearance and reality and (2) there are plenty of perspectival facts that we correct on further reflection.

In summary, we have considered three different theories of temporal passage. Dynamic theories contend that there is real change in the world and the passage of time is nothing more than this change. Property theories contend that what changes are events gaining or losing temporal properties, while ontological theories contend that what changes are events coming in or going out of existence. In contrast to these two dynamic theories are static theories of temporal passage. This theory asserts that there is no real change in the world and, as such, so time does not really pass.

14.3 THEORIES OF TIME

Now we have the three main subtheories of a theory of time, the next step is to formulate the theories of time. In each case, we will restate the theory,

refine some of subtheories that compose the theory, and present arguments for and against the theory. Here is a sneak peak.

Theory	A or B	Temp. ontology	Temp. passage
Eternalist	B theory	eternalist	static
Presentist	A theory	presentist	dynamic (ontological)
Growing Block	A theory	past-presentist	dynamic (ontological)
Moving Spotlight	A theory	eternalist	dynamic (property)

DOES ONLY THE PRESENT EXIST?

In the previous chapter, we consider our first theory of time: the moving spotlight theory (MST). In this chapter, we consider our second theory of time: presentism. Just as in the previous chapter, our plan will be to (1) articulate presentism, and then (2) consider arguments for and against presentism.

15.1 PRESENTISM

Let's begin by defining presentism:

Definition 15.1: presentism

Presentism is a theory of time that consists of the following components:

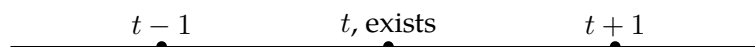
1. A-theory
2. presentist temporal ontology
3. Ontological-change dynamic theory of temporal passage: time passes because what exists changes

Some other theorists formulate presentism as consisting of two claims, but their formulation is substantially the same as they combine the fact that presentism is an A-theory with its temporal ontology. For example, Ingram writes that “I say that there are two elements to presentism as a theory of time, an ontological thesis and a dynamical thesis.”[60, p. 11]

With respect to (1), since presentists accept the A-theory, they contend that events are structured in time by being past, present, or future.



With respect to (2), the presentist claims that “the present” is coextensive with what exists. What this means is that the objects or events that are *in the present* are the same objects that *exist simpliciter*. In other words, what exists is what is present, and what is present is what exists. So, for example, if t is the present moment, then the only things that exist are those things that are in the present.



What about the objects at $t-1$ and $t+1$? For the presentist, these objects don't exist simpliciter: they don't exist without qualification. Instead, a presentist will typically say that objects in the past "did exist" and objects in the future "will exist".

There is significant debate concerning how to formulate this thesis of the presentist doctrine and whether it says anything non-trivial (for a discussion, see [60, pp. 27–33]).¹ For example, consider the following definitions of presentism (note that not all the authors accept these definitions):

1. For every x , x exists if and only if x is present.
2. For every x , x exists if and only if x exists now. [13]
3. Reality is limited to the present. [21]
4. Everything is present [7, 36]
5. All and only present things exist [100]
6. Necessarily, it's always the case that only present entities exist. [59]
7. Only present things exist. [59]
8. At the only things that have ever existed or will are ones that exist now. [88]
9. Only present entities are real [8]
10. The way things are is the way things presently are [55]
11. Reality does not extend beyond the present moment, and therefore there are no (wholly) non-present objects or events [36]
12. Necessarily, it is always the case that only present entities exist [60, p. 39]

But let's make a few points of clarification regarding the ontological thesis of presentism:

First, presentism contends not merely that the only *objects* that exist do so in the present, but the only *entities* that exist do so in the present. Thus, if there are properties or events, in addition to objects, then the only ones that exist do so in the present.

Second, what does presentism say about *abstracta*? That is, what does presentism say about objects that seem to exist outside of spacetime, e.g., mathematical objects like numbers? There seem to be a few possible answers to this question. On one approach, abstracta do not exist because the only objects that exist are those in the present moment. So, since abstracta are not spatiotemporal entities, they do not exist. On a second approach, perhaps the presentist can say that while abstracta are not spatial entities and so cannot causally interact with our senses, they are temporal entities. On this approach, the presentist might say that numbers exist in the present moment.

A third approach is simply to say that presentism has nothing to say about whether abstracta exist.

Third, let's clarify two different statements of presentism (one of them being incorrect):

1. *it is always the case* that the only entities that exist are in the present
2. the only entities that exist are in the present

Presentism is better expressed by (1) rather than (2). To see why, consider someone who accepts a past-present temporal ontology. Now imagine their perspective at the beginning of the creation of the world. At the moment the world is created, there is only one time. At this moment, the presentist and the past-presentist theory will both accept (2): "the only entities that exist are in the present". However, they will disagree about (1) since the presentist will accept (1) but the past-presentist will not.

Fourth, it is unclear whether presentism is a necessary or contingent truth. Some theorists believe that all metaphysical truths are metaphysically necessary (true at every possible world). Others, by contrast, will contend that the goal of metaphysics is restricted to capturing the truths about the actual world.

Fifth, applied to our own existence, the presentist says that "the you that are now is all the you that there is." [3] There is no prior you existing in the past. There is no future you existing in the future. The only you that exists anywhere is the one here right now.

Finally, with respect to (3), the presentist claims that the present changes or *what is present* changes. On the presentist theory, change does occur (time does pass). Here is a statement of the presentist's thesis regarding change:

"a presentist world is one that dynamically changes over time: the totality of events that exist changes as time passes, so that a different set of events comes into existence as each new present moment comes into existence, and those events then pass out of existence as that moment ceases to be the present moment." [82, p. 346]

Notice that in the above quotation, it is not necessary that for time to pass that everything in the world changes. That is, it is not necessary that all of Tek's parts undergo a complete transformation at every second. Instead, what the present asserts is that time passes when there is some change to the *totality* of what exists.

Exercise 15.97

We now have a rough sketch of presentism. Before we consider two varieties of presentism, create a diagram or picture of presentism. Use labels to aid in understanding.

15.2 VARIETIES OF PRESENTISM

There are many varieties of presentism.[cf. 32][60, pp. 23–27] We will consider two versions.

15.2.1 Solipsistic presentism

In order to understand solipsistic presentism, it is helpful to first understand what the term “solipsism” means in general.

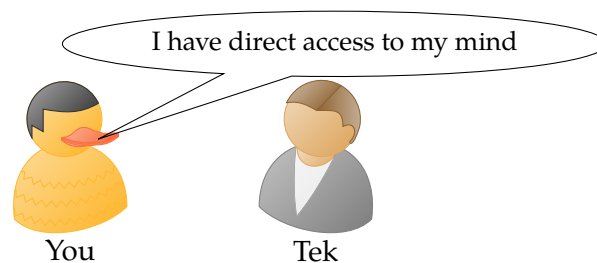
Definition 15.2: Solipsism

Solipsism is a theory the only mind for which we can be certain that it exists is our own.

Here is a simple argument for solipsism:

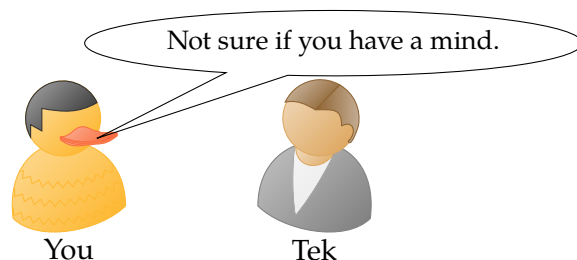
- P1: I have no direct and definitive empirical proof that other minds exist.
- P2: I do have direct and definitive empirical proof that my own mind exists.
- P3: I should only accept the truth of propositions for which I have direct and definitive empirical proof.
- C: Therefore, my mind is the only mind that exists.

The basic idea then is that we only have *direct access* to our own thoughts, feelings, emotions, perspective and not others. This direct access gives us the evidential basis for knowing that our minds exist.



Our belief that *other* organisms have minds can only be “known” indirectly. For instance, we observe a person frowning and infer that they must be feel-

ing sad. And, since sadness (rather than say pain) is an attribute of the mind, we then infer they have a mind like our own. However, our inference that they have a mind is only that: an inference. It is not a direct observation. We don't feel or experience their pain. And, since every inference is open to the possibility of error, we have no definitive proof that there are other minds.



Solipsistic presentism takes inspiration from the solipsism about other minds. First, let's define solipsistic presentism:

Definition 15.3: Solipsistic presentism

Solipsistic presentism is the theory that there is one and only one present that exists and that has ever existed. It is the one that we are in right now!

Next, let's transform the solipsistic argument against other minds into an argument against other times other than the one present that we currently experience.

- P1: I have no direct and definitive empirical proof that other times exist.
- P2: I do have direct and definitive empirical proof that this time (the present) exists.
- P3: I should only accept the truth of propositions for which I have direct and definitive empirical proof.
- C: Therefore, this is the only time that exists.

The above type of presentism is more radical than it appears. This version not only denies the existence of past and future, but it denies that there is or ever were prior present moments. Take your experience of reading about presentism here. Solipsistic presentism contends that the only time that exists is the one that is currently happening. It is not the case that you have proof you ever started reading about presentism. It is not the case that you ever were born, or were a teenager, or did anything other than read this sentence right here and right now.

There are three principal challenges for solipsistic presentism.

The **first** challenge is that it conflicts with common sense and the Common Western Metaphysic. Note that its criterion of proof is so high that it would involve denying that we ever have proof of anything that occurred in the past. There is no proof that dinosaurs ever existed, that the former president of your country was the former president, or that you were ever lived one second prior to this time. Because this theory conflicts so strongly with the Common Western Metaphysics, solipsistic presentism needs to provide strong arguments in support of its position.

The **second** challenge is that it lacks explanatory power. If there is no proof you ever existed, then how do we explain the existence of memories? Are these simply illusions of a past for which we have no proof for? If there is no proof the past ever existed, then where did fossils come from? How did the universe become so complex and organized, and did the words you are reading ever come into existence? The existence of a prior present moment gives us a basis for explaining all of these facts. For example, you have memories because you had experiences at a moment that *was* present but is now past.

The **third** challenge is that it is not entirely clear that there is anything like direct and definitive proof of our own minds or the present time. How do you know that your mind exists? A simple solipsistic will contend they have proof that they have a mind because they can directly observe (intuit) it. However, do we directly observe our own minds? Let's assume that we directly observe a thought or emotion. It seems then that we reason from this observation to (1) a mind *M* that exists and (2) that the mind *M* that exists is our mind. So, it is not the case that we directly observe the existence of our own mind but instead infer it from experience. But that is roughly the same way we know other minds. The only difference is that we reason, not from from emotions or thoughts, but actions and habits.

15.2.2 *Dynamic presentism*

A more popular form of presentism is dynamic presentism.

Definition 15.4: Dynamic presentism

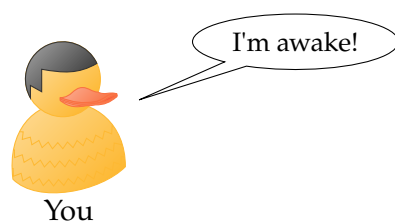
Reality consists of successive present moments.

The basic idea behind dynamic presentism is that something exists and while it exists, it is present and if it no longer exists, then it is no longer present. Reality is nothing more than a succession of present moments. In contrast to solipsistic presentists, dynamic presentists say “past events did exist”, “fu-

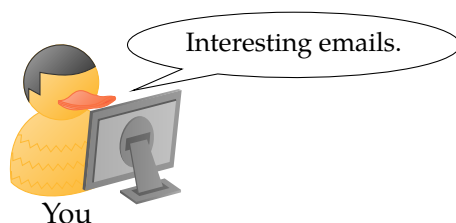
ture events will exist", but it is only present events that exist simpliciter.



Let's consider a concrete example. Suppose you wake up in the morning. The event of you waking up is the present. This event exists.



Next, you get out of bed and check your email. The event of you waking up is no longer present and no longer exists. Of course, this event *did exist*. The dynamic presentist does not entirely deny the existence of the past. However, what simply exists is what is present. And, what is present is the event of you checking your email.



Finally, you leave for work. The event of you getting out of bed is no longer present and no longer exists. Again, the event of you getting out of bed and the event of you checking your email *did exist* but it does not exist in the same way as the present event exists. The present event exists *simpliciter* but past events exist with qualification.



What exists and what is present is you leaving for work.

Exercise 15.98

You know have the theory of presentism before you. Create a picture or diagram that captures the core idea of presentism. Feel free to label the diagram to make things clear.

15.3 ARGUMENTS FOR PRESENTISM

In this section, we will consider some arguments in support of presentism.

15.3.1 *Argument: From Intuitions*

Presentism is taken to be the most intuitive theory or the theory that best accords with common sense. [60, pp. 39–43] John Bigelow writes that presentism was “believed by everyone ... until at least the nineteenth century.” [10] Similarly, de Clercq writes “This view [presentism] can be called common sense with some justification. After all, past entities seem to be entities that have ceased to exist, and future entities seem to be entities that have yet to come into existence.” [33]

Given the many different common sense beliefs about time, what is it about presentism that makes it best fit with common sense? According to Zimmerman [156]:

“The following questions go to the heart of the deepest metaphysical disagreement about the nature of time: (1) Are there objective differences between what is past, present, and future? (2) Are present events and things somehow more ‘real’ than those wholly in the past or future? I should like to respond, ‘Yes’, to both questions. Affirmative answers sound obvious and commonsensical, at least to me. Indeed, I suspect that, for many of us, belief in a deep distinction between past, present, and future can be given up briefly, if at all; and then only by a mighty effort of the will.”

In other words, our mostly deeply held beliefs about time, as they relate to what exists, are (1) there is an objective difference between the past, present, and future and (2) the present is somehow *more real* than the past and future. On this account, we might pose the following argument:

- P1: There are two primary common sense beliefs about the ontology of time: (1) objective difference between past, present, and future and (2) present is more real than past and future.
- P2: Only an A-theory best fits with these common sense beliefs.

- C: Therefore, all other things being equal, we ought to accept some version of the A-theory.

There is only one small step to take us from an A-theory to presentism. A presentist only needs to show that the best A-theory is presentism. Deciding between presentism and the other A-theories requires a comprehensive analysis of the arguments for and against these theories. However, at least according to Zimmerman, presentism is preferable to other A-theories because it is the only theory that preserves the common sense beliefs that the present is more real than the past and the future.

- Growing block theory (GBT) adopts a past-presentist ontology and so the past is just as real as the present
- Moving spotlight theory (MST) adopts an eternalist ontology and so the past, present, and future are equally real.

Expanding our argument

- P1: There are two primary common sense beliefs about the ontology of time: (1) objective difference between past, present, and future and (2) present is more real than past and future.
- P2: Only an A-theory best fits with these common sense beliefs.
- P3: The best A-theory is presentism as it best preserve the common sense datum that the present is more real than the past and the future.
- P4: All other things being equal, we ought to accept the best version of the A-theory.
- C: Therefore, we ought to accept presentism.

Generally, but not always, arguments from common sense are not thought to be decisive. More reservedly, they are thought to accomplish two ends. First, they provide a *prima facie* reason to accept a theory or motivate the initial preference for one theory over another. In other words, all other things being equal, we ought to accept presentism over any other theory. Second, they play a role in assigning the burden of proof. Since presentism best accords with our intuition, the burden of proof is on non-presentist theories to show why (1) presentism is false and (2) those theories are better alternatives.

15.3.1.1 *Argument: From free will, responsibility, and the thinning tree view*

One way to argue for a metaphysical theory is to argue that if we were to accept an alternative theory, we would have to give up some other deeply-held belief. For example, suppose someone tells me that poor people are poor because they are lazy. They claim that it is easy to become rich, you only need to buy their book and follow their ten-step method to fame and fortune.

But suppose I have a hard-working friend who I admire. My friend was born into poverty and has a serious medical condition. Now if I accept the view that poor people are at fault for their poverty because it is simple to become rich, then I have to give up on my view that (1) my friend is hard-working (his apparent hard work is an illusion) and (2) that sometimes a person's fate is the result of circumstances completely outside of their control. Since I more firmly believe that my friend is hard-working than he is not, I'm likely to deny the claim that all poor people are poor because they are lazy.

The logic of this argument is relatively simple.

- P1: Belief B is deeply held.
- P2: Theory T requires us to give up B .
- P3: Another theory T^* does not require us to give up B .
- P4: All other things being equal, we ought to accept T^* over T .
- C: Therefore, T^* is better than T .

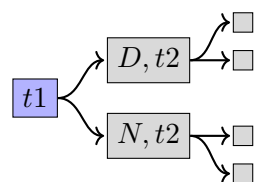
One way to argue for presentism then is to argue that other theories conflict with our deeply-held beliefs concerning (1) the thinning-tree view of the universe, (2) free will, and (3) moral responsibility.

Let's consider each of these in turn. The thinning-tree view of the universe is the view that the future contains open possibilities and what possibilities become actualized depend upon choices that we make.

Definition 15.5: thinning-tree view

The thinning-tree view of the universe is the view that (1) the future is an open set of possibilities (Open future) and (2) the passage of time plays a role in the actualization of these future possibilities.

For example, suppose Tek is at t_1 and he is faced with the choice to become a doctor (D) or nurse (N).



Tek's future will look different depending upon which choice he makes. If he becomes a doctor, he will likely have a future where he makes more money but also has more debt. If he becomes a nurse, he will likely have a future where he makes less money but has less debt. Now at t_1 , the universe consists of the actual event of Tek being neither a doctor nor a nurse. What

it consists of instead is a set of different future events that are *possibilities* rather than being *actualities*. Tek's decision to become a doctor or a nurse, transforms one of these possibilities into an actuality.

Presentism is clearly consistent with this way of viewing the world since (1) the future does not exist and (2) new events come into existence as time passes.

Why would someone believe in the thinning tree view of the universe? First, the thinning tree theory of the universe corresponds with we go about making decisions in the world. People deliberate about whether to do action A or B. If the action were already determined, then this deliberation appears entirely pointless. In other words, the thinning tree view of the universe is consistent with rational deliberation and everything that rational deliberation assumes, e.g., human freedom.

Second, we don't know anything about the future but we do know about the present. Believing that the future already exists is pure speculation.

Third, the thinning tree theory of the universe generally corresponds with how people understand the structure of the world. People believe that the world (at least at the subatomic level) is governed by some amount of randomness (indeterministic processes). If future events already existed, then no such randomness would exist.

In addition, presentism is consistent with the possibility of free will and the assignment of moral responsibility. Free will and moral responsibility typically imply the capacity to act otherwise (if my action *A* at *t*₁ is free, then I had the power at *t*₁ to do an act other than *A*). In addition, assigning moral praise and blame to a person typically requires that a person's actions are free. The assumption is that you are responsible for what you can control, but if you cannot control any of your actions, then you should not be held responsible for them. However, if the future already exists, then it is unclear that you are free since you have no ability to act in any way other than you now act.

At first glance, one argument for presentism is that presentism are the only theories that are consistent with there being an open future.

- P1: Beliefs in the open future, moral responsibility, and free will are all deeply held.
- P2: Other theories requires us to give up these beliefs.
- P3: Presentism is consistent with these beliefs.
- P4: All other things being equal, we ought to accept presentism over these theories.

- C: Therefore, presentism is better than these theories.

There are several problems with the above argument.

First, not everyone shares these "deeply-held beliefs". For example, some people have no opinion about whether the future is open or they may hold this view but it is not "deeply-held". This would mean that it would be no great sacrifice to give it up.

Second, at least at this point, we have not shown that P2 is true and so we will have to delay our evaluation about whether it is true or not until we consider the rivals to presentism.

Third, it is not clear whether P4 is the case since it is not clear whether other theories have certain advantages that make giving up these so-called deeply-held beliefs worth the sacrifice.

15.3.1.2 *Argument: From the confinement doctrine*

One of the strongest reasons to accept presentism is that it seems that the only time that we experience is the present. In other words, our conscious experience is **confined** to the present. Here is a simple argument for the truth of presentism:

- P1: The only conscious experience we have is of present events: events occurring in the present in the world (confinement doctrine).
- P2: Presentism says that the only events that exist are those in the present.
- C: Therefore, the best theory is presentism.

P1 is sometimes called the **confinement doctrine**. The confinement doctrine says that the objects and events we experience are those in the present moment. At first glance, the confinement doctrine seems obviously true. Suppose it is 2026 and Tek says he can see one hundred years into the future. Few, if anyone, would believe Tek since we think that people cannot experience the future. Similarly, suppose Tek says he not only can remember what happened 100 years ago, but he can see it as well. Again, few would believe Tek since no one can experience past events.

However, on deeper reflection, there is good reason to think that the confinement doctrine (P1) is false. Suppose that you are staring at the sun. As you stare at the sun, there is an image being burned onto your retina. You say to yourself "I only experience objects in the present. I am experiencing the sun. Therefore, the sun is in the present moment." However, consider that the light from the sun takes 8 minutes and 20 seconds to reach your eyes.

Your conscious experience is not of a present object (the sun) but of the *effects* produced by an object. More generally, since it always takes time for light to reach our eyes, or for our brain to register some sensory phenomenon (e.g., the roughness of a desk), our conscious experience of objects is of that object's effects (at time t_2) not of the object itself (at t_1).

Exercise 15.99

Create a falsifying instance of the confinement doctrine. Namely, create a case where you experience an event E or an object O , but E is in the past (or even O does not exist anymore!). Draw a picture to illustrate?

15.4 ARGUMENTS AGAINST PRESENTISM

There are many arguments against presentism. I will briefly mention some of these before focusing on one in particular.

First, there is the problem of cross-time relations. The presentist seemingly cannot account for relations between objects or events in the present and those that exist at other times. For example, if causes occur before their effects, but reality is exhausted by the present, then how does the presentist account for the fact that one event is the cause of another.[\[33\]](#)

Second, presumably, language about the past and future is meaningful. Part of language involves proper names. According to one theory of names, names get their meaning by directly-referring to their objects. For example, “George Washington” gets its meaning by referring to the object *George Washington*. But, according to presentism, past and future objects do not exist. Therefore, this seems to imply that such names have no meaning, which is counterintuitive.

Third, presentism is often said to be false since it conflicts with our best science. Presentism posits a single absolute time called “the present”. The present is thought to imply the idea of absolute simultaneity. But the special theory of relativity (STR) is said to have no absolute notion of present; instead, it only operates with relative simultaneity. If presentism conflicts with our best physics, and there are other theories of time that do not make use of the notion of the present (e.g., eternalism), then there is a problem with presentism, not with our physics (see [\[117\]](#), [\[104\]](#))

15.4.1 Objections to presentism

The primary objection to presentism is the *grounding problem against presentism*.^[5] To understand this problem, a few preliminaries are required.

15.4.1.1 Contingent propositions

Let's define what it means for a proposition to be contingent.

A proposition P is possible when it is true at least one possible world w_1 . For example, "unicorns fly" is possible. A proposition P is necessary ff it is true at every possible world. Logical truths (e.g., "John is tall or not tall") are said to be necessary. A proposition P is impossible (not possible) if there is no world at which P is true. For example, "David is both married and unmarried" is impossible. Finally, a proposition that is contingent is both not impossible (so true in at least one world) and not necessary (so false in at least one world).

Definition 15.6: Contingent

A proposition P is contingent if and only if there is some possible world w_1 where P is true and some world w_2 where P is false.

In other words, P is contingent iff:

1. P is T in some world w_1
2. P is F in some world w_2

To illustrate, consider the following sentence: "This penguin is wearing glasses".



This sentence is contingent since it true in this world (at least one world) but not true in every world (not necessary).

Definition 15.7: Contingently true

A proposition P is contingently true if and only if (1) P is true in this world and (2) there is some possible world w_1 where P is false.

In other words, P is contingently true iff:

1. P is T in our world
2. P is F in some other world w

To illustrate, consider the following sentence: “Donald Trump was the 45th President of the United States”. This proposition is true in our world but false in another possible world.

Exercise 15.100

1. Write a proposition that is contingent
2. Write a proposition that is possibly true.
3. Write a proposition that is necessarily true.
4. What is the difference between a proposition being contingently true and possibly true.

15.4.1.2 *Grounding*

There is a relation between truth and reality. At least for sentences about the world, if a sentence is true, then its truth is said to be *grounded* in reality. If a sentence is false, then it is not *grounded in* (has no basis in) in reality. This basic idea is called the *grounding principle*.

Definition 15.8: Grounding principle

Every contingently true proposition P about the world depends for its truth on the world.

Some examples. Suppose you are in college and you say “I paid my tuition”. Whether you paid your tuition depends upon the facts of this world: namely, whether you did, in fact, pay your tuition.

While the *grounding principle* may be true, it is vague. If it is true, then *what* part or parts of the world ground the proposition P . We consider two different ways of making the grounding principle more precise in the next two subsections.

15.4.1.3 *Grounding as truthmaking*

One way of precisifying it is via the *truthmaking principle*.

Definition 15.9: truthmaking principle (TMP)

Necessarily, for any contingently true proposition P , there exists something TM (its truthmaker) that makes P true.

So, if “Tek exists” is true, then there exists something (namely Tek) that

makes this sentence true. Or, if “the penguin is wearing a crown” is true, then we have the truthmaker that is the combination of the penguin, the crown, and some wearing relation that holds between them.



However, one problem with the truthmaking principle are negative existential propositions (or “negative existentials” for short). A negative existential is a proposition that denies the existence of some entity. For example, take the negative existential “Unicorns do not exist.” If the truthmaking principle is true, then there exists a truthmaker TM that makes this sentence true. But, there is no obvious candidate to serve as TM . That is, there are no objects that make “unicorns do not exist” true.

Can't the world as a whole serve as the truthmaker for a negative existential proposition? That is, can't the fact that the world itself does not contain any unicorns be the truthmaker for the negative existential “unicorns do not exist”? While this is a possibility, one issue it raises is that every negative existential would have the same truthmaker. That is, the truthmaker for “unicorns do not exist” and “Bigfoot does not exist” would have the same truthmaker: the entire world.

One way to avoid the problem of negative existentials sharing the same truthmaker is to contrast the way the world is with some other world. This approach, however, involves a different specification of the grounding principle. This alternative way of making the grounding principle precise is known as the *supervenience principle*.

15.4.1.4 Grounding as supervenience †

Let's start by defining the supervenience principle.

Definition 15.10: supervenience principle

For any proposition P and any worlds w_1 and w_2 , if P is true at w_1 but not w_2 , then either (1) something exists in one of the worlds but not the other or (2) something instantiates a property (or relation) in one world but not the other. [5, p. 342]

If the supervenience principle applies to proposition P , then we say that “the truth of P supervenes on the world.” In other words, its truth depends upon

the way that the world is.

To illustrate, suppose “Tek exists” is true at w_1 but not at w_2 . That is, suppose “Tek exists” is contingently true.



If “Tek exists” supervenes on w_1 , then there must be a difference between the two worlds (different objects existing or those objects having different instantiation patterns). This could occur by Tek existing at w_1 but not w_2 .

Now suppose the proposition “Penguins do not exist” is true at w_2 but is false at w_1 . The truth of this proposition supervenes on w_2 even though there is not an object in w_2 that makes it true (there is no truthmaker). The truth supervenes on the world because its truth depends upon the way that this world is (in contrast to another world). More specifically, it supervenes on w_2 because something exists in one of the worlds but not the other (penguins existing in w_1 but not w_2).¹

To help further clarify the supervenience principle, let's consider its contrapositive:

Definition 15.11: supervenience (contrapositive)

For any proposition P and any worlds w_1 and w_2 , if it is not the case that (1) something exists in one of the worlds but not the other and (2) something instantiates a property (or relation) in one world but not the other (in other words, both worlds have the same objects and those objects have the same properties / relations), then it is not the case that P is true at w_1 but not w_2 (that is, there is no difference in the truth of P_{w_1} and P_{w_2}).

Importantly, note that the converse isn't always true: there can be a difference in reality without a corresponding difference in the truth of two sentences.

¹To be clear, according to the supervenience principle, it is not the case that a true proposition P requires an object O in the world where P is uttered to exist to make P true. That is, the truth of the sentence “unicorns do not exist” does not depend upon there being *something* in this world that corresponds to this sentence. Instead, what grounds the truth of this sentence is that the truth of this sentence *depends upon* the way that this world is in comparison to other worlds. In particular, it depends upon some other world having something that this world lacks, namely unicorns.

Imagine two worlds w_1 and w_2 and Tek is 6'7 in w_1 and 6'8 in w_2 . Now consider the sentence "Tek is taller than 6'0". In this case, while the worlds are distinct, the sentence is true in both worlds. So, the supervenience principle is consistent with the claim that *there can be a difference in being (worlds) without a difference in truth*.

Exercise 15.101

The grounding principle is sometimes taken to be a "truism".[5, p. 342] Is the grounding principle true? Does it have any limitations? Does anyone you know ignore the grounding principle (in practice)?

15.4.1.5 *Independent of the past*

Next, consider two insights (1) there are truths about the past (past truths) and (2) past truths do not depend upon what happens in the present. With respect to (1), the following are true:

1. There were dinosaurs.
2. Napoleon died on 5 May 1821.
3. On 6 June 2013, the Guardian received information that the NSA was using PRISM to spy on the web activities of U.S. Citizens.

Let's call the claim that "past truths do not depend upon what happens in the present", the "independence of the past principle".

Definition 15.12: Independent of the past principle

Past truths do not depend upon what happens in the present.

We can state this principle in several ways (using the ideas of grounding, truthmaking, or supervenience):

1. Grounding of past truths: The past truths are not grounded in the present state of the world.
2. Truthmaking of past truths: The truthmakers of past truths are not in the present.
3. Supervenience of past truths: Past truths do not supervene on the present state of the world.²

²Let P be about a past event in this world w_1 . If P is true in w_1 but not in w_2 , then that truth does not depend upon either (1) something *presently* existing in one world or not the other or (2) something *presently* instantiating a property (or relation) in one world but not the other. More concretely, if *there were dinosaurs* is true in this world, then this does

Let's illustrate each one of these precisifications with an example:

1. Grounding of past truths: Whether or not there were dinosaurs is not grounded in the present.
2. Truthmaking of past truths: What makes it true that "there were dinosaurs" are not truthmakers (objects) in the present.
3. Supervenience of past truths: The truth of "Napoleon died" does not supervene upon the present state of the world. That is, the truth of "Napoleon died" depends upon the way that the current state of affairs.

Let's focus a little more on how past truths do not supervene on the present state of the world. Imagine two worlds:

- w_a : the present state of the actual world
- w_i : the present state of a world identical to ours but that was created ten minutes ago.

Next, notice that "there were dinosaurs" is true in w_a but not in w_i . And so if "there were dinosaurs" (a past truth) supervenes on w_a , then there should be a corresponding difference in either (1) what is (what exists) or (2) the instantiation pattern of what exists. But notice that there is no difference between w_a and w_i concerning these two features. Therefore, "there were dinosaurs" (a past truth) does not supervene on present facts.

Exercise 15.102

1. The claim is that past truths do not depend upon the present. It is not stating that the present does not depend upon the past. Create an example of (i) a case where a past event does not depend upon the present and (ii) a present event does depend upon the past.
2. Do you accept the "independence of the past principle"?

15.4.1.6 Grounding and presentism

With the above preliminary points out of the way, it appears that the presentist cannot accept of the following propositions:

not depend upon something existing presently existing in this world (or some instantiation pattern existing) but not some other world since *there were dinosaurs* could be false in that world but the same things exist (and the same instantiation pattern). As Baia[5, p. 343] points out, we could imagine a scenario where things are just as they are but the prior events leading up to the present day did not contain dinosaurs.

- **Presentism:** Reality is limited to the present: only present objects (and their instantiation patterns) exist.
- **Independence:** Past truths are not grounded in the present, do not have present truthmakers, do not supervene upon the present state of the world.

The grounding argument against presentism can thus be expressed as follows:

- P1. Contingently true sentences about the world must be grounded.
- P2. There are contingently true propositions about the past.
- P3. If there contingently true propositions about the past, then they must be grounded.
- P4. Past events are independent of present events.
- IC. Therefore, contingently true propositions about past events cannot be grounded by present events.
- P5. The presentist says that there are only present events.
- C. Therefore, presentism cannot ground contingently true propositions about the past.

Let's simplify the above argument as follows:

- P1. Truths about the past are made true by past truthmakers.
- P2. Presentism says there are no past truthmakers.
- C. Therefore, either presentism is false or there are no truths about the past.

Exercise 15.103

Is this a convincing argument against presentism? If “Napoleon died on 5 May 1821” then what makes this true?

15.4.1.7 Responses to grounding

Let's consider a few ways the presentist might respond:

1. Bite the bullet: there are no truths about the past
2. Redefine truth
3. Lucretian properties
4. Appeal to a special kind of grounding

The *first response* is to “bite the bullet”. You bite the bullet on an objection when you say that some unfortunate consequence of a theory must simply

be accepted. The presentist can bite the bullet on the grounding objection in two different ways.

First, they can say that sentences about past events do not need to be grounded since there are no past truths. This would imply that a sentence like “dinosaurs existed” is either false or neither true nor false. The problem with this claim is that it clashes with common sense and the presentist has provided no reason (besides the fact that it clashes with presentism) to accept it.

Second, they can say that true propositions about past events (or even all true propositions) do not need grounding. Past sentences can be true without needing to be grounded. This response is sometimes called “ontological cheating”.

Definition 15.13: ontological cheater

An ontological cheater is someone who says they have preserved truths without properly adhering to the intuitions underpinning the grounding principle.[54, p. 141][130, p. 429]

Since ontological cheating clashes with our intuitions, and the presentist has provided no independent reason for saying we should ignore our intuitions, the cheating strategy is not well-endorsed.³

A *second response* is to resort to redefine what it means for a proposition to be true in such a way that past truths can be grounded in presenting existing entities and their instantiation pattern. Here is one attempt.

Definition 15.14: Presentist truth

A proposition P is true iff P is derivable from presently existing objects and their instantiation patterns.

On this approach, “there were dinosaurs” is true because their existence can be derived from presently existing fossils.

There are a few objections to this response. First, this response is too vague to evaluate since it is not clear in what sense “derivable” is being used.

Second, there is the *objection from buried secrets*. Take the proposition “There were exactly n dinosaurs” where n is some precise number of dinosaurs. Now suppose that several dinosaurs (let's say the number is i) were consumed by lava. As such, there appears to be no presently existing objects

³However, see [130] and [131] for a defense.

(and pattern) that would allow for the derivation of their existence. In such a case, “there were n dinosaurs” would be true, but, in fact “there were $n + i$ dinosaurs.” Closer to home, suppose Tek murders Jon but disposes the body in acid in such a way that there is no presently existing trace. It would follow then that Tek did not murder Jon. But, Tek did murder Jon.

A *third response* is to ground past truths in presently existing objects and their instantiation patterns. This is sometimes called “present record truthmaking”:

Definition 15.15: present record truthmaking

Past tensed truths are made true by the present way the world is. [84, p. 240]

This type of response differs from the prior response because propositions about the past are not made true by what can be *inferred about the past* from the present state of affairs but by facts directly found in the world itself in the present.

To better understand this approach, let's consider a special type of property known as *Lucretian (historical) property*. What grounds past truths are not past objects but present objects containing properties known as Lucretian properties (or presently existing objects having instantiation patterns with Lucretian properties). T

To understand this response, consider the following claims:

- Consider a block of clay at t_1
- Suppose the block of clay is molded into a can at t_2 .
- At t_2 , the can has, in addition to its observable property of *being a can* but also a non-observable Lucretian (historical) property of *being a block*.

Typically, Lucretian properties are *tensed properties*. For example, suppose you were short in 2023 but tall in 2025. On the presentist account, if the present is 2025, then you have the property of *being tall*, but also the tensed property of *was short*.

There are several objections to the appeal to Lucretian properties.

First, what is a Lucretian property? At first glance it appears not to be an intrinsic property of an object but instead a relation between an object and an instantiation pattern at a prior time. If this is the case, then it appears to be committed to the existence of prior times, which conflicts with presentism.

Second, if a Lucretian property is an intrinsic property of objects, then it clashes with common sense. Suppose I love my grandma. Now suppose she died and is buried. If I still claim “I love my grandma”, then what is the object of my affection? It cannot be a past object (since no such object exists according to presentism). Instead, it is a presently existing object (the corpse) that has the Lucretian property of *being my grandma*. This is a strange result. And, it is perhaps even stranger if my grandmother is cremated.

Third, some might argue that Lucretian properties ought to be rejected as their existence is unverifiable. To illustrate, imagine two worlds w_1 and w_2 :

- In w_1 , there is a block at t_1 that is molded in such a way that it is a can at t_2 (the can thus has the Lucretian property of *being a block*.).
- In w_2 , there is a can at t_1 and it stays a can at t_2 (the can at t_2 does not have the Lucretian property of *being a block*.)

Now suppose that you are able to observe both worlds at t_2 . Presumably, there is no way to tell which can has the Lucretian property of *being a block* and so the Lucretian property is completely unverifiable.

A *fourth response* is to offer a different account of what it means for a true proposition to be grounded. Rather than claiming that every true proposition depends for its truth on how the world is (the state of the world), the presentist might propose what is called *presentist grounding* (p-grounding).

Definition 15.16: P-grounding

Every true proposition depends for its truth on how the world was, is, or will be.^a [130, p. 423]

^aFor example, Westphal writes “a proposition can be true only if there is or was something in virtue of which it is true.” [westphal2006] [cf. 5, p. 345]

On this approach, the presentist can say that “there were dinosaurs” depends for its truth not on the way that the world (presently) *is* but on the way the world *was*. With respect to future truths, the presentist can say that “Tek will eat a sandwich” depends for its truth not on the way that the world (presently) *is* but on the way the world *will be*. As Tallant puts it,

“I explain the truth of ‘Caesar crossed the Rubicon’ as follows. I say that my truthmaking principle doesn’t require there really to be Caesar, crossing the Rubicon, in the domain of my most unrestricted quantifiers to explain why it is true that Caesar crossed the Rubicon. The principle requires only that there *has* existed

the fact of Caesar's crossing. And of course, this fact has existed, as Caesar *has* crossed the Rubicon.”[**tallent2009**]

One criticism of this response is that it is ontological cheating since it does not ground truths in *existing* things or their instantiation patterns. That is, it fails to ground truths in *reality*: there is nothing that exists that *makes* the sentences true.⁴

Exercise 15.104

1. We have examined the grounding objection to presentism. How convincing is this criticism?
2. There are several other objections to presentism. Can you think of any additional objections?

⁴For a response to this criticism, see [5]

We saw that one of the principal challenges to presentism was its seeming inability to account for past truths. Rather than revising the definition of truth, trying to revise what it means for truths to be “grounded” in reality, or trying to ground past truths in present facts, another approach is to abandon presentism in favor of an alternative theory of time. One such theory is the growing block theory (MST).

16.0.1 Moving Spotlight Theory

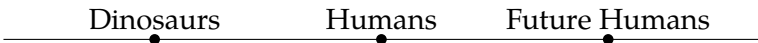
The Moving Spotlight Theory (MST) consists of the following components:

Definition 16.1: Moving Spotlight Theory (MST)

1. A-theory: a proper subset of objects, events (times), or properties instantiate the property of being present.
2. An eternalist temporal ontology: past, present, and future objects exist.
3. Time passes via events gaining and losing the property of being present (the objects, events, properties that instantiates the property of being present changes).

Let's call someone who accepts MST a “spotlighter”. Next, let's briefly review each of these components. First, the MST accepts the eternalist theory of temporal ontology. This means that it is always the case that everything that exists does so eternally. It also means that objects have their properties eternally. Let's consider a few examples:

- Dinosaurs exist (they are in the past). They exist in the same sense that you do.
- What you are doing as you are reading this slide is real (Present events are also real).
- Future Humans living on Proxima Centauri b is real (we are doing it in the future!)
- Let's say that in the future, you are one inch taller. You have the property of *being your current height plus one inch* in the future.



Relatedly, objects have their properties when they have their properties eternally.

- Example: If you were 6'0 tall on 1 July 2020, then you will always be 6'0 on that date.
- Example: If Jon was an evil person on 1 January 2021, then he will always be an evil person then.

Second, MST also accepts the A-theory. This means that there is some single, objective instant of time that is absolutely (non-relatively) present. For MST, while every time exists independent of its position in time, there is one time (one event) that is the present moment



It is worth emphasizing that there is a *single* present rather than multiple present moments. If there were multiple present moments, then MST would be inconsistent. To illustrate, suppose there are three present moments in a single A-series. If there are three distinct present moments in the A-series, then one present occurs later than another (and so is in the future). This would mean that the later "present" is both present (by supposition) and future (by the fact that it is later than some other present).

Third, MST also accepts that time passes (there is change in the world). This occurs by which objects (which proper subslice of the world) instantiates the property of being present. To illustrate this idea, spotlights characterize the flow of time as though it were a spotlight moving over that which exists.

"We are naturally tempted to regard the history of the world as existing eternally in a certain order of events. Along this, and in a fixed direction, we imagine the characteristic presentness as moving, somewhat like the spot of light from a policeman's bull's-eye traversing the fronts of the houses in a street. What is illuminated is the present, what has been illuminated is the past, and what has not yet been illuminated is the future." [12, p. 59]

In other words, imagine three different times t_1, t_2 , and t_3 . According to the spotlight, all three of these times exist but change occurs due to these different times taking on or losing the property of being present.



For time to pass, there needs to be change. If the world was frozen (no change), then time would not pass. Since MST accepts the property change dynamic theory of temporal passage, what changes are events gaining and losing the property of *being present*.

- Suppose event E1 has the property of being future.
- E1 loses this property and takes on the property of being present.
- E1 loses this property and takes on the property of being past.

The only temporal change that occurs in the world are events gaining and losing A-properties. In other words, the passage of time is accounted for in terms of events gaining and losing the properties of being present and future. To put this in a different way. For MST, temporal passage is a specific type of change to events. It is a change to the times of events; more exactly, it is a change to the temporal position of an event.

16.0.1.1 MST and change

One of the most distinguishing features of MST concerns its account of change (the passage of time). MST gets its name from its claim that the only thing that changes is the property of being present. The present is taken to be like a moving spotlight highlighting a successive set of events. This point is worth emphasizing. The only genuine change that occurs for those who accept MST is which event has the **property of being present**.

“there is genuine change in which moment is present. But notice that the spotlight theorist does not admit genuine change for anything else! For her there is no genuine change in whether I am sitting, or in whether there are dinosaurs, or whether a war is occurring, since her account of these matters is identical to the spatializers”[118]

What type of property are events gaining and losing? What type of property is the property of *being present*? To see this more clearly, let's distinguish properties into two main types:

1. An **intrinsic property** of X is a property X has in virtue of what X is.
2. A **relational property** of X is a property X has in virtue of its relation to something else.

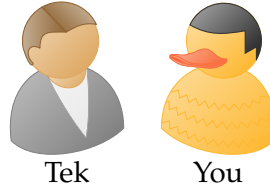
Let's consider some concrete examples. Consider Tek:



Tek

Tek's mass, brown eyes, having a heart, having a kidney, having a soul, having skin, or being kind are all (arguably) intrinsic properties of Tek. Tek

could have all of these properties in an empty world containing only Tek. In contrast, let's consider you and Tek together.

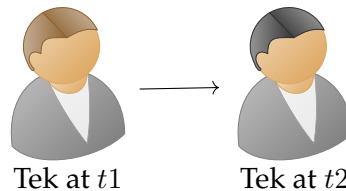


Tek being to the left of you, or 5,000 meters from you, or 150lbs on Earth, or a car owner are all (arguably) relational properties of Tek. Tek only has these properties in virtue of other objects (you, the planet Earth, a car owner, etc.). Tek could not have these properties in an empty world containing only Tek.

Second, properties are also said to either be temporary or permanent:

1. A **temporary property** of X is a property that X can gain or lose.
2. A **permanent property** of X is a property

Tek may have a skin blemish that will go away with time, or he may be angry today but happy tomorrow, or he may be hungry right before he eats a large lunch. Since these are properties (being blemished, being angry, being hungry) that Tek can gain or lose, they are temporary properties.



In contrast, Tek may be a July baby (he was born in July). This is not a property Tek can gain or lose. Or, Tek may be 5,000 meters from you on 1 July 2020. This is a property that Tek cannot gain or lose. These are thus permanent properties.

These four different properties can be combined such that there can be:

1. temporary intrinsic properties
2. temporary relational properties
3. permanent intrinsic properties, and
4. permanent relational properties

Some examples of each include:

PROPERTY	intrinsic	relational
temporary	being a person	being next to someone
permanent	being a person today	being 5m from someone at a time

The purpose of bringing up these distinctions is to highlight the fact that MST contends that the present moment is a temporary intrinsic property. For spotlights, there is no ontological change in the world. No new things are added or removed. In addition, all properties of objects are permanent except for one: the property of being present. The only change that occurs in the world is events gaining and losing the temporary intrinsic property of being present (or being past or being future).

Exercise 16.105

In your own words, explain MST. Include a diagram or drawing with your explanation.

16.0.2 Arguments for MST

Now that we have explained MST, the next step is to critically evaluate the theory. Many of the arguments that we consider for or against MST will also bear on the other theories of time.

16.0.2.1 Argument: from Intuitions

In an earlier chapter, we noted that one way to evaluate the pros and cons of a theory is by considering whether the theory fits with our intuitions. One argument in support of MST is that it is the theory that best fits with our intuitions.

- P1: There is a past, present, and future (intuition 1), this intuition is rational (intuition 2), something metaphysically special about the present moment (Intuition 3) and temporal passage is objectively real, viz., not perspectival (Intuition 4).
- P2: We ought to select a theory that best accords with our intuitions about time.
- P3: MST best explains intuitions (1)-(4) by positing the existing of the past, present, and future (intuitions 1-2), by positing that there is a temporary intrinsic property of being present that events take (intuition 3), and by explaining temporal passage by through the gain and

- loss of being present (intuition 4).
- C: Therefore, MST is true.

There are some potential problems with the above argument. First, there is a *comparative problem*. The argument is incomplete in that it fails to show how other theories do not also fit with our intuitions.

Second, if we assume that MST best fits with our intuitions like the argument shows, then this argument is insufficient for our full acceptance of MST. This is because there are other considerations beyond whether or not a theory fits with our intuitions. For example, if MST conflicts with logical laws or the best science, then we have reason to accept MST.

16.0.2.2 *Argument: Solves grounding problem*

A second argument for MST is the grounding argument for past-tense *and future* sentences. Presumably, certain past-tense sentences are true. There is, then, the question of what makes them true, viz., what are the truthmakers of past-tense sentences. One easy answer to this question is that if past events and objects exist, then it is these past events and objects that make these past-tense sentences true.

- P1: A true sentence requires something that makes it true.
- P2: There are true sentences about the past.
- P3: The best candidate for truthmakers for past-tense sentences are past events (or past objects).
- P4: MST posits the reality of past events (or past objects).
- P5: MST is the only theory that can account for the truthmakers needed for the truth of past-tense sentences.
- C: MST is true.

The core idea behind this argument involves truthmakers. If a sentence is true, then something makes it true (its truthmaker). For example, if I say "I'm six feet tall" and this is true, then David Agler needs to exist and David Agler needs to have the property of being six feet tall. This truthmaking idea applies to past-tense sentences.

Consider the following true sentence:

"George Washington was born 22 Feb 1732"

If this is true, and its truth is not determined by personal opinion, then there is a question of what fact or objects make this sentence true? MST can provide an easy answer: the event of George Washington (who exists) being born on that date (the event exists).

In addition to being able to ground ``past-tense" sentences, MST can also ground ``future-tense" sentences. Consider the following sentences:

1. Tomorrow I will eat a sandwich.
2. In two days I will sell my stock in Alphabet.
3. I will leave this apartment in 90 days.

Now, if these sentences are true, then there must be something that makes them true (some way that reality grounds these truths). In addition, if they are false, then there must be something that makes their *negations* true. For example, if (1) is false, then there must be something in the world that makes ``tomorrow I will *not* eat a sandwich" true.

It is worth pointing out that the spotlighter can argue that *even if* you don't know whether propositions about the future are true or false, the proposition is true or false. If you do eat a sandwich in the future, then the proposition was true. If you don't eat a sandwich, then the proposition was false.

There are several objections to this argument.

First, *P3 is false*. It is not clear that there needs to be past-objects or past-events to account for the truth of past-tense sentences. You could use present events and the laws of nature. For example, we could specify the truth conditions for a past-tense sentence like "X was Y" as follows: ``X is true iff what exists now along with the laws of nature is consistent with X having been true".

Second, there is again a comparative problem (so *P5 is false*). Any theory of time whose temporal ontology that accepts the existence of the past would be supported by the above argument. This would include:

1. eternalist temporal ontology (past, present, and future exist)
2. past-presentist temporal ontology (past and present exist)

With this in mind, the above argument would need to be supplemented by either (1) showing that only MST can explain how past-tense sentences are true or (2) weaken the above argument by saying that the fact that MST can account for the truth of past-tense sentences counts in its favor.

Third, it is not immediately obvious that future-tense sentences need to be grounded as true or false. For example, one might argue that the sentence ``Tomorrow I will eat a sandwich" is *neither true nor false* (indeterminate).

Exercise 16.106

Do propositions about the future (future-tensed) propositions have truth values?

16.0.2.3 Argument: from cross temporal relations

An argument closely related to the argument from truthmakers involves the relations an object or event may have with other objects in time. In short, MST is claimed to be the best theory for grounding cross-temporal relations.

Definition 16.2: cross-temporal relation

A cross-temporal relation R is a relation between two objects a and b where a exists at one time and b exists at another time.

Cross-temporal relations seem to exist. Consider, for example, the relation between a *cause and an effect*: Causes are taken to occur *before* their effects. So, suppose, we are examining a window that is broken and a baseball lying on the floor. It is natural to reason as follows:

1. A ball was thrown
2. It hit the window
3. The ball broke the window
4. The ball now rests on the floor

However, if the past does not exist, then the cause does not exist. And, if the cause does not exist, then it is unclear what sense there is saying that one event is the cause of another event.

Another type of cross-temporal relation are *attitudinal relations*, e.g., admiring, fearing, hoping. For example, suppose I admire Isaac Newton. My admiration is not merely for the work that exists of Newton nor his corpse, but for the object (person) that is Isaac Newton. In order for this relation to exist, it seems that not only do I need to exist, but so does the past object: Isaac Newton.

MST is compatible not only with present-past relations but also present-future relations. People tend to be more skeptical about the existence of the latter relations and so let's provide some motivation for their existence.

Hopes about the future are either rational or irrational. Presumably, a hope for event E occurring in the future is rational if it meets at least two criteria:

1. I have some evidence that E will occur.
2. E will occur.

If a hope fails on either of these two criteria, then it is irrational. For example, suppose I hope that there is faster-than-light travel in the next five years. This hope is not rational because it falls short of both criteria: (1) I have

no evidence that it will and (2) faster-than-light travel will not exist in the future.

Another example, suppose I am told by several physicists that faster-than-light speed is technically possible. I have evidence that faster-than-light speed is possible, but it is not physically possible. My hope for this mode of transportation is thus irrational.

But now suppose that I hope you have a happy life. This hope is rational or irrational. One criteria for the hope being rational seems to be that you exist in the future: for how can I hope that you have a happy life if you don't exist?

Insofar then as we have *rational* hopes about the future, future objects exist. This is consistent with MST, which says future objects are real.

Let's express the above considerations as an argument:

- P1: For some relations, the relation cannot hold unless both objects exist.
- P2: There are some such relations of present objects to past objects and to future objects (e.g. Tek's love for my unborn child or Tek's hope for the future).
- P3: MST best explains these facts in virtue of positing past, present, and future objects along with the property of being present.
- C: Therefore, MST is true.

This is like the truthmaking idea except we aren't talking about sentences but that there are **relations** you have to things in the past and future, and those things must be real for you to be in those relations. If I'm worried about a dental appointment, then it must exist, otherwise I'm worried about nothing!

Let's briefly consider some criticisms of the above argument.

First, there is again a comparative problem (*P3 is false*). Any theory of time that accepts an eternalist theory of temporal ontology can explain cross-temporal relations that extend into the past and into the future, e.g., Tek's love for his future child is a love for an entity that exists later than Tek (no need for the present). Similarly, any theory of time that accepts a past-presentist theory of temporal ontology can account for present-past relations in the same way -- although it cannot account for future events. However, this may be an advantage of such a theory since temporal relations between present objects and future objects are sometimes taken to be dubious.

Second, one might argue that *P2 is false*. This view is controversial. Perhaps Tek can account for his relation to past objects without positing their existence. For example, suppose Tek admires Newton. We might take this

as simply shorthand for admiring not the existent person (which does not exist) but for *the works and influence of Newton* (which does exist).

Exercise 16.107

Consider the topic of cross-temporal relations. The argument in support of MST is that we can only explain various cross-temporal relations (e.g., cause and effect, admiration, remembering, hoping) if we posit the existence of either past and future objects. One objection to this is that perhaps we can explain these relations without saying that past and future objects exist. Let's test this hypothesis. Suppose there is someone you idolize, admire, or love that does not currently exist. A critic of MST might say that you don't admire some person in the past (or future) but you admire your ideas of them or their effects. Does this capture what we mean when we say we idolize, admire, or love someone, or is it incomplete in some way?

16.0.2.4 Argument: The Composite Argument

There are several other arguments in support of MST. However, at this point, we will only consider one more.¹

In examining the arguments for a theory, it is sometimes the case that no single argument is particularly convincing but the collection of arguments is convincing. What supports the theory is not a single argument (or chain of reasoning from axioms) but a multitude of threads. With this in mind, we can develop a special composite argument. This is an argument that combines all three of the arguments that we have considered.

Theory	Intuitive	Past-tense	Past-future relations
Moving Spotlight	X	X	X
Presentist	X		
Eternalist		X	X
Growing Block		X	

In other words, we might combine all of the arguments for MST into a single composite argument:

- P1: A theory of time ought to explain (i) all of our intuitions (ii) truth-makers for past-tense and future sentences, (iii) cross-temporal relations, and (iv) how time passes.
- P2: MST does the best job of explaining or “fitting” (i)-(iv)

- P3: We ought to select a theory that best fits with (i)-(iv).
- C: Therefore, MST is true.

Exercise 16.108

Take one of the arguments in favor of MST, articulate it the best you can in your own words.

16.1 ARGUMENTS AGAINST MST

There are an assortment of problems with MST. However, our focus in this chapter will be on McTaggart's paradox.

16.1.0.1 *Objection - McTaggart's Paradox*

In an earlier chapter, we noted that one way to evaluate the quality of a metaphysical theory is to consider whether it is consistent with logical laws. A serious problem for any metaphysical theory is if it entails a contradiction. In this section, we will consider one of the most-discussed objections to MST. This is known as "McTaggart's paradox". This paradox is introduced by the idealist philosopher J. M. E. McTaggart in a 1908 article "The Unreality of Time" (published in *Mind* 17: 457--73). While McTaggart does not discuss MST in the same way we have expressed it here, contemporary philosophers take the paradox to primarily apply to MST. The paradox implies that MST is not merely false, but that it is an inconsistent theory. Since McTaggart's paradox is rather complex, let's unpack the argument in a step-by-step manner.

First, if MST is true, then the passage of time involves change in what events possess the properties of being past (P), being present (N), and being future (F). For example, say that event *E* is in 2024 and the present moment is 2022. If this is the case, then we would say the following:

1. Event *E* has the property of being in the future

Second, as time passes, event *E* will ultimately take on the property of being in the present. We can thus say the following:

2. Event *E* has the property of being in the present.

Third, as more time passes, event *E* will take the property of being in the past. We can thus say the following:

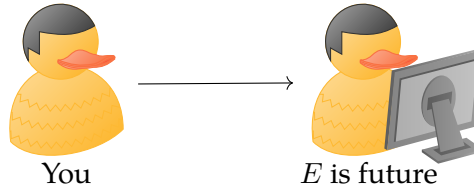
3. Event *E* has the property of being in the past.

Fourth, if you believe that the properties of being past, being present, and being future are exclusive, then no event can be past, present, and future. That is, it would be a contradiction to say that 2024 is past, present, and future.

4. No event E has the property of past (P), present (N), and future (F).

But here is the problem. Notice that we established that MST accepts (1)-(3) and also that (4) is the case. But, (1)-(4) is inconsistent. So, MST leads to contradiction and so the theory cannot be true. Before putting this argument in standard form, let's work through a concrete example.

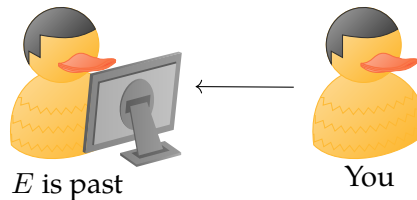
Suppose you plan to watch a video on your computer. At least from your perspective, you are not currently doing this so this event (call it E) is in the future. It is thus true to say that " E has the property of being in the future" or " E is future" for short.



Now suppose that you are currently watching a video. Since E is in present, it is true to say that " E has the property of being in the present" or " E is present".



Now suppose that you have finished watching the video and walk away from your computer. It is thus true to say " E is past".



So, we have established that it is true to say that E is future, present, and past. However, this is impossible since no event can be future, present, and past. With this concrete example in place, let's express this argument in standard form (see [81][102, p. 33]) as follows:

- P1: If MST is true, then the passage of time involves change in what events possess the properties of being past (P), being present (N), and being future (F).
- P2: No event E is P, N, F (incompatible properties)
- P3: Suppose event E is F.
- P4: Suppose time passes (as MST says it does) and E becomes N, and then becomes P.
- P4: Therefore, E is P, N, F. Contradiction with line 2.
- C: Therefore, MST is false.

The obvious *response* from a spotlihter is to say that the critic is mistaken in their characterization about what occurs when time passes.

- It is not the case that E is P, N, F (that is, it is not the case that E is future AND E is present AND E is past)
- Instead, E possesses being P , being N and being F *successively* not at the same time (E possesses P, N, F at different times)

In short, A-properties are *not* had by events simultaneously but successively. But, what does it mean for an event E to have these A-properties successively? Let's try to explain what it means for an event E to have a property successively.

First, we cannot say E is future on Sunday, present on Monday, and then past on Tuesday. This would be wrong since the critic would simply say that this is an explanation of B-properties rather than A-properties. If the spotlihter says " E is future on Sunday", they are simply saying that E is later than Sunday (B-relation) not that E has the property of being in the future.

Second, one way to express this idea is to introduce *tense* into our description of events. For example, suppose we are watching a video in the present. We would say that the event E is present (present tense), *was* future (past tense), and *will be* past (future tense). After we have completed watching the video, we will say that E is past (present tense), *was* present (past tense), and *was* future (past tense). The thought then is that by adding tense to our description of an event, we remove the apparent contradiction.

The obvious problem with this approach is that it does not resolve the paradox. Let's consider an example. Suppose there is a baseball game in 2026 and we are in 2025 (the present). Let's call the baseball game E . In 2025 all of the following are true

- E will be N
- E is F
- E will be P

In 2026 E will take the following properties:

- E is N
- E is P
- E will be N

But notice that a contradiction emerges even when we add *tense* to our characterization of E . For example, it cannot both be the case that E will be present (N) and E is present (N).

In 2025 E will be P E will be N E is F
 In 2026 E is N E will be N E was F

A second, more promising way to clarify that *E* does not have these properties simultaneous but *successively* is to temporalize when *E* has a given A-property with another A-property. The insight into this approach is to reject the idea that once an event is present, it is *always* present. Instead, an event *E* is present at a time. For example, if *E* is in 2026 and it is presently 2025, one might say that *E* is future (F) *in the present* (N). Here we are specifying *when* *E* has the property of being present using another A-property.¹

Let's consider another example of this approach. Suppose that *E* is 2026 and it is now 2026. To characterize *E*, a spotlighter would say "*E* has the property of being present in the present." Or, more compactly, *E* is present in the present or *E* is present-present. This attempt to characterize what it means for A-properties to be had successively is sometimes called an appeal to "compound tenses" since we are no longer simply saying something is present but that it is present-in-the-present.

On this approach, we would characterize an event *E* in 2026 as follows:

In 2025 E is F in N E is N in F E is P in F
 In 2026 E is N in N E is P in F E is F in P

But, again, this does not solve the problem since it may again be argued that some of the combinations of the above produce a contradiction. For example, *E* cannot be both (1) future (F) in the present (N) and (2) present (N) in the present (N).

In 2025 E is F in N E is N in F E is P in F
 In 2026 E is N in N E is P in F E is F in P

The spotlighter's response is to say "yes. I see this contradiction, but you've made the same mistake. These properties are had *successively*. It is never the case that an event is future-present and present-present." To which the critic of the spotlighter again responds, "what do you mean by successively?" To which the spotlighter can again respond by temporalizing these second-order properties into third-order properties, e.g., *E* is F in N in the N.

The end result seems to be an endless back-and-forth of objection, followed by reply, followed by objection.

¹This approach is sometimes characterized as appealing to *second-order* A-properties as they make use of a A-property that characterizes an A-property. For example, *E* is not merely future but future-in-the-present (future present).

16.1.1 Another explanation

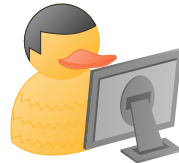
Let's consider a slightly inaccurate but perhaps more intuitive way of understanding the MST response.

While there is a contradiction if we say that an event E is both present, future, and past, there is no contradiction if we say that E is present, was future, and will be past. The solution from the spotlighter then is to *add tense* to the characterization of events in time. To see this more explicitly, consider different ways we might add tense to an event E depending upon whether it is present, past, or future:

if E is present in 2022	if E is in the future (2030)	If E is in the past (1980)
E will be past (future tense)	E will be past (future tense)	E is past (present tense)
E is present (present)	E will be present (future tense)	E was present (past tense)
E was future (past tense)	E is future (present tense)	E was future (past tense)
No contradiction	No contradiction	No contradiction

The next step of McTaggart's paradox is where things get both interesting and somewhat difficult to follow. One *counter-response* is to say that the *tense of a sentence* is a grammatical feature of language. We can thus *restate* tensed sentences in a way without their tense. The idea here is that we are taking a tensed sentence S and translating it into an equivalent untensed sentence S^* . This action is called *detensing* a sentence. When we are removing the tense, we are talking about the properties of an event as though they had these properties eternally (independent of time).

Let's consider an example. Suppose you are watching a video in the present. The spotlighter will say that the correct way to characterize this event E is that you are currently watching a video, this event will be in the past when you are finished, and this event was in the future



E will be past, is present, was future

The critic will argue that we can rephrase (without loss of meaning) what the spotlighter is saying by detensing. Each of the above sentences can be expressed in a detensed form as follows:

Tensed Sentence	Detensified
<i>E</i> will be past	<i>E</i> is past in the future
<i>E</i> is present	<i>E</i> is present in the present
<i>E</i> was future	<i>E</i> is future in the past

In fact, for any tensed sentence, it is possible to find an (equivalent) detensed version of the nine different tensed sentences that can be constructed from “*E* is past”, “*E* is present”, and “*E* is future”.

Tensed Sentence	Detensified
<i>E</i> is present	<i>E</i> is N in N (present in the present)
<i>E</i> was present	<i>E</i> is N in P (present in the past)
<i>E</i> will be present	<i>E</i> is N in F (present in the future)
<i>E</i> is past	<i>E</i> is P in N (past in the present)
<i>E</i> was past	<i>E</i> is P in P (past in the past)
<i>E</i> will be past	<i>E</i> is P in F (past in the future)
<i>E</i> is future	<i>E</i> is F in N (future in the present)
<i>E</i> was future	<i>E</i> is F in P (future in the past)
<i>E</i> will be future	<i>E</i> is F in F (future in the future)

But now notice what occurs as time passes. As time passes, events (*E*) possess all nine of these temporal properties. For example, suppose 2020 is the present and 2030 is the future. Using the same line of reasoning as above, it would follow that:²

- 2030 has the property of being future in the present.
- (Later) 2030 has the property of being present in the present
- (Later) 2030 has the property of being past in the present

But 2030 cannot both be future in the present, present in the present, and past in the present. This is the same contradiction as before when we said an event (*E*) is past, present, future. In sum, the contradiction created in the argument reemerges.

At this point, the spotlighter might respond by saying that we can remove the contradiction by taking these new temporal properties (e.g., present in the future) and *adding tense to them*. Their response will be something like this:

- An event *E* is never future in the present AND present in the present AND past in the present. This is contradictory!

- Instead, *E was* future in the present, *is* present in the present, and *will be* past in the present

But the critic of the spotlighter will quickly object to this reply. They will contend again that tense is a *grammatical feature of language* and insist that it only hides the contradiction that is a part of MST. They will again detensify the sentence, and then point out that the contradiction reemerges.

Since McTaggart's theory is somewhat complex, let's briefly summarize the main results:

1. McTaggart's paradox seems to show that MST leads to contradiction.
2. The MST theory has a reply (adding tense), but that reply is only temporary since the critic can comeback with another argument showing MST to be inconsistent.
3. We are left with a theory that leads to an infinite series of criticisms and replies.

16.1.1.1 *Composite Objection against MST*

Similar to the composite argument for MST, a composite objection can be raised against MST.

In argument standard form:

- P1: For a theory to be true, it ought not to be prone to a more serious objections / problems than its competitors.
- P2: MST is open to more serious objections / problems than its competitors. (1) epistemic argument, (2) open future argument, (3) McTaggart's paradox.
- C: Therefore, MST is false.

Exercise 16.109

McTaggart's argument against MST is quite complex. In your own words, try to simplify this argument.

16.1.2 *Summary*

For some, MST is an attractive theory of time since it accords with many of our intuitions when taken individually. People talk of the past, present, and future, and so one way of making sense of this distinction is to posit the existence of such times. In addition, the present moment seems to metaphysically specially in some way. It strikes us as somehow singular, important,

and more vibrant than other times. MST also can account for what makes past-tense sentences true (truthmakers) along with temporal relations between past and future objects.

However, MST also presents some problems. Given the fact that the present moment is non-empirical, it is unclear how we would ever know we are in the present moment. Proponents of non-compatibilist views of freedom will also object to MST on the grounds that individuals are not free in the compatibilist's picture of the world. And, finally, one distinctive problem for MST is that its synthesis of the A-theory with the eternalist ontology gives rise to McTaggart's paradox.

Finally, it is important to point out that we have only considered a small sampling of arguments for and against MST. There are many other arguments for MST as well as against MST.

FURTHER READING

For additional reading on MST, see [[18](#), [17](#), [35](#), [34](#), [83](#), [122](#), [62](#), [122](#), [121](#), [120](#), [123](#), [128](#), [135](#), [89](#), [61](#), [42](#), [29](#), [28](#), [22](#), [48](#), [80](#), [79](#), [77](#), [40](#)]

We saw that one of the principal challenges to presentism was its seeming inability to account for past truths. While the Moving Spotlight theory was able to address this problem with its eternalist temporal ontology (past, present, and future exist), this theory entails that the future already exists.

Rather than revising the definition of truth, trying to revise what it means for truths to be “grounded” in reality, or trying to ground past truths in present facts, another approach is to abandon presentism in favor of an alternative theory of time. One such theory is the growing block theory (GBT).

Definition 17.1: Growing Block Theory (GBT)

Time consists of the following components:

1. A-theory
2. Past-presentist temporal ontology
3. Ontological-change dynamic theory of temporal passage: time passes because what exists (the events) or the instantiation pattern of what exists changes.

As a theory, GBT is strikingly similar to presentism in that it is an A-theory and it denies the existence of the future (there are no future objects or events). There are two key differences.

First, in contrast to the temporal ontology of presentism, GBT takes the past to exist in just the same way as the present. Not only do you and I exist, but so do dinosaurs. Not only do present objects have the properties they do, but so do past objects. In other words, in contrast to presentism where existence is coextensive with the present, GBT takes existence to be coextensive with the past and present.



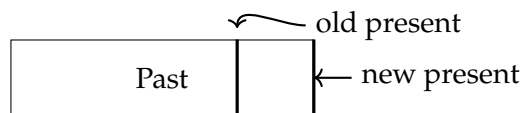
Given their acceptance of the past, GBT can overcome the grounding objection for past truths. Namely, they can hold the “past record truthmaking” principle:[84][18]

Definition 17.2: past record truthmaking

Past-tensed truths (e.g., “there were dinosaurs”) are made true by the way the world is, in the past.

So, unlike presentism which cannot posit the existence of past objects as truthmakers for past-tensed sentences, GBT can ground such truth by positing past objects and past instantiation patterns.

Second, similar to presentism, GBT takes the passage of time to be a feature of the world: new objects come into existence (or existing take on new instantiation patterns). However, in contrast to presentism, objects never go out of existence. In other words, in contrast to presentism where temporal passage is modeled in terms of addition and deletion, GBT models temporal passage wholly in terms of addition: change occurs by new objects (time-slices) being added to the block.

**17.0.1 Arguments in favor of GBT**

What are the arguments in favor of GBT?

First, the theory is intuitive in that GBT says (1) the present really exists and (2) time passes. In addition, it is *more intuitive* than the Moving Spotlight Theory (MST). Recall that MST contends that (1) the future exists and (2) that change is only a matter of existing events gaining and losing temporal properties. GBT by contrast (1) denies that the future exists and (2) posits change in terms of objects coming into existence and / or gaining new properties.

Second, insofar as GBT accepts the unintuitive idea that the past exists, it does so for good reason. When discussing presentism, we saw that presentists had difficulty with grounding contingently true propositions about the past. In contrast, GBT can ground such propositions using past events.

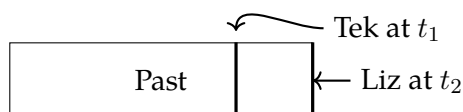
17.0.2 Arguments against GBT

The primary objection to GBT is the *epistemic objection concerning the present*. The epistemic objection is that if our world is a growing block, then we cannot know whether we are in the present or in the past. Let's consider an

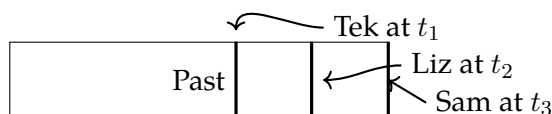
example of this problem. Suppose two penguins Tek at time t_1 and Liz at t_2 , where t_2 is later than t_1 .



Next, notice that their subjective situation of each is *nearly* identical: each likely believes they are in the objective present, each can cite their conscious awareness of events happening “now” as evidence that they are in the present, and each can cite the fact that they have no access to times happening *later* than the time they are experiencing at this moment. However, one difference between Tek and Liz is that, assuming they are in the same world, Liz can point out that she exists at a time t_2 later than Tek at t_1 .



However, notice that Liz at t_2 does not know if her subjective situation is exactly like Tek's at t_1 . Namely, Tek does not know that there exists a time t_2 that is later than the time he occupies. He does not know that Liz exists later in the block. But notice that Liz's situation is entirely the same as Tek's. She does not know if there is a third person (Sam) who exists at a time t_3 even later on in the block.



Let's formulate this problem in another way. Consider that there are two uses of the term “now”:

1. The indexical use: The use of “now” that picks out the time at which the speaker is speaking.
2. The objective use: The use of “now” is that picks out the objective present.

Now suppose a time t_i . A person may utter two sentences:

1. I am dancing now (indexical use).
2. I am dancing now (objective use).

The epistemic objection claims that while a person may know that (1) is true, they have no way of knowing whether (2) is true because the subjective situ-

ation between a person in the objective present and a person in the objective past are indistinguishable. In other words, while we can know that a time t is now (in the indexical sense), we cannot know whether this now is *now* (in the objective sense). As Miller[84, p. 224] puts it,

“The problem for the growing block theory is that while I can know that it is now, in the indexical sense, I cannot, it is argued, know that it is now now, where the first ‘now’ picks out the indexical and the second ‘now’ picks out the objectively present sense of the term.”

Even further, consider that the present is fleeting, lasting only a fragment of a second. If we accept that there are past objects, then have been many prior time-slices. If that is the case, what is more likely about the indexical ‘now’ you currently find yourself in: (1) it just so happens that it is the objective present or (2) it is one of the many past time slices?

Let's summarize this argument:

- P1: For any theory of time T that contends there is an objective present, then it ought to be epistemically possible to know whether you are in the present (it should be epistemically possible to know whether now is now).
- P2: If GBT is true, then it is not epistemically possible to know whether you are in the present (there is no way to know whether now is now since the two are subjectively indistinguishable).
- C: Therefore, GBT is false.

Exercise 17.110

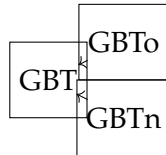
Articulate the epistemic objection to GBT.

17.0.3 Responses to the epistemic objection

There are two main responses to the epistemic objection, resulting in two different versions of GBT. The first response is that even though the present and past are subjectively indistinguishable, there is still a way to know whether you are in the present as opposed to the past (there is a way to know whether the now is now). This version of GBT is known as “old growing block theory” (hereafter GBTo).

The second response is to deny that the the past and the present are subjectively indistinguishable. In other words, there is something about our experience of the objective present that allows us to know that we are in the

present (we can detect whether the now is now). This version of GBT is known as “new growing block theory” (hereafter GBTn).



Let's consider each version of GBT.

17.0.3.1 Old GBT

Here is how one might argue for GBTo. First, we need to introduce some terminology from epistemology (the philosophical study of knowledge). Philosophers begin their analysis of “knowledge” by saying that a subject *S* knows a proposition *P* when three conditions are met:

1. *S* believes *P*,
2. *P* is true,
3. *S* has justification for their belief in *P*.

In short, knowledge is justified true belief (JTB). Generally, it is thought that a belief is justified provided certain conditions are met. There are two main theories:

1. Internalism: knowledge is justified solely by internal factors.
2. Externalism: knowledge is not justified solely by internal factors but instead involves some external factors.

Internalists claim that a belief is justified solely by one's internal (mental) states. For simplicity, we will treat internal states as *reasons* accessible within the mind of a subject for why a particular proposition is true.

Definition 17.3: internalism

Internalism asserts that justification of belief depends solely on internal mental (subjective) states, e.g., awareness of evidence, reasoning, introspection.

For example, suppose I believe that there is a barn in a certain field. My belief would be justified because I can access a reason for why that belief is true, e.g., I visited the barn yesterday.

Alternatively, suppose I believe that stock prices will go up when it is announced that interest rates will be cut. Now if I have no good reason for this

being true (e.g., I have not analyzed this in the past, or can't recall an expert telling me this is the case), then my belief is *not* justified.

Or, to use another example, consider your own existence. If you try to doubt your own existence, you only end up presupposing that you exist. So, it seems that it is impossible for you to doubt your own existence (your existence is indubitable). This short argument provides an internal justification for your belief that you exist.

In contrast to internalism is externalism. Externalists contend that justification does not solely depend upon internal states. That is, *S*'s belief that *P* may be justified even if one has no internal grasp of a reason that would make it justified. As the name implies, externalism says that justification can come from features *external* to the mind or from being in a specific relation to the *external* world.

Definition 17.4: externalism

Externalism asserts that justification of belief depends on a subject's relation to certain external states, e.g., belief-forming processes due to the world.

There are several varieties of externalism. One is the causal theory. It says that *S*'s belief that *P* is justified provided *P* is appropriately causally connected with the fact that *P*.

Another is process reliabilism. This theory says that *S*'s belief that *P* is justified provided the belief was generated by a *causal process that reliably produces true beliefs*.

A few illustrations. First, let's consider a case of a belief that is not appropriately connected to a causal process that produces true beliefs.

Consider the case where a head injury causes someone to believe. Suppose Tek is a professional fighter. He is knocked out in a fight, but does not remember being knocked out. The head injury causes his brain to produce a several beliefs: (1) aliens exist and they take over human beings and (2) he was recently knocked out in a fight. Tek's beliefs are caused by the knockout but they are not due to a causal process that reliably produces true beliefs.

Next, let's consider a case where a subject's belief that *P* is appropriately causally connected to the fact that *P*.

Suppose I learn that "Napoleon is 5'7" from reading it in a textbook and then measuring the dead body of Napoleon. Many years later I forgot how I learned this information but nevertheless believe that "Napoleon is 5'7". The

externalist will say that my belief is justified even though I don't have any accessible reason for why it is true. This is because my belief that Napoleon was 5'7 is causally connected in the right kind of way with the fact that Napoleon was 5'7.

One more example. Suppose three people: Tek, Liz, and Jon. All three want to know the temperature of a substance. Let's suppose they obtain the temperature of a substance in the following way:

1. Tek: working thermometer 32 degrees
2. Liz: broken thermometer 32 degrees
3. Jon: random-number generator 32 degrees

Clearly, Jon's belief is not justified. His use of a random number generator is not causally connected to the temperature at all. But what about Tek and Liz? Let's suppose Tek's thermometer is broken in such a way that it acts like a random number generator. So, while Tek and Liz may have the same belief (the substance is 32 degrees) and the same mental states / reasons with respect to their beliefs ("my thermometer reads 32 degrees"), only Tek's belief is justified because only his belief is appropriately causally connected to the actual temperature of the substance.

Exercise 17.111

Create an example similar to the thermometer example. In your example, you should have two people that have the same belief and the same "internal" justification for their belief but where only one person is justified via external factors.

17.0.4 GBT and externalism

Let's return to the discussion of GBT. At first glance, the epistemic objection to GBT seems to follow if internalism is true. Take two individuals at two different times: Tek at t_1 and Liz at t_2 . It seems that the internal states of both persons are, for all relevant purposes, subjectively indistinguishable. They both believe they are in present and come to this conclusion on the awareness of the same evidence. This evidence is that (1) they have no access to any future times and (2) they both know that they exist at a time. So, given the way things are for Liz and Tek is the same, if we are internalists, we ought to assign equal credence to their claims that they are in the present. [84, p. 226]

However, let's accept externalism instead of internalism. Let's say that knowledge is *justified true belief* and a belief is justified iff it is caused in a reliable manner. And, let's say that one such way my belief that P is caused in a reli-

able manner is if my belief that P is caused by the thing that P is about. So, for example, my belief that there is a house in front of me is justified if that belief is caused by the house (or my looking at the house).

Now consider a person in the objective present that believes they are in the objective present. First, my belief that I am in the present is justified since it is caused by my being in the present. Second, my belief is true (since I am in the present). Therefore, my belief that I am in the present should count as knowledge.

Contrast this with a person in the objective past that believes they are in the objective present. Their belief is justified because it was caused by their being in the present (which is now a prior present). However, their belief is no longer true and so does not count as knowledge.

So, provided externalism is true about justification, then even if the present is subjectively indistinguishable from the past, it is wrong to say that a person cannot have knowledge they are in the present. Provided their belief is caused by the present and they are in the present, they know they are in the present.

17.0.4.1 *Objections to GBTo*

There are several objections to this approach.

First, one objection to this view is that it makes an implausible assumption about how reliable beliefs are formed. Suppose FRED believes he is in the present. If the above is true, then FRED's belief is caused by FRED being at the latest slice of the growing block (let's say that FRED endures and so is wholly found there). However, this would mean that the object (FRED) that causes the belief (FRED's belief that he is in the present) are occurring at the same time. This is problematic since we typically believe that the object that causes the belief occurs before the belief.¹

Second, let's consider a person's knowledge over time. Suppose Liz is in the present and believes she is in the present. On our account, this belief is counts as knowledge since (1) Liz believes it, (2) it is true, and (3) her belief

¹As Miller puts it, "consider the token belief 'FRED is in the objective present'. If the belief that FRED is in the objective present is caused by FRED being in the objective present, then the belief is justified. Clearly, though, if [sic] can only be that FRED's being in the objective present causes the belief that FRED is in the objective present, on the assumption that causation is, in this case, non-local (i.e. on the assumption that the cause is not temporally separated from the effect). That is, indeed, a controversial assumption." [84, p. 233]

is justified on externalist grounds. But consider the next second. Liz's belief is no longer true and so no longer counts as knowledge.

Third, there is always the chance that the cure is worse than the disease. In order to avoid the epistemic objection, externalism about justification is accepted over internalism. However, externalism itself has several notable problems (independent from the philosophy of time). To point to just one such problem, consider the following case from Bonjour (1980):

Norman, under certain conditions that usually obtain, is a completely reliable clairvoyant with respect to certain kinds of subject matter. He possesses no evidence or reasons of any kind for or against the general possibility of such a cognitive power, or for or against the thesis that he possesses it. One day Norman comes to believe that the President is in New York City, though he has no evidence either for or against his belief. In fact the belief is true and results from his clairvoyant power, under circumstances in which it is completely reliable. (p. 21)

Intuitively, we think that Norman's belief is *not* justified. Norman has no reasons for thinking the President is in New York. If we were to ask Norman himself whether his belief is supported, he might reflect for a few moments and say that since he has no good reason for thinking the President is in New York, this belief is unsupported. Nevertheless, on externalist grounds, Norman's belief would be justified since it is linked to a completely unknown and potentially unique reliable process.

Exercise 17.112

A challenge for someone who supports the GBT is to try to respond to the epistemic objection in some reasonable way. We examined one way to do this by appealing to externalism about justification. There seems to be two ways forward. First, one could argue that the externalist criteria for justification we have articulated above are not the right kind of externalist criteria. So, Tek knows he is in the present but not because the belief is due to X (where X is something in the world). A second approach is to say that we know we are in the objective present but not due to externalist reasons. In other words, it is subjectively possible to know that you are in the present (so argue that GBT is true on internalist grounds).

17.0.4.2 New GBT

Another approach to epistemic objection is to argue that two persons (one in the present and one in the past) are not subjectively indistinguishable. In saying that they are not subjectively indistinguishable, what is not being asserted is that individuals that find themselves in the objective present experience the world differently than those in the past. Instead, what is being asserted is that they have access to evidence that allows them to assert they are in the objective present.

One way to argue for new GBT (GBTn) is to say to say that there is a difference between past and present objects. On this approach, there is *no way that past object are* (they exist but do not instantiate properties). Instead, the only objects that instantiate properties are present objects.

How then are past truths grounded? One answer is that presently existing objects have “temporal distributional properties” (TDPs). To illustrate, take the sentence “Tek was 5'0”. On the TDP approach, the truthmaker of this sentence is not a past object that instantiates this property but a presently existing Tek who instantiates the TDP of *being 5'0 at t_1* . To speak metaphorically, properties held by past objects are “imprinted” as TDPs on their presently existing counterparts.

If the above is the case, then we have reason to reject that past persons and present persons find themselves in the same subjective situation. This is because there is no way things are for past persons since past objects no longer instantiate any properties.

Let's consider three criticisms of this approach.

First, since only present objects instantiate properties, past objects do not play a role in grounding past-tense truths. As such, when Tek perishes, there seem to be three possible consequences: (1) past-tensed truths about Tek are no longer grounded (violating the grounding principle) or (2) Tek's corpse (ashes) will instantiate TDPs of all his previous heights at previous times. Both options clash with intuitions.

Second, GBTn might be charged with positing “spooky objects”. According to GBTn, past objects exist but they are entirely lacking in properties. This would clash with the bundle theory of objects, which contends that an object is nothing but a compresent collection of properties, or with Armstrong's theory of universals which contends that there are no bare particulars.

Third, Miller[84, pp. 242–243] points out that this version of GBTn fails to substantially distinguish itself from the presentist. The only difference be-

tween the two theories is that GBTn asserts that past objects still exist (although lacking any properties). All of the work done in terms of accounting for why past-tensed sentences are true is done by presently existing objects. And so, it isn't clear why GBTn is in any way different than a presentist who makes the same claims without appealing to property-less past objects. In fact, presentism appears preferable given that it is ontologically more parsimonious.

This third criticism raises a dilemma for both the presentist and the supporter of GBT. One motivation for moving away from presentism to GBT is that past truths need grounding (the grounding objection). By positing the existence of past objects and instantiation patterns, GBT has a straightforward way of saying how past-tense sentences are true. However, the incorporation of the past into one's ontology comes with the consequence of not being able to know whether one is in the object present (the epistemic objection). One response has been to adopt an externalist epistemology, but this comes at the cost of saying that the cause and the effect are simultaneous. Another response has been to maintain an internalist epistemology but to resort to saying that only present objects have properties and these objects are what is making past-tense sentence true. This latter type of maneuver raises the question of why do we need the past at all? In other words, in an effort to address the epistemological argument, GBT collapses back into presentism. GBT thus appears to be stuck between two choices:

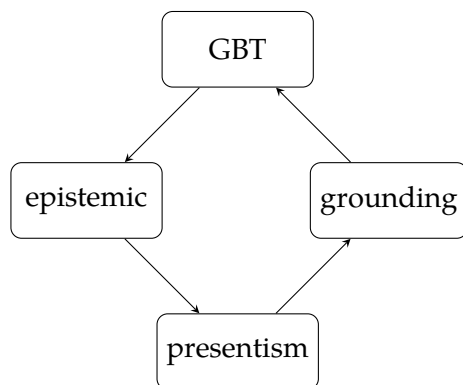
1. Unified approach: Acknowledge the existence of past objects that have properties just like present objects. This approach ensures past-tense sentences meet the grounding objection but opens the theory to the epistemic objection.
2. Disunified approach: Claim that there is something different about present objects and their instantiation pattern from past objects. This approach meets the epistemic objection but opens the theory back up to the grounding objection.

17.1 THE GBT-PRESENTIST PARADOX

Let's try to formulate a paradox involving the growing-block theory (GBT) and presentism. Let's suppose the only two theories that are true are GBT and presentism. However, if presentism is true, then there is no way to know anything about the past since there are no past facts (the grounding problem). But, there are past facts, so presentism must be false. Since GBT can account for the fact that we can know about the past, GBT must be true.

However, if GBT is true, then the present would exist but there would be no

way to know that one is in the present moment. But, if the present exists, then we should know whether or not we are in the present. So, GBT is false. Since presentism can account for the fact that we know that we are in the present, presentism must be true.



In the previous chapters, the moving spotlight theory and the presentist theory of time were introduced and evaluated. In this chapter, we will consider our third theory of time: eternalism. Just as in the previous chapter, our plan will be to (1) articulate presentism, and then (2) consider arguments for and against presentism.

18.1 WHAT IS ETERNALISM?

Let's begin by defining eternalism:

Definition 18.1: eternalism

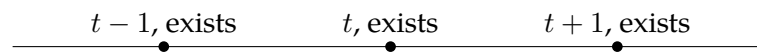
Eternalism is a theory of time that consists of the following components:

1. It is an B-theory
2. An eternalist temporal ontology
3. Static theory of temporal passage

Let's briefly articulate each of the different components of eternalism. First, eternalist accepts the B-theory. This means that events in time are completely structured by B-relations (earlier than, later than). As such, while events are related to other events by being earlier and later than them, no events are objectively present, past, or future. In other words, while people may talk as though events as “in the past” or “in the present”, events are not objectively in the past, present, or future.

Second, the eternalist accepts the eternalist view on temporal ontology. This view is sometimes called permanentism or entity-everywhere-ism. This component of eternalism is sometimes expressed as follows: *It is always the case that everything that exists does so eternally.*

Understanding this aspect of eternalism can be somewhat difficult. To aid in its understanding, suppose that there are three events in time. First, there are the objects that exist here and now: the time we take to be the present moment. Let's call this time t . Next, suppose there is an event one year earlier than t . Let's use $t - 1$ to refer to this event. Finally, there is an event that occurs one year later than t . Let's refer to this event as $t + 1$. For the eternalist, none of these events are more real than the other. The times $t - 1$, t , and $t + 1$ all exist.



To further aid in understanding this point, eternalists will sometimes make use of an analogy.

Suppose you are in the United States, specifically State College, PA. Certainly, you would accept that the objects and events that occur in State College exist and are real. But there are other places that are a certain *spatial distance* from you. Let's consider two others: Paris and Toronto. Now let's suppose the following:

1. You exist in State College, PA
2. Tek exists in Paris
3. Liz exists in Toronto

Intuitively, the fact that you are a certain spatial distance from Tek and Liz does not make them any less real. The fact that Tek is far away from you does not cause him not to exist. This is the case even though you cannot physically interact with him. To summarize the above, we might say events and objects are *ontologically space-insensitive*. The idea behind space-insensitivity is that no spatial location is metaphysical privileged. The mere fact that an object *X* is some specific spatial location does not make it *more real* than some other object in a different spatial location.

Definition 18.2: space-insensitivity

An object's precise position in space does not determine (condition) whether that object exists.

We can understand the eternalist's view on temporal ontology if we apply these same considerations to time. Let's illustrate this with an example:

1. You exist in 2024
2. Tek exists in 2030
3. Liz exists in 2020

Again, note that there are certain times t_2 and t_0 that have a **temporal distance** from you: earlier and later. The eternalist says that just like with space the fact that a time is a certain distance away from the time you occupy does not make it any less real. Events then are *ontologically time-insensitive*: no temporal location is metaphysical privileged. The mere fact that an object *X* is some specific temporal location does not make it *more real* than some other object in a different temporal location.

Definition 18.3: time-insensitivity

An object's precise position in time (earlier, later, etc.) does not determine (condition) whether that object exists.

Finally, consider (3). The eternalist adopts a static theory of temporal passage. What does this mean? There are two different accounts.

On the **first** account, it means that there is no change to what exists and time does not pass. In other words, no new objects are being added to the universe and the objects that do exist are not gaining or losing characteristics or traits.

On the **second** account, there is change to what exists and the properties of things that do exist even though time does not pass. This is sometimes called the *F-theory*. On the F-theory, the world looks exactly the same as the static approach but this approach contends that we ought to rethink how we think about change. Objects do change but they change through having and not having properties at different times. We will talk about this more when we get to one of the arguments for and against eternalism.

18.1.1 Arguments Against Eternalism

Now that we have briefly articulated eternalism, let's evaluate the theory. While popular among philosophers of time, eternalism is undoubtedly the least intuitive theory. In an earlier chapter, it was noted that if a theory runs counter to our intuitions. With this in mind, let's start by articulating some arguments *against* eternalism and consider how the eternalist might respond to these criticisms. After considering arguments against eternalism, we will then consider several reasons *in favor* of eternalism.

18.1.1.1 Argument - Denial of the Present

Eternalism is a B-theory and so eternalist deny that there is an objective present, past, or future. This clashes with the common-sense belief that the present exists, that language about the present is meaningful, and that sentences about the present are true (e.g., "I am in the present"). Given these two items, eternalism might be criticized in the following way:

- P1: If eternalism is true, then there is no objective present.
- P2: If there is no objective present, then sentences that refer to the present are meaningless and are all false.
- P3: Sentences that refer to the present are meaningful and are true.
- C: Therefore, eternalism is false.

Here is another way of putting the spirit of the above argument. How is the eternalist supposed to make sense of all of our A-beliefs and A-language about the world? First, let's suppose that sentences that refer to the present, past, and future as ``A-sentences". Second, note that we intuitively believe that some A-sentences are true (e.g., ``2022 is present" is true). Third, since truth has a relation to the world (truth supervenes on reality), if some A-sentences are true, then this seems to entail that the world contains something that makes these sentences true (truthmakers). We can call the truthmakers that make A-sentences true **A-truthmakers**. With this in place, we can express an argument against the B-theory.

- P1: There are some true A-sentences.
- P2: There must be something in the world (truthmakers) that makes these sentences true.
- P3: The only thing that would make these sentences true are A-truthmakers (A-facts).
- C: Therefore, A-truthmakers (A-facts) must exist.

To illustrate P1, consider the following A-sentences:

- ``The present is now"
- ``You are in metaphysics class in the present"
- ``You learned about growing block theory in the past."
- ``You will learn about personal identity in the future"

On this account, there are A-sentences and the truthmakers of these A-sentences are A-truthmakers (A-facts, presentness, pastness, futureness).

Since the eternalist is a B-theorist, it must contend that there are no A-facts / A-truthmakers. The eternalist has two options:

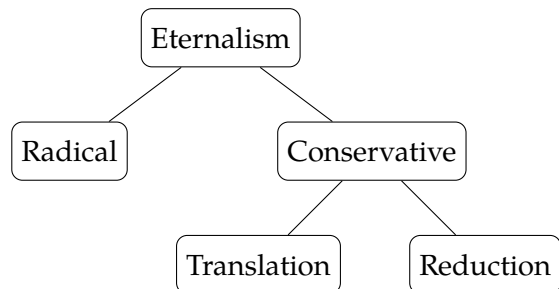
1. **Radical Solution:** A-language is nonsense (meaningless) or false.
2. **Conservative Solution:** A-language is meaningful. A-sentences can be true or false, but we have to understand what A-sentences really mean (translate them correctly) OR understand what makes A-sentences true.

Each of these two solutions will assert something different about the argument against eternalism. The radical solution will deny that there are true A-sentences, claiming that sentences that refer to the present lack meaning and so cannot be true or false. For example, the eternalist might say that just because "a four-sided triangle is an interesting shape" or "colorless green ideas sleep furiously" appear to be meaningful does not mean that they are meaningful. These sentences contain meaningful ideas but when they are combined together, they do not form a meaningful complex unit. With respect to (2).

The conservative solution is more complex. They contend that while A-sentences are meaningful, it is *not* the case that the only thing that would make these sentences true are A-truthmakers (A-facts). It will thus assert that P3 is false. In other words, eternalists can respond by saying that even though there is not an objective present moment, it does not follow that sentences that refer to the present are meaningless or neither true nor false. Rather, what makes A-sentences true or false are B-sentences or B-facts.

In what follows, let's focus our attention on the more conservative solution. Namely, we want to assert that A-language is meaningful (can be T or F) but there are no A-truthmakers. There are two ways to try to show there can be true A-sentences but no A-truthmakers:

1. **Approach 1 (Translation):** A-sentences can be translated into B-sentences without a loss of meaning. These B-sentences have B-truthmakers and so A-truthmakers are not necessary to account for the truth of A-sentences. In short, every A-sentence has a corresponding B-sentence that *means* the same thing.
2. **Approach 2 (Reduction):** A-facts are reducible to B-facts. In other words, A-sentences have B-truthmakers. For any A-sentence, what makes it true is not an A-truthmaker but a B-truthmaker.



18.1.2 Approach 1: Translation

The *translation account* contends that we can take every A-sentence and translate them into a corresponding B-sentence without loss of meaning. Thus, if we can translate every A-sentence into a B-sentence, then we don't need the A-truthmakers. In other words, we don't need A-facts / A-truthmakers (presentness, pastness, or futureness) to state what facts in the world make A-sentences true.

Let's look at a few examples. First, consider the following two sentences:

- S1: E is in the future (A-sentence)
- S2: E is later than t (B-sentence)

According to the translation account, we can translate S1 into S2. When someone says S1, what they mean is S2. So, if Tek says "my future life will be great", this sentence can be translated into the B-sentence "there is a time t1 later than the time at which I am uttering this sentence t0, such that my life at t1 is great."

Here is a second example. Consider the following sentences:

- S3: E is in the present (A-sentence)
- S4: Event E occurs at the same time as the utterance of E (B-sentence)

We can translate S3 into S4. When someone says S3, what they really mean is S4.

However, the main problem with the translation approach is that this translation approach *does not seem to work* since A-sentences are context-sensitive but B-sentences are not. To see how this works, consider the following two sentences:

1. I am hungry now.
2. I am in New York.

The meaning of both sentences depends upon the context in which they are used. For example, if I say (1) in 2025, then it means "David is hungry in 2025". But, if you say (1) in 2026, then it means "[Your Name] is hungry in 2026."

In contrast, B-sentences and our translations of (1) and (2) are not context-sensitive. Consider the following:

3. 2025 is earlier than 2026.
4. 2026 is later than 2025.

The meaning of (3) and (4) does not depend upon who said them or when they were said, unlike sentences (1) and (2).

So, given that A-sentences are context-sensitive and B-sentences are not, there seems no way to translate A-sentences into B-sentences without losing some aspect of their meaning.

18.1.3 Approach 2: Reduction

The previous attempt by the B-theorist to reduce the A-theory to the B-theory involved the following:

- Take an A-sentence
- Translate it into a B-sentence

- Unpack the truthmaker of the B-sentence using a B-truthmaker (relation between event)

Another way to account for the fact that A-sentences are meaningful, can be true, and have truthmakers is to *simply identify a B-truthmaker for every A-sentence*. That is, we do not need the controversial claim that the meaning of every A-sentence can be captured by a corresponding B-sentence. Instead, all we need is to show that every A-sentence can be made true by a B-truthmaker.

On this approach, there are meaningfully true A-sentences, but this does not commit us to there being A-facts.

To really understand what is going on here, we need to introduce distinguish between a **type** and a **token**.

- A type: general thing, abstract, a concept
- A token: a physical or concrete instance of something.

Consider the following sentences:

- S1: The man rode the bicycle.
- S2: A rose is a rose is a rose.

In S1, there is one word "the" (type) but there two tokens of that type in S1. There is one general thing but two instances of that general thing. In S2, there is one word "rose" but there are three tokens of that type in S2.

With the type-token distinction in place, the B-theorist's plan is to **devise a B-fact for every token A-sentence**. That is, we can't do the *type* A-sentence translation because the meaning of "Today is sunny" changes based on when it is uttered ("today" is context sensitive). So, instead of type-type translation, we instead opt for a token-token translation. This can be performed in at least two ways:

1. **Token-reflexive method:** put the event E in a B-relation with a token of the A-sentence (put it in relation to the event of uttering the sentence)
2. **Date-relation method:** put the E in a B-relation with some other event, e.g. a date

Let's look at some examples. Consider the following sentence:

- S3: "Metaphysics class is now" (A-sentence)

Now we can specify a B-truthmaker for this sentence using both methods

- B-truthmaker: S3 is true iff metaphysics class occurs at the same time as the token (utterance) of S3.

- B-truthmaker: S3 is true iff metaphysics class occurs on date you uttered S3 (28 Jan).

We can account for the truth of the A-sentence by using a B-truthmaker! Let's consider another example. First, we begin with a sentence:

- S4: "Metaphysics class is tomorrow" (A-sentence)

Next, we pick either one of our methods for obtaining the B-fact that makes this A-sentence true:

- B-truthmaker: S4 is true iff metaphysics class occurs a day later than the token (utterance) of S4.
- B-truthmaker: S4 is true iff metaphysics class occurs on date X (X=date a day after the day you uttered S4).

In other words, our new approach is to take A-sentences and specify a B-truthmaker that makes that A-sentence true. We avoid the intermediary step of trying to find an adequate translation of an A-sentence into a B-sentence. B-theorist: If A-sentences are true, then what makes them true are not A-truthmakers (an event being present) but B-truthmakers.

- S5: "Metaphysics class is in the present" (A-sentence)

Now we look for the B-fact:

- B-truthmaker: S5 is true iff metaphysics class occurs at the time at which the speaker of S uttered the token instance of S
- B-truthmaker: S5 is true iff metaphysics class occurs on date X (X=date the speaker uttered S).

Of course, the B-theorist might make further qualifications, e.g., simultaneous to a reference frame or at the same time relative to their clock in the reference frame.

18.1.4 *Returning to Argument - Denial of the Present*

Let's return to our original argument.

- P1: If eternalism is true, then there is no objective present.
- P2: If there is no objective present, then sentences that refer to the present are meaningless and fail to be true.
- P3: Sentences that refer to the present are meaningful and are true.
- C: Therefore, eternalism is false.

The eternalist now has a straightforward response to this argument. They will assert that **P3 is false**. However, note the following from our discussion of the reduction of A-sentences to B-truthmakers:

- A-sentences are not meaningless.
- A-sentences can be true.
- If we think to ourselves ``what I am experiencing right now is the present'', this does not imply that there is the A-fact of some event being present. It only implies the B-relation: ``my experience of this event occurs at the same time as my having this thought (or the same time as some other event happening now)."

We can, perhaps, but this more simply by saying that when we talk about ``the present'', we are not talking about some objective feature of the world but a *relation* between two events, e.g., my utterance of a sentence and a clock reading in 2021.

Exercise 18.113

In this section, we considered the argument that (1) there are some A-sentences that are meaningful and true but (2) eternalism denies this. Try to articulate the eternalist's response using the reduction approach.

18.1.4.1 Argument 2 - from Experience of the Present

Let's look at an argument that purportedly supports presentism. One of the strongest reasons to accept presentism is that it seems that the only time that we experience is the present. In other words, our conscious experience is **confined** to the present. Here is a simple argument for the truth of presentism:

- P1: The only conscious experience we have is of present events.
- P2: Presentism says that the only events that exist are those in the present.
- C: Therefore, the best theory is presentism.

P1 is sometimes called the **confinement doctrine**.

Definition 18.4: confinement theory 1

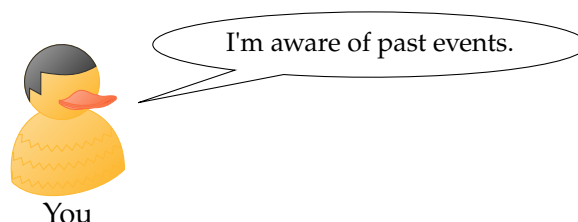
Our conscious experience is only of present events.



You

I'm only aware of present events.

The eternalist can respond to this by pointing out that P1 is false. Suppose, for example, that you are staring at the sun. Given that the light from the sun takes 8 minutes and 20 seconds to reach your eyes, your conscious experience is not of a present event (the sun shining). It is instead, of an event that happened at a prior time. More generally, since it always takes time for light to reach our eyes, or for our brain to register some sensory phenomenon (e.g., the roughness of a desk), all of our experiences are of events in the past. In short, our conscious experience of the sun is not of the sun itself but of the sun in the past, which we experience through its effects.



With this in mind, let's revise **P1** of the above argument by saying that it is not that we experience events that occur in the present but *our experience of an event is in the present*. Alternatively, we might say that we don't see the star in the present. What we instead see are the effects of the star *in the present*.

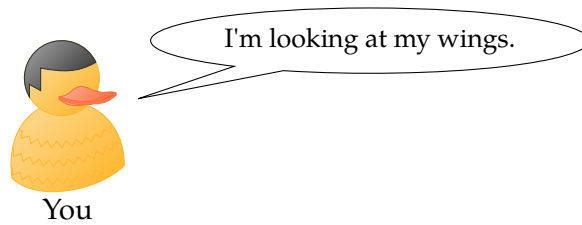
Definition 18.5: confinement theory 2

Our conscious experiences of events are confined to the present moment.

With this new version of the confinement theory, we can revise the argument against eternalism as follows:

- P1*: While we may not experience present events in the present, all of our experiences of events are in the present moment.
- IC: Therefore, only present events are real.
- C: Therefore, eternalism is false.

How might the eternalist respond? First, the eternalist might argue that P1 is not obviously true. Our experience does not tell us that there are no other events earlier or later than the event we are experiencing. For example, suppose you are looking at your hands.



The experience you have of *seeing your hands* does not tell you that there are no other spatial objects in the universe. Instead, it only tells you that you have hands. Relatedly, your experience that you have hands does not tell you that there are no other events in time (ones occurring later or earlier). For all you know, there could be events earlier or later than the one you are experiencing.

Second, your experience that you have hands does not tell you that your experience is *in the present*. Rather, the only thing you can conclude about your experience is that it is happening *at a time* or *it is happening when it is happening*. So, while you know you are having an experience, no part of your experience tells you *when* that experience is happening.

In other words, the correct statement of the confinement doctrine (P1) is the following

Definition 18.6: confinement theory 3

Our experience happens when it happens. Our experiences are confined to the time in which they occur.

On this latest version of confinement theory, if an experience happens at t_1 then it happens at t_1 . I don't know if it is the absolute present. The true confinement theory is that experiences are confined to when they happen. With this new version of the confinement theory, let's consider the updated argument against eternalism:

- P1**: Experiences happen when they happen.
- IC: Therefore, only present events are real.
- C: Therefore, eternalism is false.

This updated version of the argument against eternalism is invalid. It is possible that P1 is true and IC is false. Since earlier experiences happen when they happen, then they happen at an earlier (non-present) time. Since later experiences happen when they happen, then they happen at a later (non-present) time.¹

18.1.4.2 *Argument - From Disgust*

Let's introduce a few definitions.

Definition 18.7: Determinism

Determinism is a theory about the world that every future event is determined by the current state of affairs and the laws of nature.

The idea is that if we had a complete specification of the world at a time t and a complete knowledge of the laws of nature, then we could derive future events. If determinism is true, then future events (or events later than one designated at t) are determined by the details of the world at a time t and the laws of nature.

A closely related theory to determinism is fatalism.

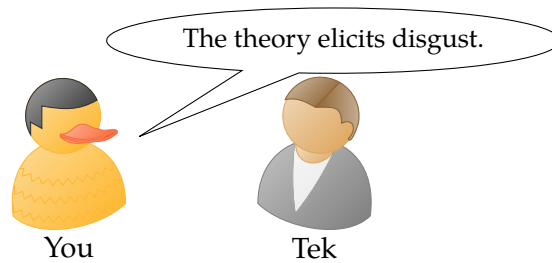
Definition 18.8: Fatalism

Fatalism is the theory that future events occur regardless of any choice you might make.

In the case of determinism, future events are determined by two things (a state of affairs at a time and the laws of nature / physics). In a fatalistic world, future events are fixed by any number of things. They may be fixed due to a variety of factors:

1. because the world is deterministic
2. by God who has a plan and will institute that plan regardless of the laws of nature
3. the logical structure of the universe: every sentence is true or false (one or the other, not both and not neither), there are future-tensed sentences, these sentences are made true or false by future events, therefore, the future is already determined.

Theories that deny free will are often rejected. These theories are rejected for a variety of reasons, but let's say that they are rejected because they elicit *disgust*. The general idea is that since fatalism evokes a deep-seated disgust (or ``yuck" factor), this is grounds for believing that fatalism is false.



To see this more clearly, let's first consider a general argument from disgust and then apply it to fatalism.

- P1: Theory X implies fatalism.
- P2: Fatalism elicits disgust.
- P3: Any theory that evokes disgust is false (or ought to be rejected).
- C: Therefore, theory X is false.

Now let's apply this argument to eternalism. Notice that eternalism entails all events, regardless of their place in time, are equally real. From our current position in time, there is nothing we can do about future events. If there is nothing we can do about future events, then we lack freedom. Moral responsibility implies the possibility of freedom. Therefore, there is no such thing as moral responsibility in this world. This is a horrific and disgusting outcome. Therefore, eternalism is flawed. In sum,

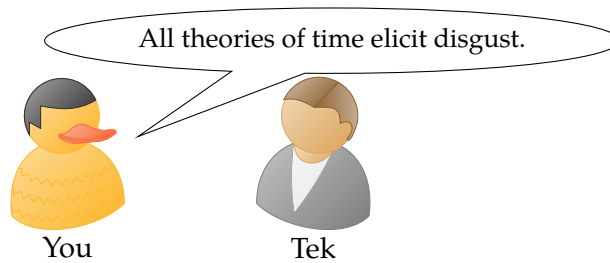
- P1: Eternalism implies fatalism.
- P2: Fatalism elicits disgust.
- P3: Any theory that evokes disgust is false (or ought to be rejected).
- C: Therefore, eternalism is false.

How might the eternalist respond to the argument from disgust?

First, let's put aside the overall quality of arguments from disgust. Let's assume that they are an acceptable way to argue for a position. One response that the eternalist might make is that **even if** arguments from disgust are a legitimate way to argue for a theory of time, nearly every theory of time elicits disgust.

- The moving spotlight theory has an eternalist temporal ontology and so is equally fatalistic.
- One way that presentism explains the truthmakers of past-tense sentences by invoking present evidence and the laws of nature and if the laws of nature are symmetrical (apply backwards and forwards), then then the fatalism objection also applies for presentism.

In sum, the argument thus provides no relative advantage for other theories over eternalism.



A second response is to reject arguments from disgust in general. That is, to reject **P3**. On this approach, there is no reason to accept or reject a theory based upon a feeling of disgust. Consider that feelings of disgust have been used to support such things as:

1. Anti-Semitic propaganda
2. That African Americans are inferior
3. Mistreatment of homosexuals

Individuals might cite disgust concerning these individuals as a way of justifying inhumane treatment of these groups of people. But, at least to our modern perspective, no one would accept these arguments as legitimate since, in these cases, a person's feelings of disgust are rooted in bias or prejudice rather than an understanding of objective reality.

Exercise 18.114

Do you think that arguments from disgust have validity in any context? That is, if you are skeptical of their use to prove some fact about the nature of time, are there other contexts in which they can apply (e.g., moral issues, questions concerning beauty)?

18.1.4.3 Argument: from Experience of Temporal Passage

Another argument against eternalism is drawn from our experience of change or our experience of the passage of time. In short, what explains that we experience time passing? Dynamic theorists have a relatively simple answer to this question. Our **experience** of time passing (experiential temporal passage) is explained by **time passing** (ontological temporal passage). This insight forms the basis of the following argument:

- P1: We have an experience of the passage of time (change).
- P2: What would best explain our experience of the passage of time is that time actually passes (change is an objective feature of reality).
- P3: According to eternalism, the passage of time is only a perspectival (relative) fact.

- C: Therefore, eternalism is false.

The eternalist has at least two possible responses:

1. Deny we experience change (we don't really experience the passage of time)
2. Accept that we experience change but this experience is not best explained by real temporal change in the world. What causes our experience of change is not change in the world (not ontological change).

Let's consider both.

18.1.5 Response 1: We don't actually experience the passage of time

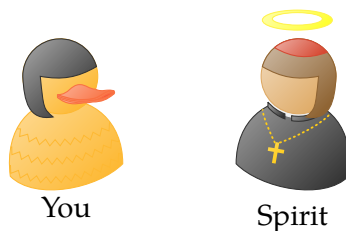
One response to the above argument is to say **P1 is false**. On this response, our experiences do represent reality, but we do not actually experience the passage of time (or change). On this approach, there is a disconnect between:

- What we actually experience
- What we think (or believe) about what we experience

To illustrate, let's consider some examples.

First, suppose you think you saw a ghost. You didn't really see a ghost, but you think you saw one. In this case, there is nothing *wrong* here with your experience. What you saw is what is there in the world. What is wrong is your *interpretation* of your experience. You incorrectly interpreted what you saw.

Second, let's start with a more radical example. Suppose you are just awaking from a deep slumber. You believe you are having a religious experience in that you believe you see a spiritual being.



After seeing the being, you fall back to sleep. Now while you *think* you saw a spiritual being, let's suppose that what you really saw was a person on your computer screen (you left it on last night and it was autoplaying videos).



So, what happened was the following:

- You perceived a person on your computer screen
- What you *think* you saw was a real-life spiritual being.

What is being pointed out here is that there is sometimes a difference between (1) appearances gained via sensation or perception, (2) our beliefs about what we experience, and (3) reality (or what actually occurred).

Third, it is common for an illusionist ("magician") to make us think we see one thing but when we actually saw no such thing. An illusionist might take three separate unbroken rings and connect them together. What we believe we are seeing is the illusionist separate three unbroken rings, but we did not actually experience this fact. We do not really see that the rings are unbroken.

Exercise 18.115

Find a video of the 'snap change card illusion'. For someone who is fooled by this trick, describe (1) what they see and (2) what they think they see.

Third, suppose I believe I am having an experience of seeing a space in front of me that is continuous. I say to myself "I see a continuous (unbroken) sheet of paper. There may not be any paper before me, but my experience is an experience of a continuous sheet of what I call 'paper'." Note that this is inconsistent with other things that we believe about what we actually experience. We know that the back of the eye is composed of millions of retinal cells and we know that this is not a continuous sheet but many cells next to each other. We also know that there is a blind spot in the back of the eye where there are no retinal cells. And so, it makes more sense to say that what we actually experience are millions of tiny dots (bits of information) but our brain fills in the gaps between these empty spaces. That is, the structure of our retina and the existence of the blind spot suggests that what we are experiencing is not an experience of continuous (unbroken) space. Again, there is a disconnect between what we believe we are experiencing and what we are actually experiencing.

Let's return to the passage of time. The eternalist can say this about our so-called experience of time.

1. Belief about experience: People believe they are experiencing the passage of time passing (change)
2. Actual experience: for any given time t , we experience what happens at that time t and at that time t , this experience contains content from (memory) an event designated as earlier. We don't actually experience change or passage or flow.

Let's consider one possible illustration. Consider the following two experiences:

- Experience 1: You are looking at a clock and it reads 10:15 AM.
- Experience 2: Looking at a clock and seeing it says 10:18AM and the memory of Experience 1.

In this example, there is no actual experience of time passing. Instead, for each moment in time, there are moments where you (a) have an experience and (b) have an experience and a memory. When you have an experience and a memory of an earlier event, you are simply having a more complex experience at a time. You are not actually experiencing the movement or flow of time. Instead, there is simply an ordered set of increasingly complex experiences that happen when they happen. All you have are discrete experiences coupled with memories.

Another example. Suppose I drag a piece of chalk over a chalkboard. You may think that you are actually experiencing change or temporal flow, but the eternalist can argue that this is not what you are actually experiencing. What you experience is a set of individual events with the memory of events that occurred before each later event.

18.1.6 Response 2: We experience the passage of time, but this is not due to the world

Recall the earlier argument raised against the eternalist:

- P1: We have an experience of the passage of time (change).
- P2: What would best explain our experience of the passage of time is that time actually passes (change is an objective feature of reality).
- P3: According to eternalism, the passage of time is only a perspectival (relative) fact.
- C: Therefore, eternalism is false.

The eternalist has at least two possible responses:

1. **Deny P1:** deny we experience change (we don't really experience the passage of time)
2. Accept that we experience change but this experience is not best explained by real change in the world.

Let's consider the second response. This response involves *denying P2*. On this response, the eternalist contends that we do really have *an experience of the passing of time*, but this experience is not due to (caused by, explained by) the existence of temporal passage (change) in the world. That is, our experience of the passage of time does not directly correspond to what goes on in the world. To better understand this approach, it is helpful to define a *phenomenal illusion*.

Definition 18.9: phenomenal illusion

A **phenomenal illusion** is a genuine experience of something but that something does not correspond to reality

My favorite example of a phenomenal illusion is the simultaneous contrast illusion. Take a look at the image below and answer how many gradients are there?



The simultaneous contrast illusion is an illusion that occurs when adjacent colors influence the perception of these colors.

“Our perception of shades (and colours) changes according to their visual context, a process known as simultaneous brightness contrast. If a shape is surrounded by something dark, we perceive it as brighter than if they're amidst lighter shades.”²

Other examples may include a gray square looking different on different backgrounds (black, white background with a black border, white background with no border). In the above case, we experience (see) two gradients when there is only one. In this case, we do not need to consult **what you believe about what you experience** but instead can consult what you **actually experience** when you look at the illusion. In this case:

1. You actually experience two gradients
2. But, your experience of two gradients is not best explained by there being two gradients (since there is, in reality) only one gradient
3. It is instead better explained by the workings of your visual system (aka how you experience the world, not the world itself)

The eternalist then can point to illusions like this and contend that **we do experience the passage of time** but the experience of time is better explained by **how you experience the world and not the world itself**.

One problem with this response is that it might apply to our perception of contrasting colors, but how do we know it also applies to time? In short, the response seems speculative. It is one thing to point out that our experience of the passage of time could be due to our subjective interaction with the world and another to show that it is, in fact, due to this fact. In addition, examples involving visual perception (e.g., the simultaneous contrast illusion) seem somewhat remote from those involving time. In other words, we need some positive reason to think that change (temporal passage) is an illusion.

Here is a closer example. There are some phenomenal illusions where we experience (change / motion) but where there is no motion. One such example are motion aftereffect illusions (waterfall illusions). A **motion aftereffect** occurs when you *see motion in a static image*. This often occurs when viewing a moving visual stimulus for an extended period of time. One famous example is the waterfall illusion. Find a waterfall. If you stare at a waterfall for a while, then look at the rocks next to the waterfall, it will appear that the rocks are moving upward.

Another example involves rotating images. After watching a set of rotating disks, the disks appear to move even when they are not moving. In this case, you are experiencing motion, but this motion is due to how your perceptual system interacts with the world and not due to the world itself.

The eternalist will thus respond that just as we can have an **experience of motion in a static image**, we can have an **experience of temporal passage in a static universe**. And, this fact is reflected in various other commonsense beliefs we have about time:

- Fun: Time passes when you are having fun (does it really speed up?)
- Fear slows down perception of time: anachrophobic testers were presented with spider-stimuli that frightened them (Droit-Volet).
- Time slows down when you are bored (does it really slow down?)

18.1.6.1 *Argument 5 - McTaggart's Argument from Unchanging Facts*

Some critics of eternalism contend that eternalism involves a universe where there is no change.³ But, clearly there is change. So, eternalism must be false. More precisely:

- P1: B-facts are timeless: if O is P at t1, then O is always P at t1 (e.g., if the poker is hot on Monday, it is always hot on Monday)
- P2: Things do not change unless the facts change (e.g., an O gaining or losing a property: a poker going from hot to cold).
- IC: Therefore, since eternalism asserts that the world consists of there is no change to B-facts (or the properties of objects), there is no change.
- P3: There is change.
- C: Therefore, eternalism is false.

First, begin with P1, the premise that B-facts are timeless. To illustrate,

- It is always the case that the poker is hot on Monday
- It is always the case that the poker is cold on Tuesday.

There is no change to these facts or the truth values of these sentences.

Second, P2 asserts that there is no change to objects without a change to the facts. For example, suppose you put a poker into a fire on Monday and remove it on Tuesday. Intuitively, the poker changes. On Monday, when you put the poker into the fire, the poker was hot. But, then Tuesday came and the poker is no longer hot. In this example, we have one object O that gains and loses the property of being hot. We have one object O, for which it is a fact that it is hot, and then it is a fact that it is not hot.

Third, it appears then that eternalism is committed to saying that since the eternalist's world is a world of B-facts, and these facts never change, there is no change in the eternalist's world. For the eternalist, the poker is a four-dimensional object and so it has a Monday-part that is hot and a Tuesday-part that is not hot. Since these are timeless facts, it is never the case that the poker itself goes from hot to cold.

P3 asserts the common sense datum that there is change, and so the conclusion is that since eternalism (as a B-theory) is incompatible with this datum, it follows that eternalism is false.

How can the eternalist respond? The eternalist responds in a straightforward way:

The poker goes from being hot on Monday to cold on Tuesday. So, it does change. It goes from being hot at a time t_1 to cold at a time t_2 . "Change *just* is the variation in the (intrinsic) properties of a thing between one time and another"[117, p. 214]

In other words, the eternalist contends that there is change, but change occurs "in virtue of unchanging facts about [the] temporal parts" of an object ([117, p. 214]). In other words, change is not a single object O gaining and losing properties but an object's temporal parts instantiating different intrinsic properties. When we talk about the change in the poker, we are talking about the poker's temporal parts instantiating different properties (one part is hot, another is cold). The presence and absence of the property of being hot is what change involves.

18.1.6.2 *Argument 6 - McTaggart's Argument from Spatial Analogy*

For the eternalist, an object's temporal parts are able to instantiate different (incompatible) properties (e.g., the person may be young at t_1 and old at t_2). Analogously, an object's spatial parts are able to instantiate different (incompatible) properties: a road may be bumpy, smooth, or patchy because different spatial parts instantiate these properties. Given this analogy, this seems to imply that temporal change is analogous to spatial variation. But, this yields an absurd result since we don't talk about spatial objects as "changing" just because their parts have different properties. Therefore, there must be something wrong with the eternalist's understanding of change.⁴

To understand this argument, let's make two points:

- The Greenwich (prime) meridian is a reference line that passes north to south through a series of degrees of latitude.
- Let S_1 and S_2 be two different points of latitude and let it be true that S_1 passes through the UK and let it be false that S_2 passes through the UK.

With these points in mind, consider the following argument:

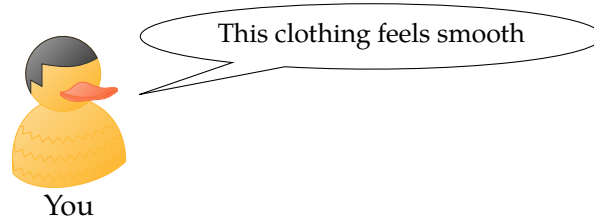
- P1: Eternalism takes space and time to be analogous.
- P2: Eternalism says that the poker **changes** (or there is change) because one of its parts is hot and another part is cold.

- P3: No one would say that the meridian (a spatial object) changes because one of its spatial parts is in the UK and another part is not in the UK. An object doesn't change because its parts have different properties.
- IC: Therefore, eternalism has no coherent sense of change.
- C: Therefore, eternalism is false.

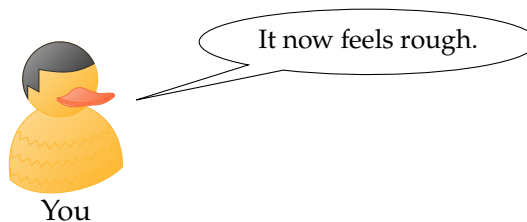
The eternalist responds as follows:

“The four-dimensionalist ought to grant that spatial variation is in many ways analogous to temporal change; it is unclear why this should be thought problematic.”[117, p. 216]

The response is that it not at all strange to talk about spatial objects as “changing” due to the variation in the parts of that spatial object (so P3 is false). Here is an example. Suppose you are touching a piece of clothing and the clothing feels smooth.



As you are running your hand over the clothing, the texture changes from smooth to rough. You might say that the texture of the clothing changed from smooth to rough.



In short, it is not strange to talk about spatial objects as “changing” due to the variation in the parts of that spatial object. Some other examples include:

1. The road changed from smooth to bumpy[117, p. 216]
2. The color gradient on the page changes from red to green.
3. The limb of the tree changed from strong to weak (assume you are walking out to the edge of the limb)
4. This paper on time went (changed) from exciting and coherent to esoteric and incomprehensible.

In sum, the eternalist sees nothing counterintuitive with understanding temporal change as analogous to spatial variation.

18.1.7 *Arguments for Eternalism*

We examined several arguments against eternalism as well as several (somewhat) lengthy replies from the eternalist. Let's suppose that the eternalist's responses are convincing. If so, then we have no direct reason to reject eternalism. However, we also have no positive reason to accept eternalism either. In what follows, several reasons will be given to support eternalism.

18.1.7.1 *Argument 1 - From Science*

One of the strongest reasons is from science (from the theory of special relativity). In short, A-theories seem to require the existence of a single present that applies to all events and objects in space. This seemingly entails that there is an absolute notion of simultaneity. That is, if event E_1 and E_2 are both in the present, then E_1 is simultaneous with E_2 . However, the special theory of relativity contends that there is no absolute notion of simultaneity. Therefore, this implies that there is no objective notion of the present.

If A-theories run into conflict with the best science, then we ought to search for an alternative. Since the best alternative is eternalism (a B-theory), we have an argument in favor of the B-theory.

There is an immense amount of literature on this topic, most of which focuses on (1) presenting the special theory of relativity and (2) how to understand it in relation to the different theories of time. As such, there is considerable discussion about the technical details of the special theory of relativity (this is outside the scope of this course). For some overviews, see chapters 4 and 5 from [117].

18.1.7.2 *Argument 2 - From Leibniz's Laws*

We tend to think that an object can retain its identity over time. For example, suppose I am looking at a billiard ball B_1 on a pool table at t_1 and then again at the ball B_2 one second later t_2 . It is a strongly held intuition that I am looking at one and the same ball both times. In other words, the balls are identical to each other: $B_1 = B_2$. One argument in support of eternalism is that it better fits with our beliefs about identity.

We will examine this argument in three parts. First, we will clarify what it means for two objects to be identical by introducing Leibniz's law. Second,

two different accounts of how objects persist through time are introduced. These are endurantism and perdurantism. Third, it will be argued that perdurantism better aligns with Leibniz's laws and that this combination better aligns with eternalism.

18.1.8 Leibniz's Law

What do we mean when we say that two objects are identical? What we don't mean is that they are closely similar. Twins are sometimes said to be "identical" but this does not mean they are one and the same object. Two pens in a box of pens may be said to be "identical" but this just means "qualitatively identical" or "identical for all practical purposes". Instead, presumably, what we mean when we say that two objects x and y are identical is that they have the same properties. And, we also mean to say that if two objects have the same properties, then they are identical. This idea is captured by what is known as **Leibniz's Law**:

Definition 18.10: Leibniz's law

For every object x and y , $x = y$ iff x and y have all of the same properties.

This law is the conjunction of two laws:

1. Indiscernibility of identicals: If two things x and y are identical to each other (the same thing) then they have all of the same properties. If $x = y$, then for any property P , if x is P , then y is P .
2. Identity of indiscernibles: If two things have all of the same properties, then they are identical to each other (the same thing). For every property P , if x is P and y is P , then $x = y$.

Let's illustrate Leibniz's law by using it to show that two things A and B are *not identical*. Suppose there are two objects A and B (and you don't know if they are identical). Now suppose that A has a property that B lacks. According to Leibniz's Law, it follows that A is not identical to B.

Here is a concrete example:

- Suppose there is a murderer is on the loose and that the murderer is lefthanded.
- Now suppose you suspect that Tek is the murderer but you discover Tek is righthanded
- By Leibniz's law, you know that Tek is not the murderer
- The murderer and Tek have different properties so they are not identical.

In this section, we introduced Leibniz's law and used the law to conclude that (1) when two objects *A* and *B* have different properties, it follows that (2) they are not identical.

18.1.9 Persistence

Objects are said to ``persist" through time.

Definition 18.11: persistence

An object *O* persists through time if and only if *O* exists at various times.[cf. 70, p. 202].⁵

The pen in my hand persists since it exists at various times. It exists in my hand now, a few minutes earlier, and it will continue to exist in a few minutes from now. My daughter persists through time. She exists now, at an earlier time, and will (hopefully) exist at many times in the future. In contrast, a fictitious unicorn does not persist since there are no times in which it exists. Similarly, a cube that exists at a single instant in time does not persist since there are not various times in which it exists.

One more example. Suppose today is the day that Tek is standing trial for murder. At his trial, he is said to be the *one and the same* Tek who, at an earlier point in time, murdered someone. In saying that Tek is *one and the same* Tek who murdered someone at an earlier date, what is not meant is that there exists an object that is relevantly similar to Tek who murdered someone. For Tek is standing trial for his crime, not the crime of someone who looks like him, or his identical twin, or a clone made in a laboratory. Rather, what is being said is that there is one object Tek that persists through time and the object that is standing trial is the same object that is said have (at an earlier time) performed a murder.

There are two main accounts of persistence through time. That is, there are two different accounts of how one and the same object can exist at different times t_1 and t_2 . Let's consider both of these in the following subsections.

18.1.9.1 Endurantism

The first theory of persistence is called "endurantism" (also known as "three-dimensionalism"). On this account, *O* persists through time by enduring through time and an object *O* endures through time by being wholly found at each time it exists.

Definition 18.12: Endurantism

Endurantism (three-dimensionalism): is the view that objects persist by enduring. An object O **endures** if it can be wholly present at each time it exists (an object O is wholly present in a region R of spacetime if and only if O “fits exactly into” R , cf. [141, p. 246] Objects are three-dimensional (3D) entities.

Here is another definition of endurantism: “Necessarily, for any x and any times, the ts : if the ts are more than one and x exists at each of the ts , then x is wholly present at each of the ts .”[30]

One key concept in the definition of “endurantism” is that of an object being “wholly present” or “wholly found” in the times where it exists. This concept has been notoriously difficult to make explicit.[30]. The approach here won't be to make this idea precise but instead to give its intuitive sense.

On the endurantist view, objects are thought to “sweep through” time. Imagine a timeline with the times 2023, 2024, and 2025. On the endurantist view, an object O sweeps or moves through these times (first 2023, then 2024, and then 2025) such that whatever time an object is at, that is the only time you can find the object.

An object's persistence can be thought of like a car traveling to a destination. The car begins its journey in Los Angeles, then passes through Chicago, and ultimately arrives in New York. At any time in the journey, the whole of the car can only be found in one spatial location. Similarly, if Tek is halfway through a marathon, the whole of Tek is located in one and only one point: at the halfway point. The same is true for a person's life. Take the entire history of your own life. For the endurantist, the you that is here and now is wholly here at this moment. As Mellor put it “things are wholly present throughout their lifetimes.”[81]

My favorite metaphor comes from Ted Sider [117] who asks what it would mean to touch an enduring object. If endurantism is true, when you meet an enduring object O at time t , you are meeting the whole object. To touch an enduring object O , you only need to touch all of its spatial parts at a time t since the whole of the object is found at t . To spend a lifetime with an enduring object, you need to spend each successive instant with them.

Endurantism is called the “three-dimensional view of persistence” because objects are extended in three spatial dimensions but have no extension in time. The 3D object moves along the temporal axis but is not extended along it.

Endurantism enjoys a lot of intuitive support. It corresponds with several beliefs we have about our own existence through time and beliefs we have about rewards and punishments.

- The whole of me moves through time. I survive change.
- I am one and the same person at successive instants of time.
- Endurantism Corresponds with beliefs about ourselves (I'm here all at once. I'm nowhere else).
- Rewards and punishments (we punish the one and the same person who committed the crime days ago).

Exercise 18.116

Draw an image or diagram that illustrates the endurantist position.

18.1.9.2 *Perdurantism*

Endurantism is one theory of how objects persist through time. The other major theory is perdurantism (also known as four-dimensionalism). On the perdurantist theory, objects persist by perduring.

Definition 18.13: perdurantism

Perdurantism (four-dimensionalism): objects perdure rather than endure through time. An object *O* **perdures** if at every time $t_1, t_2, \dots, t_n$, a proper part of *O* (called a “temporal part”) is found at each time *t*. Objects are four-dimensional (4D) entities.

Perdurantism is called the “four-dimensional view of persistence” because objects are extended in four dimensions (three spatial and one temporal). It is also sometimes called the worm theory since the whole of you is a 4D spacetime worm: you are the totality of your spatial and temporal parts.

For the perdurantist, how do objects persist through time? Recall that for the endurantist, *the whole of you moves through time*. That is, you *sweep through time*. For perdurantists, *parts of you are located at different moments in time*. That is, whole of you is *spread out over time*. For example, imagine that there are three points in your life: your birth ($t - 1$), you right now (t), and you right before you die ($t + 1$), then your existence would consist of $t - 1, t, t + 1$ and all times in between these.

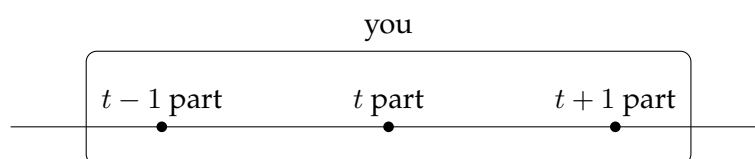
Perdurantism is often associated with the **doctrine of temporal parts**. This is the view that objects not only have spatial parts, but also have temporal parts.

To understand temporal parts, it is helpful to think of objects as protracted **events**. An American football game is an event that consists of four parts (quarters). These are its temporal parts: 1st quarter, 2nd quarter, 3rd quarter, and 4th quarter. The game can also be divided up in terms of the minutes elapsed. There is the 1st minute, the 2nd minute, the 3rd minute and so on. These, again, are its temporal parts. Now the whole of the object is the sum of its parts. The game itself is not any one part but the collection.

Here is another way of thinking about perdurantism. As a material being, you have spatial parts (arms, legs, torso, head). As an material object, you likely do not think that you are any one of these parts, e.g., you are not merely your left hand. Instead, as a material being, you take yourself to be the collection of these spatial parts.

The perdurantist contends that not only do you have spatial parts, but you also temporal parts.

1. You have the temporal part of you that consists of the 2022-you
2. You have the temporal part of you that consists of the 2021-you
3. You have the temporal part of you that consists of the 1 Jan 2021 you



Consider that if we were to meet an enduring object at a time, we would meet the entire object. And, to touch the whole enduring object, we would only need to touch all of its spatial parts at a time. In contrast, when you meet a perduring object O at a t , you are not meeting the whole object. Instead, you are only meeting a proper part of the object. If you touch a perduring object, you are not touching all of them. To really touch the whole person, you would need to hold them in an all-consuming embrace for their entire lives.

Let's clear up a few misconceptions about perdurantism.

First, perdurantism says that not only do you have spatial parts but you also have temporal parts. However, note that a temporal part is not simply a spatial part of an object at a time. For example, your hand at a t_1 is not a temporal part of you. Rather a temporal part is as big as you are at that time. So, suppose you are just your body. A temporal part would be your entire body at a time t_1 . This idea that a temporal part is as big as you are at a time is reflected in the following definition of "temporal part" (cf. [117]):

Definition 18.14: temporal part

x is a temporal part of y at t if (i) x is a part of y at t ; (ii) x exists at, and only at, t , (iii) x overlaps at t everything that is part of y at t .

Second, another common misconception about perdurance (the doctrine of temporal parts) is that objects maintain their existence by being constantly created. Thomson writes:

“this [4D] seems to me a crazy metaphysic. It seems to me that its full craziness comes out only when we take the spatial analogy seriously. The metaphysic yields that if I have had exactly one bit of chalk in my hand for the last hour, then there is something in my hand which is white, roughly cylindrical in shape, and dusty, something which also has a weight, something which is chalk, which was not in my hand three minutes ago, and indeed, such that no part of it was in my hand three minutes ago. As I hold the bit of chalk in my hand, new stuff, new chalk keeps constantly coming into existence *ex nihilo*. That strikes me as obviously false.” [133, pp. 201–220], [117]

Thomson argues that if we accept eternalism, then we accept an ontology of objects where the continued existence of an object (e.g., a piece of chalk) is being constantly created out of nothing (*ex nihilo*). That is, when I am holding a piece of chalk in my hands, its continued presence in my hand is not due to that object persisting or maintaining its presence *through* time but it being constantly recreated at every instant.

Sider's [117, p. 217] response is as follows:

“in saying that temporal parts come into existence *ex nihilo*, Thomson makes it sound as if a *miracle* is constantly occurring. That isn't right. The sensible four-dimensionalist will claim that current temporal parts are caused to exist by previous temporal parts. The laws that govern this process are none other than the familiar laws of motion.”

For Sider, there is nothing causally strange about perduring objects since later temporally parts are causally determined in the normal way by earlier temporal parts.⁶

Exercise 18.117

1. Explain the perdurantist view. Include a diagram or picture with your explanation.
2. The perdurantist view (doctrine of temporal parts) has been articulated. Suppose that you believe in perdurantism, what practical implications (if any) for how you understand yourself or others.

18.1.10 The Argument from Leibniz's Law

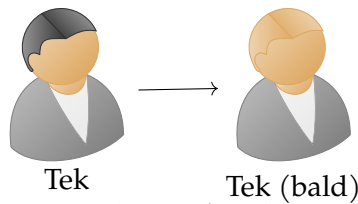
We have described Leibniz's law, the notion of persistence, and two accounts of persistence (endurantism and perdurantism). Let's begin to formulate the argument from Leibniz's laws. First, we will start with the simplified version of the argument. Suppose you have 300 beans in a jar. Now if you subtract 100 beans, you surely do not have the same number. Similarly, if you have a cake, eat a piece of it, the size of the cake is surely not the same. Finally, now take a person. That person grows or decays. Thus, just like the beans and the cake, they are not one and the same person.[\[115, p. 255\]](#). In short, "[c]hange seems to require difference: if something has changed, it is different from how it was. But if it is different, it cannot be identical, on pain of contradiction."[\[27\]](#)

- P1: 3D-endurantist theory of persistence conflicts with Leibniz's laws ($x=y$ iff x and y have all the same properties).
- P2: 4D-perdurantist theory of persistence does not conflict with Leibniz's laws.
- P3: Eternalism is the only theory of time that accepts the 4D-perdurantist theory of persistence.
- C: Therefore, eternalism is true.

Let's focus on **P1** by considering Tek who has shaved his head.

- t1: Tek with hair (Tek)
- t2: Tek without hair (Bald-Tek)

If 3D is true and Leibniz's laws are true, then **Tek is not identical to Bald-Tek** (not the same). They don't have the same properties and therefore they are not the same object. In other words, if endurantism is true, then objects never survive any change to their parts. Tek did not survive the loss of his hair!



Next, consider **P2**. If we accept the 4D (perdurantism) and Leibniz's laws, then Tek and Bald-Tek are **not identical**. But this is not a problem for the perdurantist since they say that Tek and Bald-Tek are non-identical. These two "Teks" are *parts* of Tek. Just like your hand and your foot are not identical but they are still part of you, so too Tek and Bald-Tek are not identical. They are different temporal parts of a larger Tek (the 4d spacetime Tek).

Finally, note (P3) that most eternalists accept the 4D (persistence) view of persistence, while non-eternalists tend to accept the endurantist (3D) account. The eternalist contends that all events in time exist and so the so-called "future", "past", and "present" you exist but do so as temporal parts.

18.1.10.1 Argument - from paradoxes of material constitution

According to some, one of the strongest arguments in favor of eternalism is its ability to solve the paradoxes of material constitution. In order to understand how Eternalism accomplishes this feat, we need to first (1) state some common beliefs people have about material objects and (2) how these common beliefs lead to paradox.

18.1.10.2 The Standard View on Material Objects

First, let's consider some beliefs that people have about material objects.

1. One Thing in One Place: It is impossible that two *distinct* material objects are ever *wholly found* in the same spatial location at the same time.
2. Constitution is identity: If there are two objects *A* and *B* and objects *A* and *B* are wholly constituted by exactly the same material, then they are the same object, viz., $A = B$.
3. Diachronicity of material objects: Material objects can survive change.
4. Leibniz's Law: If there is something true of object *A* that is not true of *B*, then $A \neq B$.
5. Composite Objects: There are composite material objects. There are tables, ships, cats, and human beings.
6. Transitivity of Identity: If *A* is the same object as *B* and *B* is the same object as *C*, then *A* is the same object as *C*.

Let's call the above propositions, "The Standard View". (1) contends that

if an object is fully found in a location, then there is only one object fully found in that location. In other words, for any given region of space, there is at most one object that wholly occupies that space. Take a book lying on your desk. Other objects may partially be found in the space occupied by the book, e.g., the cover, the pages, but only one object (the book) fully occupies that space. There is only one book (not two or ten or thirty) books in that space.

For (2), material objects are made up of ("constituted by") stuff. A marble statue is made up of marble. Humans are made up of blood, bones, organs, and so on. A book is constituted by its cover, the pages, and glue. Alternatively, a book is made up of atoms. If there are two objects A and B and they are constituted by the exact same stuff (the same matter), then these two objects are the same object ($A = B$). Metaphysicians sometimes say that for material objects, the relation of constitution is the relation of identity. Or, more simply "constitution is identity".

For (3), material objects can survive change. Suppose I have a notebook on my desk. I pick up this book, write a note in it, and then put it down. The book is different than it was before because now it has a note in it, but it is not a *new* object. It is one and the same book but now with a small change. Here is another example. I'm hungry. So, I eat a sandwich. Now I'm not hungry. In going from hungry to not-hungry, there is not a new person before you. I am the same person as I was before. I didn't die after eating the sandwich. I am still "David W. Agler". In sum, (3) says material objects can survive change.

To illustrate (4), suppose there are two objects A and B , if A is 6'0 tall but B is 5'0 tall, then surely they are different objects. In fact, this seems universally true. In addition, the equivalent contrapositive is true: if $A = B$, then everything true of A is true of B .

(5) says that there are objects that have parts. A table has legs, a top, bolts, etc. A human being has a heart, kidneys, skin.

(6) says that identity is transitive. If Clark Kent is Superman, and Superman is "The Man of Steel", then Clark Kent is "The Man of Steel." Similarly, if Tek at age 21 is the same person as Tek at age 20, and Tek at age 20 is the same person as Tek at age 19, then Tek at age 21 is the same person as Tek at age 19.

Now that we have the standard way of looking at material objects in front of us, our next step is to consider several puzzles that seemingly require us to give up at least one of the propositions in the Standard View.

18.1.10.3 *The Puzzles of Material Constitution*

Let's consider the most famous puzzle of material constitution: the ship of Theseus puzzle. Our focus will not be on this puzzle, but it is worth mentioning initially because it is so famous.

18.1.10.4 *The Ship of Theseus*

Imagine a ship called "the ship of Theseus" (or SOT for short). As the planks of the ship decayed, they were gradually replaced with new planks in their place. Over time, all of the planks (and parts) of the ship were replaced such that the ship today contains none of the original planks of the original ship. Let's call this ship "Replace".

1. SOT and Replace are the same object (by Diachronicity of Identity)

Next, we believe that some objects can survive being disassembled and reassembled. Suppose you take a watch in for your repair. The watchmaker takes the watch apart, cleans a piece, and then reassembles it. Surely, it is the same watch. Or suppose your cell phone has water damage. If you buy a new phone, that is surely not the same phone as your phone. But, if a technician repairs your phone by disassembling and reassembling it, then the phone you receive back from the technician is the same phone that was damaged.

With respect to the SOT, suppose that whenever SOT needs to be repaired, an old man collects the old planks (and parts). He stores them in a warehouse. Once all of the parts of SOT have been removed, the old man takes those parts and reassembles them to form a ship in the exact same configuration as the SOT. Let's call this ship "Planks". According to the standard view, if two objects are constituted by the same stuff, then they are the same object. And, since Planks and SOT are constituted by the same stuff, there is reason to believe the following:

2. SOT and Planks are the same object (by Constitution is Identity)

By Leibniz's Laws, Planks and Replace are not the same object: they are made of different material, they are in different locations, and one is sea-worthy and the other would likely sink if put to sea.

3. Replace and Planks are not the same object: $\text{Replace} \neq \text{Planks}$ (by Leibniz's Laws)

However, if identity is transitive, then Replace and Planks are the same object.

4. Since $SOT=Replace$ and $SOT=Planks$, $Replace=Planks$ (by Transitivity of Identity).

But this is a contradiction since we have $Replace \neq Planks$ and $Replace=Planks$.

So, either our reasoning above is mistaken or there is something wrong with the Standard View.

Exercise 18.118

Create a picture / diagram of the Ship of Theseus problem.

18.1.10.5 *Dion and Theon*

Consider two human beings named “Dion” and “Theon”. Dion and Theon share material parts. In fact, Dion is constituted by the same stuff as Theon except for Theon's right foot. So,

- $Dion = Theon - Theon's\ right\ foot.$
- $Theon = Dion + Theon's\ right\ foot.$

By Leibniz's laws, Dion is not Theon because Dion lacks a property that Theon has. Dion does not have a right foot, but Theon does.

1. Dion and Theon are not identical.

But now suppose that Theon's right foot becomes gangrenous and is surgically amputated. What happens to Dion and Theon? According to the Standard View, material objects can survive change, and human beings can survive the amputation of a foot. So, Theon survives. And, since Dion never had a right foot to begin with, surely he survives.

2. Dion survives the amputation since none of Dion's parts are altered.
3. Theon survives the amputation since an object can survive minor changes.

But now notice that Dion and Theon are made up of the same stuff. And, under the standard view, if constitution is identity, then two objects a and b that are wholly made up of the same stuff S are the same object $a = b$.

4. Dion and Theon are identical.

Finally, notice that there is something true of Theon that is false of Dion. Namely, Theon once had a right foot. Dion never had a right foot. Therefore, they are not identical.

5. Dion and Theon are not identical.

(4) and (5) result in a contradiction. Therefore, either something is wrong with our reasoning or there is something wrong with the Standard View.

18.1.10.6 *Tibbles the Cat*

The Tibbles the Cat problem is a sophisma (paradox) developed by Peter Geach, one that he says he developed from William Sherwood.^[15] Here is a presentation of the puzzle:



“Suppose there exists a 10-pound cat named ‘Tibbles’. In looking at Tibbles, let’s name the 9-pound part of Tibbles that includes all of Tibbles except his tail ‘Tib’. Just like your heart and you do not wholly overlap, notice that all the parts of Tibbles and Tib do not overlap. And, just like you are not your heart, Tibbles is not Tib for there are some things that are true of Tibbles that is not true of Tib, e.g., that Tibbles has a tail.”

Now suppose Tibbles loses its tail.

1. Before the loss of the tail of Tibbles and Tib are not identical. Some things are true of Tibbles but false of Tib (and vice versa). For example, Tibbles once was a 10-pound cat but is now a 9-pound tailless cat. In contrast, Tib was always a 9-pound tailless cat.
2. After the loss of the tail, Tibbles and Tib wholly occupy the same space and are composed of the same matter.
3. Tibbles survives the loss of his tail (Diachronicity of material objects)
4. Tib survives the loss of Tibbles’ tail (nothing about Tib has changed)
5. If constitution is identity, then Tibbles and Tib are identical.
6. But Tibbles and Tib are not identical since (by Leibniz’s Laws) something is true of Tibbles that is false of Tib (Tibbles had a tail).

So, just like the case of Dion and Theon, we are led to paradox and we seemingly must reject some part of the Standard View.

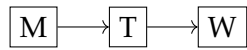
Exercise 18.119

How would you resolve the Tibbles/Tib or Dion/Theon paradox?

18.1.10.7 *The Statue and the Lump*

Take a lump of clay on Monday. Let’s call it “Lump”. The Lump is just the continuous piece of clay made up of clay particles. Now take Lump and craft it into a statue on Tuesday. It is now a Statue. So, let’s call it “Statue”. Now

smash it back into a lump on Wednesday.



Lump Lump Lump

Statue

While we can describe the paradoxical situation involving the Statue and Lump in the same way we described the previous scenarios, let's describe it in a slightly different way.

On Tuesday, how many distinct material objects occupy the space of the statue? Obviously there is *One!* There is just the Statue. If there were two objects, then they would wholly occupy the same exactly space. But this cannot be true given the *One Thing in One Place* principle. Similarly, if there were two objects occupying the statue on Tuesday, then there would be two objects made up of the same stuff. But this cannot be true given that *Constitution is identity*.

On the other hand, there cannot be *One*. There must be *Two!* Intuitively, the statue exists on Tuesday. But there is good reason to think that the lump also exists: The Lump is just the clay and surely a Lump of clay can survive being reorganized in all sorts of ways. This follows from the *Diachronicity of material objects*. However, the Lump and Statue are surely not identical. For one, the Lump exists on Monday, Tuesday, and Wednesday. But, the lump and the statue are distinct things since there are sentences true of the lump that are false of the statue, e.g., the lump exists on Monday but the statue does not. In addition, there are some things true of the Lump that are not true of the Statue. For example, the Lump can survive a smashing but the Statue cannot.

18.1.11 Temporal parts

Rather than think of objects as 3D material things moving through time, think of them as 4D objects *spread out* over time. Material objects have spatial parts while 4D objects have "temporal parts". If Tek is a human being who lived from 1980 to 2026, then there is the 1980 temporal part (consisting of all of his spatial parts at this time) but also his 2020 temporal part and 2026 temporal part.⁷

Now let's make some important observations about the parts of objects. First, a proper part of an object is distinct from the whole. Some examples:

1. Consider a jacket and the zipper of the jacket. These are distinct objects

- even though the zipper is part of the jacket.
2. The GPU of a computer is distinct from the entire computer.
 3. My hand is distinct from my entire body.

Second, two distinct objects of the same type can share spatial parts without being the same object.

Definition 18.15: Part-sharing principle

Two distinct objects A and B (of the same type) can share some but not all of their parts without being the same object.

Let's consider some examples:

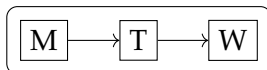
1. Two distinct intersecting roads share the some of their material parts without being the same road
2. Conjoined twins can share an organ without being the same person.

Third, just as two distinct objects x and y can share some (but not all) of their spatial parts without being one and the same object (e.g., conjoined twins sharing an organ, intersecting roads), the temporal part theorist claims that two objects can also share some (but not all) of their temporal parts without being identical.

Let's apply this insight to the Statue-Lump puzzle. On Tuesday, the Statue and Lump occupy the same space at the same time. According to the Standard View, since two distinct things cannot wholly occupy the same place at the same time and two distinct things cannot be made up of exactly the same stuff, there are not two distinct objects present on Tuesday. On the other hand, if material objects can survive change, then there are two distinct objects on Tuesday.

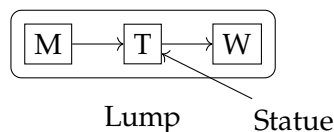
When we take on the temporal-part perspective, we look at the Statue and Lump as 4D entities with spatial and temporal parts.

First, consider the Lump. For the temporal-part theorist, the Lump is an object that exists four-dimensionally. It not only has material parts (clay particles) but also temporal parts: a Monday-part, a Tuesday-part, and a Wednesday-part.



Lump

Now consider the Statue. For the temporal-part theorist, the Statue is also an object that exists four-dimensionally. It not only has material parts (clay particles) but also temporal parts. However, unlike the Lump, the Statue only has a Tuesday-part.



For the temporal-part theorist, there are two distinct objects: Statue and Lump. They reject that these objects ever fully overlap or ever share all of the same parts and they hold the Part-sharing principle (that two distinct objects *A* and *B* can share some, but not all, of their parts):

1. Statue and Lump never completely overlap. Lump occupies Monday-Tuesday-Wednesday, but Statue only occupies Tuesday.
2. while Statue and Lump share a *temporal part* (Tuesday-part), they never share *all* of the same *temporal* parts. Since they do not share all of their parts, they are not the same object.

In short, the Statue and Lump are analogous to a zipper of a jacket. The Statue is a single part of the three-part Lump.

Let's summarize our results in the form of an argument:

- P1: There are paradoxes (contradictions) involving our understanding material objects (the Standard View).
- P2: If one accepts Eternalism and the view that material objects have temporal parts, then our understanding of material objects is no longer paradoxical (contradictory).
- C: Therefore, Eternalism is true.

The main problems with the above argument are (1) potential negative consequences associated with accepting temporal parts (the cure might be worse than the disease) and (2) whether this argument offers Eternalism any relative advantage over the Moving Spotlight theory.

Exercise 18.120

1. What does the temporal theorist say about the ship of Theseus?
2. Potential problem: Take the ship of Theseus when it is first created (time t_1). Given that the ship will become Planks and Replace, what am I looking at when I stare at the ship at t_1 ?

18.1.11.1 *Argument 4 - from Death's Badness*

In an earlier chapter, it was noted that if a theory meshes well with our intuitions, then this counts in the theory's favor. However, it was also noted that our intuitions do not always tell a consistent story. In some cases, our intuitions about the world conflict. Arguments that show our intuitions are contradictory are sometimes referred to as "paradoxes". One merit for a theory is its ability to better make sense of the world or to better organize the beliefs we have about the world. In other words, if there is a paradox (a conflict of intuitions) and a theory is able to eliminate that paradox, then this counts in favor of the theory.

In what follows, we will put forward an argument inspired by the philosopher Epicurus. Our approach will be to use this argument to formulate a paradox and then show how eternalism but not presentism can resolve that paradox. Let's begin our discussion with the articulation of the paradox before turning to how the paradox can be solved by eternalism.

18.1.12 *An epicurean timing paradox*

Let's begin with the first premise of the argument.

- P1: Death is bad for the person who dies.

Let's first articulate what P1 says and then briefly outline some misconceptions about what it says.

What P1 is saying is that when a person dies, there is a harm, evil, or badness that happens to the person who has died. Our death makes us worse off: it negatively impacts our well-being. Generally, P1 has qualified scope in that it is bad for people whose life still has prudential value. This is because when we die something of value is taken from us. And so, we are worse off for dying than if we didn't die.

For example, if Tek is young, enjoys life, and would have many positive future experiences, then death is a badness that befalls Tek as these positive experiences are taken away. In contrast, if Tek is older, does not enjoy life, and is in terrible pain, then death may not be a badness for him since it does not take away anything of value from Tek.

P1 does not say any of the following:

1. The deceased person's death is bad for the deceased person's family members or friends.
2. The deceased person's death is bad for society.

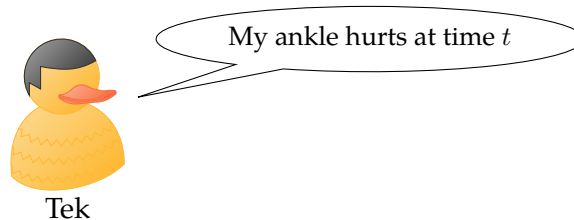
3. The act of dying was not bad for the deceased person
4. It is morally acceptable to kill a person for whom death would not be an evil.

All of the above may be true, but P1 is only about death being a badness for Tek.

With P1 in place, let's consider the second premise of the argument.

- P1: Death is bad for the person who dies.
- P2: If something is bad for someone, then it must be bad for that person at a particular time.

Let's consider an example. Suppose Tek twists his ankle. This is instrumentally bad for him. He cannot do the things he once enjoyed doing: run, walk, or sit without pain. But, it appears that insofar as Tek is harmed, then these harms occur at a particular time. The duration of the badness seems to fall sometime around when the injury occurs and the pain, suffering, and frustration that results as a consequence of the injury. The badness also appears to end once he has recovered, when the pain no longer is present, and when he can return to the activities he enjoyed prior to the injury.



Here is another example. Suppose Liz and her boyfriend break up. She is really sad about it for the first week (things are bad during this time). She is less sad in the second week even though there are moments when she is sad (things are less bad during this time). Finally, things are no longer bad for her when she's no longer sad at all. The time in which the badness occurred for Liz was the duration of her sadness.

If P2 were false, then this would imply there are some events that are bad for people but don't occur at particular times. This is to say that there are atemporal evils (timeless evils). But, it's hard to think of any evil that is atemporal in nature (bad but not bad at any particular time).

So, not only is P2 intuitively true but the implication of denying P2 (the existence of atemporal evils) is intuitively false.

With P1 and P2 in place, let's consider P3:

- P1: Death is bad for the person who dies.
- P2: If something is bad for someone, then it must be bad for that person at a particular time.
- P3: There is not time at which death is bad for the person who dies.

In considering P3, let's consider the three of the four possible times for which death might be bad for someone.

Theory	when death is bad
priorism	before the person dies
concurrentism	at the time the person dies
subsequentism	after the person dies
eternalism	always

First, someone might say that death is bad for someone *before* they die. This view is known as *priorism*.

Definition 18.16: priorism

Death is bad for a person before they die.

Priorism is implausible for several reasons. The primary reason has to do with backward causation. Suppose it is 2024 and Tek is alive. In addition, suppose that Tek will die sometime in the future. Priorism suggests that an event in the future (which may not have happened) is harming Tek in the present. This is counterintuitive since (1) it entails that events in the future determine events in the past and (2) events that have not happened are influencing events that are happening. More concretely, suppose that in three days, Tek will eat some poisoned meat. Priorism seems to say that Tek is harmed now by some future meat he will eat.

So, the conclusion seems to be death is not bad for someone before they die. Priorism is false.

A second account of the time at which death is bad for a person is known as concurrentism. This view says that death is bad for a person *at the moment in which they die*. In other words, the harm of death is concurrent with the moment of their death.

Definition 18.17: concurrentism

Death is bad for a person at the time the person dies.

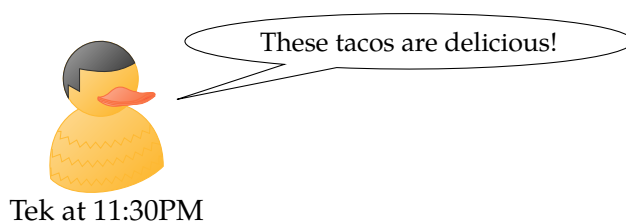
There are several reasons to think that concurrentism is false.

First, while dying may take a consider amount of time, death itself seems momentary or instantaneous. It may strike us as unintuitive that the totality of the badness of death can happen to us in such a short time. Consider Tek's injury or Liz's sadness after her breakup. Both of these harms have a duration where the badness begins, peaks, and gradually diminishes. In short, the moment of death is too fast it to be bad for you.

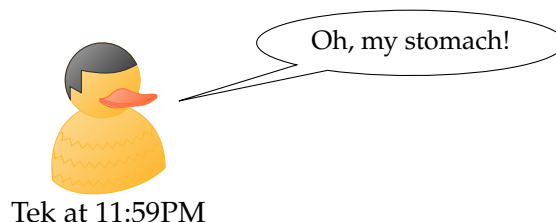
Second, death is bad for people because people "miss out" on experiences they would have had if they did not die. For example, suppose Tek is looking forward to going to his daughter's wedding. Unfortunately, Tek dies a month before the wedding. But when does Tek "miss out" on his daughter's wedding? It seems more natural to say that he misses out *the day that the wedding occurs* not on the day he dies. In other words, the badness of death occurs *after* he dies.

Third, it is claimed that concurrentism does not distinguish between *when the bad event occurs* and *when the bad event is bad for me*. It is thought that when Tek twists his ankle, the moment in which the ankle twists is the moment in which the bad event occurs. However, the badness of the event for Tek does not occur in this moment. Instead, it occurs after the event. That is, this event becomes bad for Tek once the twisting of the ankle has consequences for Tek's well-being. That is, the event is only bad for Tek when his ankle starts swelling, when he feels the pain, and when he can no longer walk easily.

Here is another example. Suppose that Tek eats a piece of rotten meat in a taco at precisely 11:30PM.



The bad event occurs at 11:30PM, but the event is not bad for Tek at 11:30PM. It is instrumentally bad for him an hour later when he becomes sick.

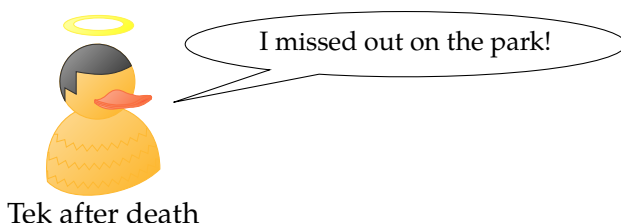


Based on these considerations, the time at which death is bad is not when death occurs. Therefore, concurrentism is false.

The last theory that we will consider is known as *subsequentism*. According to the priorist about the badness of death is bad for someone *after* they die.

Subsequentism has some initial support. Intuitively, if I bang my finger, the badness occurs after the event: when the pain kicks in or when I can no longer use my finger. With respect to death, the badness of death is due to the fact that there is something of value after our death that we could have had or experienced if we had not died.⁸ On this account, death is bad because it *deprives* us of certain goods after our death.

Let's consider an example. Suppose Tek plans to go to an amusement park tomorrow. He loves amusement parks and if he goes, he will have a great time. Tragically, Tek dies before getting to go to the amusement park. For subsequentism, the badness of death for Tek occurs after his death. Namely, on the day he was supposed to go to the amusement park, Tek is harmed by not being able to go.



While intuitive, one of the main reasons that subsequentism is implausible is because (1) once you die you don't exist and (2) it is impossible to harm non-existent beings. Let's formulate both of these concerns as two principles:

Definition 18.18: death-existence principle

If object O is dead, then O does not exist.

Definition 18.19: harm-existence principle

If O does not exist, then O cannot be harmed.

First, the death-existence principle is intuitive. If Tek dies, then while Tek's corpse may exist, Tek does not.

Second, the harm-existence principle is also intuitive. Consider that you cannot harm unicorns since they don't exist. Similarly, you cannot harm a person after they die (since they don't exist).⁹

Time of harm	Objections
Before death	Death has not happened yet
At death	Too fast, future bad events, when the event is bad for me
After death	cannot harm non-existent things

Recall that P3 says that “there is not time at which death is bad for the person who dies.” We have shown above that it is implausible to say that the badness of death occurs at a time before death (priorism), at death (concurrentism), or after death (subsequentism). The conclusion of the epicurean argument can now be added:

- P1: Death is bad for the person who dies.
- P2: If something is bad for someone, then it must be bad for that person at a particular time.
- P3: There is not time at which death is bad for the person who dies.
- C: Therefore, death is not bad for the person who dies.

Notice that the above argument proceeds by intuitive premises to a conclusion that contradicts one of the premises (P1). We are seemingly committed to (1) death is bad for the person who dies and (2) death is not bad for the person who dies. This is a contradiction. If a theory of time can remove the paradox, then this would count in favor of the theory. With that in mind, let's briefly consider two approaches: the presentist and the eternalist.

18.1.13 *Presentist and eternalist solutions to the epicurean paradox*

Recall that P3 of the epicurean argument says “there is *not* a time at which death is bad for the person who denies.” One way to deal with the epicurean argument is to deny this premise. Doing so would seemingly involve accepting either priorism, concurrentism, or subsequentism.

As a theory of time, presentism only seems compatible with concurrentism. The presentist cannot accept priorism because this would involve positing a future event (one's death) that does not exist. In addition, they seemingly cannot accept subsequentism since it would involve positing a past object (the deceased person) who no longer exists. The only option then is to say that the time at which death is bad for a person is the precise moment that they die. However, as we pointed out, there are several reasons why concurrentism should be rejected. It is also unclear how the presentist would respond to these objections.

In contrast, eternalism appears compatible with subsequentism (death is bad after the person dies). Recall that subsequentism says that the badness

of death occurs when the deceased is deprived of various experiences that would have made their life go well. To illustrate again, suppose Tek dies in 2023 and, if he had lived, he would have gone onto do the following in 2024:

1. had fun at a wedding on 2 Jan 2024
2. enjoyed 200 walks throughout 2024
3. received a great gift on 25 Dec 2024

The badness of Tek's death happens when these events occur in 2024. However, as we noted, the main reason for rejecting subsequentism is because (a) you don't exist after you die (death-existence principle) and (b) only existent entities can be harmed (harm-existence principle). However, the eternalist can argue that (a) and (b) are false. Here we will only consider the rejection of (a).

For the eternalist, *after you die, there exists times t where you exist.*¹⁰ These times are the times before you die. Suppose again that Tek is born in 2000 and dies in 2023. If Tek dies in 2023 and misses out on events in 2024 that would have made his life go well, then the person that is harmed in 2024 is the person that exists in 2000-2023. That is, Tek is harmed by events that occur after his death.

Let's summarize the argument in favor of eternalism as follows:

- P1: The Epicurean argument leads to paradox.
- P2: If a theory of time can resolve this paradox, then this counts in favor of the theory.
- P3: Presentism cannot resolve the paradox.
- P4: Eternalism can resolve the paradox by contending that death is bad for a person after they die.
- C: Therefore, there is a reason to rationally prefer eternalism over presentism.

Exercise 18.121

1. Try to articulate the eternalist's argument in your own words.
2. In order for the eternalism-subsequentism combination to work, it must also show that posthumous harm is possible (that people can be harmed after death). Try to give an argument or an example that makes posthumous harm plausible.

The evidence in support of the claim that people have traveled through time is dubious. However, the absence of evidence of time travels does not entail that time travel has not occurred. To argue that time travel has not occurred or cannot occur is a different matter. Let's consider some arguments against time travel.

19.1 FROM THE NON-ABUNDANCE OF TIME TRAVELERS

Before considering the next argue, ask yourself the following question: if you could travel through time, where in time would you go? Would you visit some important event in the distant past or travel to the future to enjoy the fruits of advanced technology? Would you revisit a time in your life or would you travel to some time you could only dream of? With that location in mind, what sort of costs would you bear to travel to those locations in time? Would you be willing to take a one-way trip or suffer some serious disease? Would you be willing to abandon your family or part with your entire life savings?

In Arthur C. Clarke's essay "About Time" in *Profiles of the Future*, Clarke^[23] notes:

"The most convincing argument against time travel is the remarkable scarcity of time travelers. However unpleasant our age may appear to the future, surely one would expect scholars and students to visit us, if such a thing were possible at all. Though they might try to disguise themselves, accidents would be bound to happen -- just as they would if we went back to Imperial Rome with cameras and tape-recorders concealed under our nylon togas. Time-travelling could never be kept secret for very long; over and over again down the ages, chronic argonauts (to use the original and singularly uninspiring title of Wells' *The Time Machine*) would get into trouble and inadvertently disclose themselves. As it is, the chief evidence of a security leak from the future appears to be the notebooks of Leonardo da Vinci. Their parade of inventions from the succeeding centuries is astonishing, but hardly conclusive proof that fifteenth-century Italy had visitors from elsewhere."

Similarly, in Jack McDevitt's "Time's Arrow" (1989), McDevitt writes:

“If [time travel] could be done, someone will eventually learn how. If that happens, history would be littered with tourists. They’d be everywhere. They’d be on the Santa Maria, they’d be at Appomattox with Polaroids, they’d be waiting outside the tomb, for God’s sake, on Easter morning.”

Finally, in order to gather experimental evidence that time travel is impossible, in 2009 the theoretical physicist Stephen Hawking sent out a written invitation to time travelers to attend a party he was throwing. He gave the exact time (12:00 UT on 06/28/2009) and location. However, no time travelers attended the party. Hawking is reported as saying the following:

“I gave a party for time-travelers, but I didn’t send out the invitations until after the party. I sat there a long time, but no one came.”

These remarks and others suggest the following argument:

- P1: If time travel has occurred in the future, then there would be evidence in the present of time travelers.
- P2: There is no evidence in the present of time travelers.
- C: Therefore, time travel has not occurred in the present or in the future.

The claims and argument made about time travel here is analogous to criticisms of extraterrestrial visitors (see Fermi paradox). The reasoning is that if aliens exist, then they would have either visited humans (or at least given humans a sign of their existence) by now. But they have not, so this counts as evidence that they do not exist (evidence of their absence).

In terms of evaluating this argument, the primary problem involves P1. The argument relies upon evidence in the present (no present evidence of time travel) to make a claim about the future (no time traveling going on in the future). In other words, a critic of P1 might argue that it is possible that while time travel is occurring in the future, time travelers have a reason for not showing up during the time we exist (or any time before our existence). Some reasons include:

1. Time travelers have no desire to return to our time period as there is nothing about our time period that would make them want to visit our time period. Hawking assumes that future scientists with their huge advancements in knowledge and technology would want to visit a theoretical physicist in the early part of the 21st century. In the future where time-travel is possible, physics is probably so advanced that a theoretical physicist from the 21st century might seem like a cave man.

2. McDevitt and Clarke both assume that there would be a multitude of time travelers at various historical events. However, the mere *existence* of a technology does not entail its *abundance*. Time travel is sometimes characterized as a dangerous activity, whose success is low (or by mere accident), and whose trips are one-way.[39] Nuclear weapons and reactors exist, but it isn't the case that everyone has a nuclear weapon or reactor.
3. The argument assumes that future time travelers, with their advanced knowledge and technology, would feel the need to send humans (on a potentially one-way mission) through time to satisfy their personal intellectual curiosities. Such an act would be sufficient to satisfy our curiosity but surely isn't necessary. Another way this intellectual curiosity could be satisfied would be simply to view the distant past from the future. For example, future time travelers might use a "Chronovisor" that is capable of displaying past events on a monitor. This mirrors our approach to trying to obtain knowledge about distant objects in space (e.g., we send satellites, rather than people, on one-way missions into outer space).

Exercise 19.122

Many people are persuaded by the argument from the non-abundance of time travelers. Evaluate it!

19.2 ARGUMENTS FROM PARADOX

In an editorial titled "Time-Travel" in *Asimov's Science Fiction Magazine* (April 1984), Isaac Asimov wrote the following:

"The dead give-away that time travel is flatly impossible arises from the well-known 'paradoxes' it entails ... So complex and hopeless are the paradoxes ... so wholesale is the annihilation of any reasonable concept of causality, that the easiest way out of the irrational chaos that results is to suppose that true time travel is, and forever will be, impossible."

What Asimov is asserting is that the very possibility of time travel generates paradoxes (contradictions). And since there are no true contradictions, there is no time travel. Before considering these paradoxes, let's first define time travel a little more precisely.

19.2.1 Defining time travel

Let's consider some different ways of defining time travel:

Definition 19.1: time travel (simple)

Time travel is traveling through time.

This definition is not what people mean when they talk about time travel because it would imply that if we persist through time, then we are all time travelers. Let's revise our definition. Essentially, a person counts as traveling through time in at least two different cases. First, if there is a discrepancy between the *speed* at which the traveler moves through time and the speed of everything else, then that person has traveled through time. Second, if there is a discrepancy between the traveler's *direction* of travel and the direction of everything else, then that person has traveled through time.

Definition 19.2: time travel (movement)

Time travel occurs in one of two ways: (1) leaving one time dt (departure time) and arriving at another time at (arrival time) and the interval of time for the time traveler is different than the interval of time for non-time-travelers or (2) the temporal relations (e.g., earlier than) for the time traveler differ from the temporal relations of non-time-travelers.^a

^aSo, if $(It_1 = [dt, at])$ is the interval of time for the time traveler and $(It_2 = [dt, at])$ is the interval of time for non-time-travelers, then time travel occurs when $It_1 \neq It_2$. Or, if e_2 occurs later than e_1 , then time travel occurs if the traveler experiences e_2 earlier than e_1 .

Let's consider the two ways that time travel can occur.

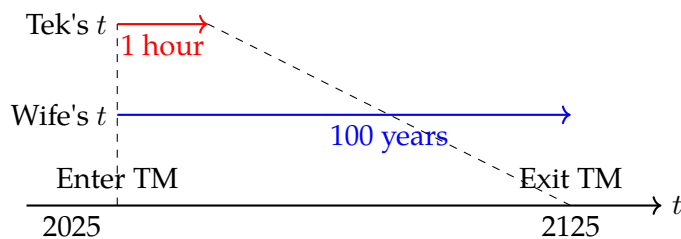
First, suppose Tek enters a time machine on 1 January 2025. He sits in the machine for approximately one hour. He monitors the amount of time he sits in the time machine by periodically looking at his wristwatch and counting how many times his heart beats (he knows that his heart beats around 60 times per minute and so he plans to leave the machine after 3600 beats). When he exits the machine, the interval of time between when he entered the machine and exited the machine was 1 hour.

However, when Tek exits the machine, the calendar reads 2125. His wife is dead. His wife died several years ago. Even his children are dead. Countless events have occurred, wars, rotations of the earth around the sun. While his

heart beat 3600 times, the hearts of those outside of the time machine beat hundreds of millions of times. In short,

1. Tek traveled 100 years (2025 to 2125) in one hour.
2. Everyone else traveled 100 years (2025 to 2125) in 100 years.

Some argue that the above is inconsistent (the time discrepancy paradox): It is impossible for the interval of time between two dates to be larger or smaller than it is. If the time between 2025 and 2125 is 100 years, then it 100 years. It is absurd to say that it is also 1 hour.¹



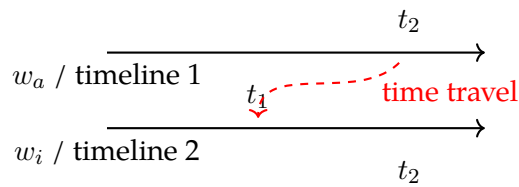
To resolve the time discrepancy paradox, a distinction is typically drawn between *personal time* and *external time*. *External time* is “time itself” or time for the vast majority of objects in the universe (e.g., people, animals, rocks, planets) and the personal time of the traveler is “the time measured by his wristwatch.” [71, p. 146] This does not mean that by simply changing the time on your watch, you are traveling through time. It also does not mean that if your watch stops, you are frozen in time. Instead, Tek's personal time can be measured in several ways: by a wristwatch, by counting the number of heart beats, by counting the number of times he need to breath, by counting the number of ideas pass through his mind, or by considering the how long it takes his fingernails to grow or any other aspect of Tek's biological growth, development, and decay.

Unlike external time, which is thought to have a relatively fixed structure, personal time is said to have a much more flexible structure, one that permits the possibility of time travel: “personal time can run in reverse, repeat itself, and generally get tied up into all sorts of knots.” [148, p. 5]

So, it is not the case that when Tek travels to the future, the interval between his departure time and arrival time is *both 1 hour and 100 years*. Rather, there are two intervals: 1 hour of personal time and 100 years of external time.

¹As Holt[56] puts it, “If the time traveler departs at t_0 and arrives at his destination five minutes later, and if the time at which he arrives is forty years after his departure, then the time of his arrival as reckoned by his own clocks and calendars and by the clocks and calendars of the stay-at-homes is both t_0 plus five minutes and t_0 plus forty years.”

Two points. First, the difference between external time and personal time does not involve two different time dimensions (different timelines). This is sometimes called “multiverse time travel”. To illustrate multiverse backward time travel, suppose you are in this world w_a at t_2 . When you travel back in time from t_2 to t_1 , you travel to a different world w_i that is identical in every way that this world was at t_1 . As you act in world w_i , differences emerge between our world w_a and this new world w_i . The usual response to this type of “time travel” is that it is not time travel at all. It is just traveling to a different possible world.



In contrast, when talking about time travel, personal time is on the same single timeline as external (real) time.

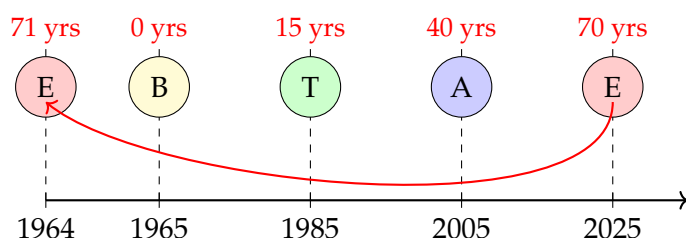
It was mentioned that there were two ways to travel through time. The first was that the amount of time it takes you to travel through time differs from the rest of the world. The second is if there is a discrepancy between your temporal relations and the temporal relations that apply to everyone else.

Suppose Tek has previously traveled forward in time 100 years. His wife and children are dead. He regrets his decision and so he hops back into his time machine on 1 January 2125, wishing to travel to the past. Just as before, he sits in the machine for one hour. He looks at his wristwatch, he counts his heart beats, he thinks about the normal amount of experience (boredom) one would endure in an hour. When he exits the time machine, he notices that he has only traveled 1 hour into the past. It is now 31 December 2124. While the *speed* at which he has traveled through time is the same speed at which everyone else travels, what differs is the *direction* and the temporal relations between events.

1. For Tek, the order of events are: 1 January 2025, 1 January 2125, then 31 December 2124
2. For everyone else, the order of events are: 1 January 2025, 31 December 2124, 1 January 2125

Let's consider an example where both forms of time travel are at play. Humans are babies, then adolescents, then adults, then elderly. Suppose Tek is

70 years old and travels 81 years into the past (taking 1 year to do so). After traveling into the past, Tek dies the second before he is born.



First, Tek's travels 80 years in external time in 1 year of personal time, which differs from the speed in which other people travel through time. Second, Tek's temporal relations are also out of wack. The stages of his biological development do not correspond to normal people. Relative to external time, Tek is elderly, then a baby, then an adolescent, then elderly again. Rather than claim that Tek is a unique kind of object that is born elderly, then becomes a baby, and then develops as normal, we say that Tek has traveled through time. That is, to accommodate the normal stages of objects, different temporal coordinates are assigned to the traveler (their personal time) so we can say that the traveler is a baby, then a child, then an adult, then elderly. As Wasserman[148, p. 7] puts it, "the revised proposal is that personal time is an assignment of coordinates to person-stages that preserves the regularities we ordinarily see in people."²

Exercise 19.123

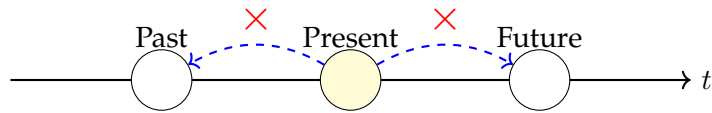
Create an example where you travel through time. Use the distinction between external and personal time. Include a diagram with your explanation.

19.2.2 Time travel paradoxes

Let's briefly articulate and illustrate several time travel paradoxes. These paradoxes suggest that time travel is *impossible*.

First, there are *temporal ontological paradoxes*. You can only travel to some place if that place exists. I cannot travel to Biverpool if Biverpool doesn't exist. Similarly, I can only travel to a time if that time exists. Now suppose that only the present time exists. If only the present exists, then backward and forward time travel is impossible because it would imply traveling to times that do not exist.

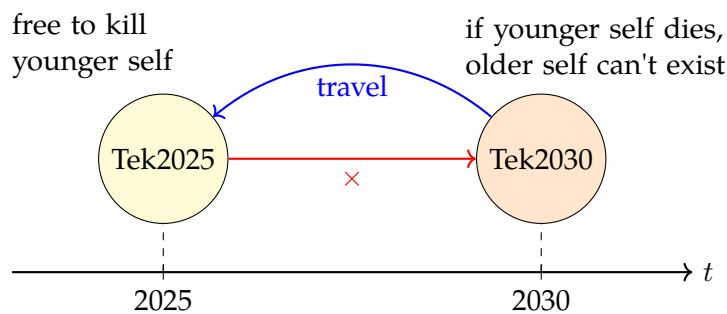
²For more on the distinction between personal and external time, see [56, pp. 7–10]



Second, there are *paradoxes of freedom*. In these paradoxes, the existence of time travel and free will conflict. Namely, our free will allows us to do things that are impossible. One example is the *self-suicide paradox*. Suppose it is 2030. Tek has been taking shooting lessons and is accurate with a rifle up to 200 meters. Now imagine that Tek in 2030 travels back in time to 2025 to his college graduation. Tek sees that his younger self is receiving his diploma, traveler Tek shoots and kills the younger version of himself. Both of the following seem true:

1. Tek is free to kill his younger self. He has the means and the opportunity to do so. He is a skilled shooter. It would be true to say *2030-Tek can shoot 2025-Tek*
2. *2030-Tek cannot shoot 2025-Tek* kill his younger self. If Tek dies in 2025, then he would not survive to 2030, and so would no longer be able to kill his younger self.

It seems that (1) and (2) are both true but they contradict each other. This contradiction is made possible by time travel and so we have a reason to believe that time travel cannot exist (as it would make this contradiction possible).

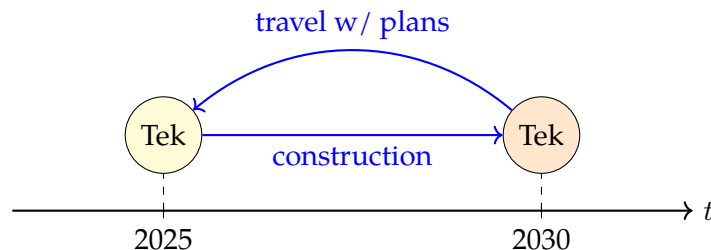


Exercise 19.124

The self-suicide paradox is a simpler version of the *grandfather paradox*. Try to explain why going back in time to kill your grandfather is paradoxical.

Third, there are *causal paradoxes*. These paradoxes emerge when the cause of an event is seemingly *after* its effect. A classic example is the *bootstrap paradox*.

Imagine Tek in 2025 receives a book that contains plans for how to build a time machine. He uses these plans to build the time machine. As soon as the time machine is completed, he returns to 2025 to give himself the plans he used to build the time machine.



There are at least two ways of teasing out the paradox here. First, consider the following:

1. What caused Tek to build the time machine in 2025 is Tek's traveling back in time in 2030 with the plans.
2. But, what caused Tek to travel back in time in 2030 with the plans is that Tek built the time machine in 2025 using the plans.

We thus seem to have a causal loop, where a future event E_F is caused by a past event E_P but that past event E_P is caused by that future event E_F .

Another way of teasing out the paradox is to ask the following question to 2025 and 2030 Tek: *where did the plans for the time machine come from?* If we ask 2025-Tek, he will say "I was given them by my future 2030-self." If we ask 2030-Tek, he will say "I was given them in 2025 by 2030-Tek, who, I guess is me. So, I guess I gave them to myself." But, again, if we ask *where did these plans come from, who wrote these plans?* Tek won't have an answer. From his perspective, these plans always existed.

A related problem with causal loops involves the *wear-and-tear paradox*. Suppose that when Tek first gets the time-travel plans, they are in perfect condition. As he uses them to construct the time machine, they suffer minor wear. So, by the time he arrives in 2030, the plans are slightly damaged. Since 2030-Tek will take these plans back in time, again and again, these plans will continuously suffer more damage until they are unusable. In short, objects caught in time-loops appear to wear away to nothing.

Let's consider another example of a causal loop. Suppose Tek is a time traveler. He goes back in time, falls in love, impregnates Liz, and then immediately abandons her. Liz has a baby that she names "Tek", who grows up without a dad. Once he becomes an adult, he wishes to go back in time to

find his father. So, he enters a time machine, goes back in time, meets Liz, impregnates her, and then leaves her as he is in search of his father. Again, we have a causal loop with no clear starting point: Tek must be go back in time to be born, but he must be born to go back in time.

Exercise 19.125

Create another example of the bootstrap paradox. Here is an example. Suppose an old man teaches you a song. You then go back in time and teach that song to the old man when he is a child.

Fourth, there are *identity paradoxes*. One example of an identity paradox is the *self-visitation paradox*.

First, it seems plausible to say that an object cannot be in two distinct places at the same time. For example, if Tek is in New York, he cannot be in Chicago at the same time. However, suppose Tek is in Chicago in 2025 and Tek in 2030 travels to 2025 arriving in New York. This would imply that Tek is in two places at the same time, which is impossible.

Second, suppose Tek has black hair in 2025 but white hair in 2030. If Tek in 2030 travels back to 2025, then it is true that Tek's hair is wholly black and wholly white. But this is a contradiction, so time travel must be impossible.[cf. [126](#), p. 157]

Exercise 19.126

Do you think time travel is impossible?

Part V

Back Matter

SOLUTIONS TO THE EXERCISES

Solution 14.96

2. Here are some options: Pastism: only the past exists, the present is so fleeting it doesn't exist. Past-Futurism: The present is so fleeting, it does not exist. We are half in the past and half in the future.

NOTES

CHAPTER 1 -- WHAT EXISTS?

1. The standard response is to argue that there is no empty set. Spencer^[127], in a review of Coggins' *Could There Have Been Nothing? Against Metaphysical Nihilism* rejects Coggins' argument in favor of this conclusion and provides the following: "Here is a stronger argument for the conclusion that there is no empty set. First, all sets existentially depend on something. Second, the only things that sets existentially depend on are their members and whatever their members in turn depend on. It follows from these two claims that there is no empty set. For suppose there is an empty set. Then, from the first claim it follows that it existentially depends on something. But, from the second claim it follows that the only things it depends on are its members and whatever those members in turn depend on. But, by hypothesis, it has no members. So, there is no empty set. The same kind of reasoning can be used to show that there is no uninstantiated universal."

CHAPTER 2 -- DOES GOD EXIST?

1. An exception is Robert M. Price who contends that "quite likely there never was any historical Jesus." (p.55)
2. The argument thus commits what is called the "fallacy of composition", which says that just because the parts of something have a property does not mean the whole has that property.
3. "No state shall make or enforce any law which shall abridge the privileges or immunities of citizens of the United States."
4. For example, children were taught to pray with their hands folded together and looking upward. This contrasts with facing Mecca, raising hands, prostrating on the ground, or restricting prayer to a synagogue.
5. The Court expressed slippery slope worries noting that after the British made the Book of Common Prayer the only accepted form of religious service, they later passed the Act of Uniformity, which compelled "all Englishmen to attend those services and to make it a criminal offense to conduct or attend religious gatherings of any other kind" *Engel v. Vitale*. For a video interviewing one of the children involved in the *Engel v. Vitale* case, see https://www.youtube.com/watch?v=Gz7_J1gTtms
6. "A wise man [...] proportions his belief to the evidence."^[58], sec. 10]

7. First, suppose there is some hypothesis H . For example, suppose we think that a particular coin is a biased rick coin that is more likely to turn up 'heads'. We flip the coin 10,000 times and it turns up heads 9,999 times. We now have reason to believe H . However, also note that our experiments also count against the hypothesis that the coin is a 'fair coin'. It is still possible that the coin is fair, but it is extremely unlikely.

Second, what evidence do we have for the 'laws of nature', e.g. gravity? The laws of nature are similar to our experiment involving the coin above. We run repeated experiments and these experiments repeatedly support the presence of certain regularities in nature ('laws governing certain phenomena'). In addition, the more observations we make, the probability that they will not hold in the future increasingly diminishes.

Just as the evidence in support of the probability of the coin being biased counts against the probability of it being fair, the evidence in support of the probability of the laws of nature holding is said to count against the probability of someone's account of a miracle occurring being genuine. It is not the case that no miracle could ever occur, but instead that our past experience so strongly supports the laws of nature that it is more likely that the person testifying to the miracle is mistaken rather than there being a violation of such a law.

To simplify, P5 asks us to consider these two options:

- H1: Laws of nature continue to hold and person reporting the miracle is mistaken
- H2: Person reporting a violation of the laws of nature.

P5 says that since there is so much evidence in support of the laws of nature, H1 is more likely than H2.

8. Here is how Pierre Bayle put it: "[A] religion, however false it might be, passes from fathers to children without any obstacle, and it passes by way of education as easily to a thousand people as to one, in such a way that, if education is the sole means of propagation, the reason why one religion has only one hundred thousand adherents, while another has two hundred thousand, is solely that there have been fewer children in the one than in the other." - Bayle, *Oeuvres*, 3:207b-208a.

9. There is much more to say concerning the argument from common consent. For a delightful history of the common consent argument, see [106]. For more on primitive cultures and their beliefs in myths, see [138]. In addition, there is a debate among those who accept the common consent argument concerning *how* the consent comes about. Do we come to know God

through the senses? Is the idea of God innately impressed upon us? Or, perhaps, it depends upon knowing some other argument for God's existence. In addition, we sort of assumed that our evolutionary explanation for nature spirits (and God for that matter) was accurate. Some might point out that this explanation is not a good one, pointing to prisoner-dilemma scenarios where *selfishness* is more beneficial than *cooperation*. Further, some anthropologists suggest that it is belief in the afterlife (not belief in nature spirits) that generates the sense of "being watched" and so our evolutionary explanation misses its target.

10. As Park[cf. 97, p. 326] puts this point, 'Even though all of our experiences demonstrate that all physical objects we come in contact with have a cause to their existence, this allows us to only induce that all physical objects must have a cause. We cannot conclude based on our experience with physical objects that all physical *and abstract* objects must have a cause to their existence.'

CHAPTER 4 -- DO PERSONS EXIST?

1. This example is taken from [142, pp. 135–136]
2. For an overview of efforts to detect extraterrestrial intelligence, see [153]

CHAPTER 5 -- HOW DO WE EXIST THROUGH TIME?

1. It is important to note here that this question might be too specific since we might consider a slightly more general question of what allows *me* to maintain my identity over time. That is, if *A* is a me at t_1 and *B* is me at t_2 , under what conditions is *A* identical to *B*? In this case, we are not assuming *A* and *B* are persons.
2. Swinburne[116, p. 30] provides a semi-formal representation of his argument. See [1]
3. There are at least two interpretations of this option. On what I take to be the most radical interpretation, contrary to our senses, the fetus no longer continues as a human organism. Humans develop from biological beings to persons and, when they die, back to biological beings. On a more sensible interpretation, what occurs is a "fight" over what fundamental *type* of being the human being is. The idea here is that in the same place and time, there are two competing fundamental types: the human organism and the person. Once the person appears, the human organism "dies" in that it loses its status as the fundamental type to which the being belongs.
4. Red blood cells live approximately **4 months**, white blood cells live ap-

proximately **1 year**, skin cells only last 2-3 weeks. Some cells (neurons) in the cerebral cortex are not replaced upon death. That is, these cells remain with the human organism throughout the entirety of its life.

5. Locke[72, Bk. II, xxvii, 11], for example, took it to be obvious that a person could survive the loss of some of its minor parts, writing that there is “no question about the same person, though the limbs, which but now were part of it, be cut off.”

6. These approaches are also sometimes referred to as “Lockean theories” as the philosopher John Locke[72, Bk.II, Ch. XXVII, Sec.9] defined a person as “a thinking intelligent Being, that has reason and reflection, and can consider it self as it self, the same thinking thing in different times and places; which it does only by that consciousness, which is inseparable from thinking.” For example, Parfit[96, p. 6] refers to psychology-based approaches to personal identity as “Lockean”.

7. Swinburne[116, p. 9], for example, says that “A person is who he thinks that he is.”

8. This example is from Locke who accepts this as a consequence of his theory of personal identity. Locke[72, Bk.II, xxvii, 16]writes: “Had I the same consciousness, that I saw the Ark and *Noah's* flood, as that I saw an overflowing of the *Thames* last winter, or as that I write now, I could no more doubt that I, that write this now, that saw the *Thames* overflowed last winter, and that viewed the flood at the general deluge, was the same *self*, place that *self* in what Substance you please, than that I that write this am the same *my self* now whilst I write (whether I consist of all the same substance, material or immaterial, or no) that I was Yesterday. For as to this point of being the same *self*, it matters not whether this present *self* be made up of the same or other substances, I being as much concern'd, and as justly accountable for any action was done a thousand years since, appropriated to me now by this self-consciousness, as I am, for what I did the last moment.”

9. So, what does it mean for *A* and *B* to have the same personality?

10. For the available space of theories, see [66, p. 55] who notes that there are five options and Parfit[94, p. 5] who points out three possible options.

11. Note that Parfit also says that rather than there being no fact to the matter, he often suggests instead that there are true answers. In cases involving division, for example, he contends that neither of the resulting people are identical with the divided person.[95, p. 262]

12. Parfit [94, p. 3] briefly notes that some might contend that hypothetical

cases involving personal identity “could never occur.” The substance of his paper “Personal Identity” is that even if they could occur, they would not present a problem.

13. As Sauchelli[112, p. 409] puts it “if the holding of R consists in the holding of (strong) psychological connectedness *and* continuity, then (1) R depends on the holding of psychological connections, and (2) the holding of some of these connections (e.g., memory connections) require certain appropriate types of causal dependence.”

14. Katherine Hawley[53, p. 610] contends that “[p]roblematic fission cases are those in which we compare non-branching continuity with branching continuity.” According to Hawley, a problematic fusion case would “compare the non-branching evolution of an earlier object into a later one with a process which differs only in that the later object has an extra causal ‘source’ in a further object also continuous with it.”[53, p. 610] Hawley gives two ways in which a problematic fusion case could occur.

In the first, Tek’s body is destroyed but his brain states are transferred into a new body. Let’s call the brain states post-transfer BS_1 . This is to be compared with a fusion case where Tek and his doppelganger are simultaneously transferred into the same body. Let’s call the brain states post-transfer BS_2 and let’s say that $BS_1 = BS_2$. For Hawley, while the brain states are intrinsically the same, what differs between the non-fusion and fusion scenarios is that in the fusion scenario, Tek’s brain states are causally overdetermined (the result of two rather than one psychology). That is, in the fusion case there is an *extra cause* determining the brain states at t_2 .

In the second case, suppose Tek can survive the destruction of his body into a new body via the transfer of some of his brain states. He can survive this even though some of his brain states are destroyed in the process (via the use of the technology that allows for the transfer). In other words, we have an imperfect transfer of Tek’s brain states from body B_1 to body B_2 . Intuitively, Tek survives. Now consider a fusion case where instead of some of Tek’s brain states being destroyed, they are replaced with another person (Keanu Reeves). In the fusion case, there is a single body B_2 with two rivals for identity: Tek and Keanu Reeves.

CHAPTER 7 -- ARE WE FREE?

1. An example is Frankfurt. Kind[66, pp. 17–19] considers Frankfurt’s account. In brief, this theory contends:

1. some entities have desires (first-order desires)

2. some entities that have desires also have desires about their desires (second-order desires), e.g. I want to do drugs, but I don't want to want to do drugs
 3. some entities that have first-order desires don't have second-order desires (they don't have desires about their desires)
 4. we have a *first-order volition* (a *will*) when our desires move us to act, e.g. when I sleep late because I want to sleep late, this is a case where my desire moved me to act and so it is an exercise of my will
 5. a *second-order volition* is when we have desires about which of our desires moves us to act (it is when we care about which desire we act upon (e.g. I may desire to eat a cookie but also desire to stay on my diet, if I have a desire about which of these two desires wins out, then I have a second-order volition, a volition about my first-order volitions))
 6. an entity is a person only if they have second-order volitions
2. This type of freedom is sometimes called: surface freedom, freedom of choice, freedom to act, free action, etc

CHAPTER 8 -- IS FREE WILL POSSIBLE?

1. Putting the argument into argument standard form gives us the following:
 - P1: Every proposition takes exactly one truth value: true or false (principle of bivalence).
 - P2: There are propositions about the future.
 - IC: Therefore, every proposition about the future takes exactly one truth value: T or F.
 - P3: If every proposition about the future P takes exactly one truth value (T or F), then it must be the case that P or not- P is true (exactly one is true, the other is false).
 - P4: If P is true, then there is nothing anyone can do to change that P is true. And, if P is false, then there is nothing anyone can do to change that P is false. The truth value of future propositions are unchangeable.
 - P5: If the truth value of future propositions are unchangeable, then if P is true, then we have no power to do not- P .
 - P6: If the truth value of future propositions are unchangeable, then if P is false, then we have no power to do P .
 - IC2: Therefore, either we have (1) no power to do not- P or no power to do P .
 - P7: If we are free concerning some state of affairs expressed by P (e.g.,

lifting our hand), then we have (1) the power to do P and (2) the power to do not- P .

- IC3: IC2 and P7 are contradictory.
- C: Therefore, we are not free.

CHAPTER 9 -- CAN WE BE FREE IF WE ARE DETERMINED?

1.

- P1: $N(P_2)$
- P2: $N(P_2 \rightarrow P_3)$. Intuitively true.
- C: $N(P_3)$. Rule β .

CHAPTER 10 -- LIBERTARIANISM

1. In other words, an action / choice $E1$ is free if and only if S there is a nondeterministic cause for the reasons $R1$ (beliefs, knowledge, intentions, desires) and $R2$, which lead, respectively, to two different actions / choices.

2. As O'Connor and Franklin put it: "On the event-causal picture, the agent's causal contribution to her actions is exhausted by the causal contribution of her reasons, and yet her reasons leave open which decisions she will make, and this seems insufficient for self-determination."

3. As Taylor puts it, "One can hardly affirm such a theory of agency with complete comfort, however, and not wholly without embarrassment, for the conception of agents and their powers which is involved in it is strange indeed, if not positively mysterious." (p.218)

CHAPTER 15 -- DOES ONLY THE PRESENT EXIST?

1. Despite some arguments, it is standardly thought that the disagreement between presentism and non-presentist theories is a substantial, rather than purely linguistic, disagreement. As Ingram puts it, "Presentists and eternalists disagree about what exists; eternalists think that there are dinosaurs, Roman Emperors, and Mars outposts, for instance, while presentists deny that there are any such wholly past or merely future entities at all. This is a genuine ontological disagreement, not a merely verbal dispute." [60, p. 12]

CHAPTER 16 -- THE MOVING SPOTLIGHT THEORY

1. For example, the argument from ontological simplicity is put forward by Power [102, p. 53], who points out that one reason to accept MST is because it is simple from an ontological perspective. Power writes: "The advantage

of the moving spotlight is that it folds both temporal passage and the real present into one metaphysical entity that has a simple-seeming metaphysical feature: it is a unique present that changes (it "moves")."

2. Note: Not all these properties conflict. There is no contradiction between:
 - (1) 2030 is future in the present (if 2022 is present, the year 2030 is future in the present), AND - (2) 2030 is future in the future (if 2022 is present, then 2025 is future, and so 2030 is future in the future)

CHAPTER 18 -- DO ALL TIMES EXIST?

1. But why is my experience confined to specific times? Why is this time the only one I have access to? Why is my experience locked into the time that it occurs? Dainton's[32, p. 30] answer is that our consciousness is only capable of being aware of a limited amount of time. As an analogy, consider spatial separation. Notice that you can only be in one space. You are locked into the space that you occupy. In terms of temporal separation, your consciousness is locked to the time that it occurs. It cannot occupy (or see) more than a limited amount of time. The implication is that if there were a consciousness that were capable of being aware of more time, it would be aware of all the time we are aware of. It would see 10 years like we see 1s, in an instant.

2. <https://www.sciencealert.com/here-s-why-you-re-fooled-by-this-classic-visual-illusion> For more, see [119].

3. See [79, ch.XXXIII, sec. 31][117, p. 213]

4. See [79, ch.XXXIII, sec. 315-6][117, pp. 213-216].

5. "An object *o* persists over time if and only if it is present at multiple distinct times (where *o* is present at a time *t* iff its complete spatiotemporal path--i.e., its career trajectory--passes through time *t*.)"

6. People love to criticize philosophy for being detached from reality and having zero practical implications. So, let's pause here and consider that if you accept endurantism and perdurantism you are seeing the world in profoundly different ways. One view conceptualizes the object of their perception as whole of the object (assuming they are perceiving the whole of the object in the present). The other conceptualizes the object of their perception as partial. They take the object to consist of parts that they cannot see due to the confinement of their conscious experience to a given time. Here is another practical (humanistic) consequence of endurantism vs. perdurantism. Let's say something bad happened to me in 2019. Let's call this "the tragic event". Now let's consider two different viewpoints. For the **endurantism**, that tragic event is not a part of who you are. The event is "a part of you"

(figurative) only to the extent as you feel its effects. For the **perdurantist**, that tragic event is a part of who you are. The you of 2019 still exists and it is just as much of a part (literal) as your hand or leg. That you of 2019 is still suffering. If you are a perdurantist, you might feel for someone who has suffered even if the present part of them isn't suffering. Presumably, this conceptualization impacts the **compassion** we would show another person. For example, if something bad happened to my child, it would make sense for me to hug her (even if she didn't remember the event). If you are an endurantist, then this type of sympathetic response seems strange since the trauma is no longer part of her. Relatedly, if you are a perdurantist, **to love** the whole of a person is to love not merely who they are right now but the entire 4D object (the good temporal parts and the bad temporal parts). In contrast, for the endurantist, to love the whole person is just to love who they are right now.

7. In other words, according to the temporal-part theorist, objects that exist at multiple times do not exist by enduring through time but perduring through time. They don't move through time but are spread out through time. That is, just as you have spatial parts, but you also have temporal parts (the part of you in 2021, the part of you in 2022, and the part of you in 2023). So, when we talk about objects over time, we are talking about a complex entity made up of spatial and temporal parts.

8. The death of S at t1 is bad for S at t2 when the life of S at t2 would have been better for S than no life at all.

9. This applies to a broad variety of harms. With respect to physical harm, since I can't causally interact with an object that does not exist in a way that would do them physical damage (e.g., drawing their blood), I cannot physically harm them. Similarly, I cannot take away the rights of a non-existent entity since only existent beings have rights. I cannot steal from them since they have no possessions.

10. Note that the eternalist is not saying that *there are times after your death where you exist*.

BIBLIOGRAPHY

- [1] William P. Alston, Thomas W. Smythe, and The Society of Christian Philosophers. "Swinburne's Argument for Dualism:" in: *Faith and Philosophy* 11.1 (1994), pp. 127–133. ISSN: 0739-7046. DOI: [10.5840/faithphil199411125](https://doi.org/10.5840/faithphil199411125). (Visited on 10/15/2024).
- [2] Thomas Aquinas. *Summa Theologica*. Aquinas Institute, 2020.
- [3] Inio Asano. *Goodnight Punpun*. Viz signature edition. Vol. 5. San Francisco, CA: Viz Media, 2016. ISBN: 978-1-4215-8620-5 978-1-4215-8621-2 978-1-4215-8622-9 978-1-4215-8623-6 978-1-4215-8624-3 978-1-4215-8625-0 978-1-4215-8626-7.
- [4] Scott Atran and Ara Norenzayan. "Religion's Evolutionary Landscape: Counterintuition, Commitment, Compassion, Communion". In: *Behavioral and Brain Sciences* 27.6 (Dec. 2004), pp. 713–730. ISSN: 0140-525X, 1469-1825. DOI: [10.1017/S0140525X04000172](https://doi.org/10.1017/S0140525X04000172). (Visited on 07/13/2022).
- [5] Alex Baia. "Presentism and the Grounding of Truth". In: *Philosophical Studies* 159.3 (July 2012), pp. 341–356. ISSN: 0031-8116, 1573-0883. DOI: [10.1007/s11098-011-9711-8](https://doi.org/10.1007/s11098-011-9711-8). (Visited on 07/07/2019).
- [6] Thomas Baldwin. "There Might Be Nothing". In: *Analysis* 56.4 (1996), pp. 231–238.
- [7] Luca Banfi and Daniel Deasy. "Presentism and Times as Propositions". In: *Philosophical Studies* 179.3 (Mar. 2022), pp. 725–743. ISSN: 0031-8116, 1573-0883. DOI: [10.1007/s11098-021-01684-3](https://doi.org/10.1007/s11098-021-01684-3). (Visited on 10/13/2022).
- [8] Sam Baron. "How to Endure Presentism". In: *Inquiry* 62.6 (July 2019), pp. 659–673. ISSN: 0020-174X, 1502-3923. DOI: [10.1080/0020174X.2018.1487883](https://doi.org/10.1080/0020174X.2018.1487883). (Visited on 10/13/2022).
- [9] Herbert Benson et al. "Study of the Therapeutic Effects of Intercessory Prayer (STEP) in Cardiac Bypass Patients: A Multicenter Randomized Trial of Uncertainty and Certainty of Receiving Intercessory Prayer". In: *American Heart Journal* 151.4 (Apr. 2006), pp. 934–942. ISSN: 0002-8703. DOI: [10.1016/j.ahj.2005.05.028](https://doi.org/10.1016/j.ahj.2005.05.028). (Visited on 07/11/2025).
- [10] John Bigelow. "Presentism and Properties". In: *Philosophical Perspectives* 10 (1996), pp. 35–52.
- [11] D Bourget and D. J. Chalmers. "Philosophers on Philosophy: The 2020 PhilPapers Survey". In: *Philosophers' Imprint* 23.11 (2023).

- [12] C. D. Broad. *Scientific Thought*. New York: Harcourt, Brace & Company, 1923.
- [13] Michael B Burke. "Liberated Presentism". In: *Review of Metaphysics* 73 (2020), p. 36.
- [14] Michael B. Burke. "Cohabitation, Stuff and Intermittent Existence". In: *Mind* LXXXIX.355 (1980), pp. 391–405. ISSN: 0026-4423, 1460-2113. DOI: [10.1093/mind/LXXXIX.355.391](https://doi.org/10.1093/mind/LXXXIX.355.391). (Visited on 10/17/2024).
- [15] Michael B. Burke. "Tibbles the Cat: A Modern Sophisma". In: *Philosophical Studies* 84.1 (Oct. 1996), pp. 63–74. ISSN: 0031-8116, 1573-0883. DOI: [10.1007/BF00364673](https://doi.org/10.1007/BF00364673). (Visited on 06/13/2022).
- [16] Randolph Byrd. "Positive Therapeutic Effects of Intercessory Prayer in a Coronary Care Unit Population". In: *Southern Medical Journal* 81.7 (1988), pp. 826–829.
- [17] Ross Cameron. "Infinite Regress Arguments". In: *The Stanford Encyclopedia of Philosophy* (2018). Ed. by Edward Zalta.
- [18] Ross Cameron. *The Moving Spotlight*. Oxford: Oxford University Press, 2015.
- [19] Ross P. Cameron. "Much Ado About Nothing: A Study of Metaphysical Nihilism". In: *Erkenntnis* 64.2 (Mar. 2006), pp. 193–222. ISSN: 0165-0106, 1572-8420. DOI: [10.1007/s10670-005-3637-5](https://doi.org/10.1007/s10670-005-3637-5). (Visited on 12/09/2025).
- [20] Joseph Keim Campbell. "Free Will and the Necessity of the Past". In: *Analysis* 67.294 (Apr. 2007), pp. 105–111. ISSN: 0003-2638, 1467-8284. DOI: [10.1111/j.1467-8284.2007.00658.x](https://doi.org/10.1111/j.1467-8284.2007.00658.x). (Visited on 11/15/2024).
- [21] Ben Caplan and David Sanson. "Presentism and Truthmaking: Presentism and Truthmaking". In: *Philosophy Compass* 6.3 (Mar. 2011), pp. 196–208. ISSN: 17479991. DOI: [10.1111/j.1747-9991.2010.00380.x](https://doi.org/10.1111/j.1747-9991.2010.00380.x). (Visited on 07/07/2019).
- [22] Ferrel Christensen. "McTaggart's Paradox and the Nature of Time". In: *The Philosophical Quarterly* 24.97 (Oct. 1974), p. 289. ISSN: 00318094. DOI: [10.2307/2217823](https://doi.org/10.2307/2217823). (Visited on 10/11/2022).
- [23] Arthur C. Clarke. *Profiles of the Future: An Inquiry into the Limits of the Possible*. Harper & Row, 1962.
- [24] Cynthia B. Cohen et al. "Prayer as Therapy: A Challenge to Both Religious Belief and Professional Ethics". In: *The Hastings Center Report* 30.3 (May 2000), p. 40. ISSN: 0093-0334. DOI: [10.2307/3528046](https://doi.org/10.2307/3528046). JSTOR: [3528046](https://www.jstor.org/stable/3528046). (Visited on 07/11/2025).

- [25] Alan Cooperman, Gregory A Smith, and Stefan S Cornibert. "U.S. Public Becoming Less Religious". In: *Pew Research Center* (2015).
- [26] Alan Cooperman et al. "Religion in Everyday Life". In: *Pew Research Center* (2016).
- [27] Damiano Costa. "Persistence in Time". In: *Internet Encyclopedia of Philosophy* (2024).
- [28] W. Lane Craig. "McTaggart's Paradox and Temporal Solipsism". In: *Australasian Journal of Philosophy* 79.1 (Mar. 2001), pp. 32–44. ISSN: 0004-8402, 1471-6828. DOI: [10.1080/713659176](https://doi.org/10.1080/713659176). (Visited on 10/11/2022).
- [29] William Lane Craig. "McTaggart's Paradox and the Problem of Temporary Intrinsics". In: *Analysis* 58.2 (1998), pp. 122–127.
- [30] Thomas M. Crisp and Donald P. Smith. "'Wholly Present?' Defined". In: *Philosophy and Phenomenological Research* 71.2 (Sept. 2005), pp. 318–344. ISSN: 0031-8205, 1933-1592. DOI: [10.1111/j.1933-1592.2005.tb00451.x](https://doi.org/10.1111/j.1933-1592.2005.tb00451.x). (Visited on 09/24/2024).
- [31] Richard R. La Croix. "The Hidden Assumption in the Paradox of Omnipotence". In: *Philosophy and Phenomenological Research* 38.1 (Sept. 1977), p. 125. ISSN: 00318205. DOI: [10.2307/2106521](https://doi.org/10.2307/2106521). JSTOR: [2106521](https://www.jstor.org/stable/2106521). (Visited on 06/12/2026).
- [32] Barry Dainton. *Time and Space*. 2nd. Montreal & Kingston: McGill-Queen's University Press, 2010.
- [33] Rafael de de Clercq. "Presentism and the Problem of Cross-Time Relations". In: *Philosophy and Phenomenological Research* 72.2 (2006), pp. 386–402.
- [34] Daniel Deasy. "Philosophical Arguments Against the A-Theory". In: *Pacific Philosophical Quarterly* 99.2 (June 2018), pp. 270–292. ISSN: 02790750. DOI: [10.1111/papq.12151](https://doi.org/10.1111/papq.12151). (Visited on 12/29/2019).
- [35] Daniel Deasy. "The Moving Spotlight Theory". In: *Philosophical Studies* 172.8 (Aug. 2015), pp. 2073–2089. ISSN: 0031-8116, 1573-0883. DOI: [10.1007/s11098-014-0398-5](https://doi.org/10.1007/s11098-014-0398-5). (Visited on 09/15/2019).
- [36] Daniel Deasy. "The Triviality Argument against Presentism". In: *Synthese* (Dec. 2017). ISSN: 0039-7857, 1573-0964. DOI: [10.1007/s11229-017-1601-y](https://doi.org/10.1007/s11229-017-1601-y). (Visited on 07/07/2019).
- [37] Philip K. Dick. "The Skull". In: *If*. Ed. by Paul W. Fairman. Buffalo, NY: Quinn Publishing Company, 1952, pp. 107–123.

- [38] Douglas Ehring. "Fission, Fusion and the Parfit Revolution". In: *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition* 94.3 (1999), pp. 329–332.
- [39] Harlan Ellison. "Soldier From Tomorrow". In: *Fantastic Universe* 8.4 (Oct. 1957), pp. 4–22.
- [40] Nina Emery, Ned Markosian, and Meghan Sullivan. "Time". In: *The Stanford Encyclopedia of Philosophy* (2020). Ed. by Edward N. Zalta.
- [41] Jonathan Evans et al. *Spirituality and Religion: How Does the U.S. Compare With Other Countries?* 2025. DOI: [10.58094/3X69-ET45](https://doi.org/10.58094/3X69-ET45). (Visited on 07/11/2025).
- [42] David John Farmer. *Being in Time: The Nature of Time in Light of McTaggart's Paradox*. Lanham: University Press of America, 1990.
- [43] Ann Fienup-Riordan and Eddy Carmack. "The Ocean Is Always Changing": Nearshore and Farshore Perspectives on Arctic Coastal Seas". In: *Oceanography* 24.3 (Sept. 2011), pp. 266–279. ISSN: 10428275. DOI: [10.5670/oceanog.2011.78](https://doi.org/10.5670/oceanog.2011.78). (Visited on 07/13/2022).
- [44] Owen J. Flanagan. *The Science of the Mind*. 2nd ed., rev. and expanded. Cambridge, Mass: MIT Press, 1991. ISBN: 978-0-262-27279-7 978-0-585-02346-5.
- [45] A. G. N. Flew. "Farewell to the Paradigm-Case Argument": A Comment". In: *Analysis* 18.2 (1957), pp. 34–40.
- [46] Antony Flew. "The Paradigm Case Argument: Abusing and Not Using the PCA". In: *Journal of Philosophy of Education* 16.1 (July 1982), pp. 115–121. ISSN: 0309-8249, 1467-9752. DOI: [10.1111/j.1467-9752.1982.tb00602.x](https://doi.org/10.1111/j.1467-9752.1982.tb00602.x). (Visited on 11/21/2025).
- [47] Harry G Frankfurt. "The Logic of Omnipotence". In: *The Philosophical Review* 73.2 (1964), pp. 262–263.
- [48] Edward Freeman. "On McTaggart's Theory of Time". In: *History of Philosophy Quarterly* 27.4 (2010), pp. 389–401.
- [49] Kayla Goggin. "11th Circuit Upholds Ruling against Florida Christian School Proposing a Public Pregame Prayer". In: *Courthouse News Service* (Sept. 2024).

- [50] Tyler S. Greenway, Sarah A. Schnitker, and Abigail M. Shepherd. "Can Prayer Increase Charitable Giving? Examining the Effects of Intercessory Prayer, Moral Intuitions, and Theological Orientation on Generous Behavior". In: *The International Journal for the Psychology of Religion* 28.1 (Jan. 2018), pp. 3–18. ISSN: 1050-8619, 1532-7582. DOI: [10.1080/10508619.2017.1406790](https://doi.org/10.1080/10508619.2017.1406790). (Visited on 07/10/2025).
- [51] Edward C. Halperin. "Should Academic Medical Centers Conduct Clinical Trials of the Efficacy of Intercessory Prayer?" In: *Academic Medicine* 76.8 (Aug. 2001), pp. 791–797. ISSN: 1040-2446. DOI: [10.1097/00001888-200108000-00008](https://doi.org/10.1097/00001888-200108000-00008). (Visited on 07/11/2025).
- [52] Mehmet Harma et al. "Vaginal Delivery of Dicephalic Parapagus Conjoined Twins: Case Report and Literature Review". In: *The Tohoku Journal of Experimental Medicine* 205.2 (2005), pp. 179–185. ISSN: 0040-8727, 1349-3329. DOI: [10.1620/tjem.205.179](https://doi.org/10.1620/tjem.205.179). (Visited on 10/20/2024).
- [53] Katherine Hawley. "Fission, Fusion and Intrinsic Facts". In: *Philosophy and Phenomenological Research* 71.3 (2005), pp. 602–621.
- [54] Chris Heathwood. "On What Will Be: Reply to Westphal". In: *Erkenntnis* 67.1 (July 2007), pp. 137–142. ISSN: 0165-0106, 1572-8420. DOI: [10.1007/s10670-007-9050-5](https://doi.org/10.1007/s10670-007-9050-5). (Visited on 03/25/2025).
- [55] Mark Hinchliff. "The Puzzle of Change". In: *Philosophical Perspectives* 10 (1996), pp. 119–136.
- [56] Dennis Charles Holt. "Time Travel: The Time Discrepancy Paradox". In: *Philosophical Investigations* 4.4 (1981), pp. 1–16. ISSN: 0190-0536, 1467-9205. DOI: [10.1111/j.1467-9205.1981.tb00244.x](https://doi.org/10.1111/j.1467-9205.1981.tb00244.x). (Visited on 04/24/2025).
- [57] Hud Hudson. "The Resurrection and Hypertime". In: *Paradise Understood: New Philosophical Essays about Heaven*. Ed. by T. Ryan Byerly and Eric J. Silverman. First edition. Oxford: Oxford University Press, 2017, pp. 263–274. ISBN: 978-0-19-183581-0. DOI: [10.1093/acprof:oso/9780198794301.001.0001](https://doi.org/10.1093/acprof:oso/9780198794301.001.0001).
- [58] David Hume. *An Enquiry Concerning Human Understanding*. Ed. by Tom L. Beauchamp. Oxford: Oxford University Press, 1999.
- [59] David Ingram. "The Virtues of Thisness Presentism". In: *Philosophical Studies* 173.11 (Nov. 2016), pp. 2867–2888. ISSN: 0031-8116, 1573-0883. DOI: [10.1007/s11098-016-0641-3](https://doi.org/10.1007/s11098-016-0641-3). (Visited on 10/13/2022).
- [60] David Ingram. *Thisness Presentism: An Essay on Time, Truth, and Ontology*. New York: Routledge, 2019. ISBN: 978-0-367-66526-5.

- [61] Rögnvaldur D. Ingthorsson. *McTaggart's Paradox*. 1 [edition]. Routledge Studies in Contemporary Philosophy 84. New York: Routledge, Taylor & Francis Group, 2016. ISBN: 978-1-138-67724-1.
- [62] Nihel H Jhou. "The Moving Spotlights Way of 'Unfreezing the Spotlight'". In: *Analysis* 81.3 (Dec. 2021), pp. 439–447. ISSN: 0003-2638, 1467-8284. DOI: [10.1093/analys/anaa082](https://doi.org/10.1093/analys/anaa082). (Visited on 10/11/2022).
- [63] Dominic D.P. Johnson. "The Error of God: Error Management Theory, Religion, and the Evolution of Cooperation". In: *Games, Groups, and the Global Good*. Ed. by Simon A. Levin. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, pp. 169–180. ISBN: 978-3-540-85435-7 978-3-540-85436-4. DOI: [10.1007/978-3-540-85436-4_10](https://doi.org/10.1007/978-3-540-85436-4_10). (Visited on 07/13/2022).
- [64] Robert Kane. *A Contemporary Introduction to Free Will*. Fundamentals of Philosophy Series. New York, N.Y.: Oxford Univ. Press, 2005. ISBN: 978-0-19-514970-8 978-0-19-514969-2.
- [65] Robert Kane. "Incompatibilism". In: *Contemporary Debates in Metaphysics*. Ed. by Theodore Sider, John Hawthorne, and Dean Zimmerman. Malden, MA: Blackwell, 2008, pp. 285–302.
- [66] Amy Kind. *Persons and Personal Identity*. Cambridge, UK ; Malden, MA: Polity, 2015. ISBN: 978-0-7456-5431-7 978-0-7456-5432-4.
- [67] Nathaniel M. Lambert et al. "Invocations and Intoxication: Does Prayer Decrease Alcohol Consumption?" In: *Psychology of Addictive Behaviors* 24.2 (June 2010), pp. 209–219. ISSN: 1939-1501, 0893-164X. DOI: [10.1037/a0018746](https://doi.org/10.1037/a0018746). (Visited on 07/10/2025).
- [68] Henry Laycock. "Some Questions of Ontology". In: *The Philosophical Review* 81.1 (Jan. 1972), pp. 3–42. ISSN: 00318108. DOI: [10.2307/2184224](https://doi.org/10.2307/2184224). JSTOR: [2184224](https://www.jstor.org/stable/2184224). (Visited on 08/11/2026).
- [69] Lee. *Anti-Scepticism*. London: R. Clavel and C. Harper, 1702.
- [70] David Lewis. *On the Plurality of Worlds*. Oxford: Basil Blackwell, 1986.
- [71] David Lewis. "The Paradoxes of Time Travel". In: *American Philosophical Quarterly* 13.2 (1976), pp. 145–152.
- [72] John Locke. *An Essay Concerning Human Understanding*. Ed. by Peter H. Nidditch. Oxford: Clarendon Press, [1689] 1975.
- [73] Arthur Oncken Lovejoy. *The Great Chain of Being: A Study of the History of an Idea*. The William James Lectures 1933. Cambridge (Mass.) London: Harvard university press, 1964. ISBN: 978-0-674-36153-9.

- [74] E. J. Lowe. "On the Identity of Artifacts". In: *The Journal of Philosophy* 80.4 (Apr. 1983), pp. 220–232. ISSN: 0022362X. DOI: [10.2307/2026005](https://doi.org/10.2307/2026005). (Visited on 08/11/2026).
- [75] J. L. Mackie. "Evil and Omnipotence". In: *Mind* 64.254 (1955), pp. 200–212. ISSN: 0026-4423, 1460-2113. DOI: [10.1093/mind/LXIV.254.200](https://doi.org/10.1093/mind/LXIV.254.200). (Visited on 06/11/2026).
- [76] George I. Mavrodes. "Some Puzzles Concerning Omnipotence". In: *The Philosophical Review* 72.2 (1963), pp. 221–223. ISSN: 00318108. DOI: [10.2307/2183106](https://doi.org/10.2307/2183106). JSTOR: [2183106](https://www.jstor.org/stable/2183106). (Visited on 06/12/2026).
- [77] Kris McDaniel. "John M. E. McTaggart". In: *The Stanford Encyclopedia of Philosophy* (2020). Ed. by Edward N. Zalta.
- [78] Jeff McMahan. *The Ethics of Killing: Problems at the Margins of Life*. Oxford Ethics Series. New York: Oxford University Press, 2002. ISBN: 978-0-19-507998-2.
- [79] J. M. E. McTaggart. *The Nature of Existence*. Cambridge: Cambridge University Press, 1921.
- [80] J. M. E. McTaggart. "The Unreality of Time". In: *Mind* 17 (1908), pp. 457–473.
- [81] D. H. Mellor. *Real Time II*. London and New York: Routledge, 1998.
- [82] Kristie Miller. "Presentism, Eternalism, and the Growing Block". In: *A Companion to the Philosophy of Time*. Ed. by Heather Dyke and Adrian Bardon. John Wiley & Sons, Inc., 2013, pp. 345–364.
- [83] Kristie Miller. "The Cresting Wave: A New Moving Spotlight Theory". In: *Canadian Journal of Philosophy* 49.1 (Jan. 2019), pp. 94–122. ISSN: 0045-5091, 1911-0820. DOI: [10.1080/00455091.2018.1519770](https://doi.org/10.1080/00455091.2018.1519770). (Visited on 03/27/2019).
- [84] Kristie Miller. "The New Growing Block Theory vs Presentism". In: *Inquiry* 61.3 (Apr. 2018), pp. 223–251. ISSN: 0020-174X, 1502-3923. DOI: [10.1080/0020174X.2017.1385528](https://doi.org/10.1080/0020174X.2017.1385528). (Visited on 07/14/2019).
- [85] David Moher, Kenneth F Schulz, and Douglas G Altman. "The CONSORT Statement: Revised Recommendations for Improving the Quality of Reports of Parallel-Group Randomised Trials". In: *The Lancet* 357 (2001), pp. 1191–1194.
- [86] Justin Mooney. "The Possibility of Resurrection by Reassembly". In: *International Journal for Philosophy of Religion* 84.3 (Dec. 2018), pp. 273–288. ISSN: 0020-7047, 1572-8684. DOI: [10.1007/s11153-018-9669-y](https://doi.org/10.1007/s11153-018-9669-y). (Visited on 08/12/2026).

- [87] Isaac Newton. *Quæstiones Quædam Philosophiæ*. Cambridge, UK: Cambridge University Library, 2003.
- [88] Harold W. Noonan. "Presentism and Eternalism". In: *Erkenntnis* 78.1 (Feb. 2013), pp. 219–227. ISSN: 0165-0106, 1572-8420. DOI: [10.1007/s10670-011-9303-1](https://doi.org/10.1007/s10670-011-9303-1). (Visited on 07/07/2019).
- [89] L. Nathan Oaklander. "McTaggart's Paradox and Crisp's Presentism". In: *Philosophia* 38.2 (June 2010), pp. 229–241. ISSN: 0048-3893, 1574-9274. DOI: [10.1007/s11406-009-9222-4](https://doi.org/10.1007/s11406-009-9222-4). (Visited on 10/11/2022).
- [90] Eric T. Olson. *The Human Animal: Personal Identity Without Psychology*. Philosophy of Mind Ser. New York: Oxford University Press, Incorporated, 1999. ISBN: 978-0-19-513423-0 978-0-19-802647-1.
- [91] Eric T. Olson. "Was I Ever a Fetus?" In: *Philosophy and Phenomenological Research* 57.1 (1997), pp. 95–110.
- [92] Graham Oppy, Joshua Rasmussen, and Joseph Schmid. "Ontological Arguments". In: *The Stanford Encyclopedia of Philosophy* (2025). Ed. by Edward N. Zalta and Uri Nodelman.
- [93] Michael Oren. "Are Politicians Who Talk About God Crazy?" In: *Tablet* (2021).
- [94] Derek Parfit. "Personal Identity". In: *The Philosophical Review* 80.1 (1971), pp. 3–27.
- [95] Derek Parfit. *Reasons and Persons*. Oxford: Clarendon Press, 1984. ISBN: 978-0-19-824615-2.
- [96] Derek Parfit. "We Are Not Human Beings". In: *Philosophy* 87.1 (Jan. 2012), pp. 5–28. ISSN: 0031-8191, 1469-817X. DOI: [10.1017/S0031819111000520](https://doi.org/10.1017/S0031819111000520). (Visited on 06/06/2024).
- [97] John J. Park. "The Kalām Cosmological Argument, the Big Bang, and Atheism". In: *Acta Analytica* 31.3 (Sept. 2016), pp. 323–335. ISSN: 0353-5150, 1874-6349. DOI: [10.1007/s12136-015-0273-9](https://doi.org/10.1007/s12136-015-0273-9). (Visited on 07/15/2025).
- [98] Blaise Pascal. *Pascal's Pensées*. New York: E. P. DUTTON & CO., 1958. ISBN: 0-525-47018-2.
- [99] Hervey C. Peoples, Pavel Duda, and Frank W. Marlowe. "Hunter-Gatherers and the Origins of Religion". In: *Human Nature* 27.3 (Sept. 2016), pp. 261–282. ISSN: 1045-6767, 1936-4776. DOI: [10.1007/s12110-016-9260-0](https://doi.org/10.1007/s12110-016-9260-0). (Visited on 07/13/2022).

- [100] Robert E. Pezet. "A Foundation for Presentism". In: *Synthese* 194.5 (May 2017), pp. 1809–1837. ISSN: 0039-7857, 1573-0964. DOI: [10.1007/s11229-016-1024-1](https://doi.org/10.1007/s11229-016-1024-1). (Visited on 10/13/2022).
- [101] Nelson Pike. "Divine Omniscience and Voluntary Action". In: *The Philosophical Review* 74.1 (Jan. 1965), p. 27. ISSN: 00318108. DOI: [10.2307/2183529](https://doi.org/10.2307/2183529). JSTOR: [2183529](https://www.jstor.org/stable/2183529). (Visited on 11/05/2024).
- [102] Sean Enda Power. *Philosophy of Time: A Contemporary Introduction*. Routledge Contemporary Introductions to Philosophy. New York (N.Y.): Routledge, 2021. ISBN: 978-1-138-24049-0.
- [103] Marjorie S Price. "Identity Through Time". In: *The Journal of Philosophy* 74.4 (1977), pp. 201–217.
- [104] Hilary Putnam. "Time and Physical Geometry". In: *The Journal of Philosophy* 64.8 (1967), pp. 240–247.
- [105] Anthony Quinton. "The Soul". In: *The Journal of Philosophy* 59.15 (1962), pp. 393–409.
- [106] Jasper Reid. "The Common Consent Argument from Herbert to Hume". In: *Journal of the History of Philosophy* 53.3 (2015), pp. 401–433. ISSN: 1538-4586. DOI: [10.1353/hph.2015.0055](https://doi.org/10.1353/hph.2015.0055). (Visited on 07/12/2022).
- [107] Thomas Reid. *Essays on the Powers of the Human Mind*. Edinburgh: Bell & Bradfute, 1803.
- [108] G. Rodriguez-Pereyra. "There Might Be Nothing: The Subtraction Argument Improved". In: *Analysis* 57.3 (July 1997), pp. 159–166. ISSN: 0003-2638, 1467-8284. DOI: [10.1093/analysis/57.3.159](https://doi.org/10.1093/analysis/57.3.159). (Visited on 12/09/2025).
- [109] Chip Rotolo. "Support for Christian Prayer in U.S. Public Schools Varies Widely by State". In: *Pew Research Center* (June 2025).
- [110] Bertrand Russell. "Is There a God?" In: *The Collected Papers of Bertrand Russell*. Ed. by John G. Slater. Vol. 11. London and New York: Routledge, 1997, pp. 542–548. ISBN: 978-0-415-09409-2.
- [111] Lydia Saad. "Record Few Americans Believe Bible Is Literal Word of God". In: *Gallup* (2017).
- [112] Andrea Sauchelli. "Personal Identity and Trivial Survival". In: *Theoria* 85.5 (Oct. 2019), pp. 402–411. ISSN: 0040-5825, 1755-2567. DOI: [10.1111/theo.12205](https://doi.org/10.1111/theo.12205). (Visited on 06/05/2024).
- [113] Ernst Friedrich Schumacher. *A Guide for the Perplexed*. 1. ed. New York, NY: Harper & Row, 1977. ISBN: 978-0-06-013859-2.

- [114] Betty Marshall Scott, ed. *The History Of Torture*. Hoboken: Taylor and Francis, 2013. ISBN: 978-1-136-19160-2.
- [115] David Sedley. "The Stoic Criterion of Identity". In: *Phronesis* 27.27 (1982), pp. 255–275.
- [116] Sydney Shoemaker and Richard Swinburne. *Personal Identity*. Reprinted. Great Debates in Philosophy. Oxford: Blackwell, 1989. ISBN: 978-0-631-13208-0 978-0-631-13432-9.
- [117] Theodore Sider. *Four Dimensionalism: An Ontology of Persistence and Time*. Oxford: Oxford University Press, 2001.
- [118] Theodore Sider. *Writing the Book of the World*. Oxford : New York: Clarendon Press ; Oxford University Press, 2011. ISBN: 978-0-19-969790-8.
- [119] Pawan Sinha et al. "Mechanisms Underlying Simultaneous Brightness Contrast: Early and Innate". In: *Vision Research* 173 (Aug. 2020), pp. 41–49. ISSN: 00426989. DOI: [10.1016/j.visres.2020.04.012](https://doi.org/10.1016/j.visres.2020.04.012). (Visited on 10/20/2022).
- [120] Bradford Skow. *Objective Becoming*. Oxford: Oxford University Press, 2015. ISBN: 978-0-19-871327-2 978-0-19-877669-7.
- [121] Bradford Skow. "On the Meaning of the Question ``How Fast Does Time Pass?'" In: *Philosophical Studies* 155.3 (Sept. 2011), pp. 325–344. ISSN: 0031-8116, 1573-0883. DOI: [10.1007/s11098-010-9575-3](https://doi.org/10.1007/s11098-010-9575-3). (Visited on 10/11/2022).
- [122] Bradford Skow. "Relativity and the Moving Spotlight:" in: *Journal of Philosophy* 106.12 (2009), pp. 666–678. ISSN: 0022-362X. DOI: [10.5840/jphil20091061224](https://doi.org/10.5840/jphil20091061224). (Visited on 10/11/2022).
- [123] Bradford Skow. "Some Questions about The Moving Spotlight". In: *Analysis* 77.4 (Oct. 2017), pp. 800–810. ISSN: 0003-2638, 1467-8284. DOI: [10.1093/analysis/anx100](https://doi.org/10.1093/analysis/anx100). (Visited on 10/11/2022).
- [124] Brian Smart. "How to Reidentify the Ship of Theseus". In: *Analysis* 32.5 (1972), pp. 145–148.
- [125] Brian Smart. "The Ship of Theseus, the Parthenon and Disassembled Objects". In: *Analysis* 34.1 (1973), pp. 24–27.
- [126] Nicholas J.J. Smith. "The Problems of Backward Time Travel". In: *Endeavour* 22.4 (Jan. 1998), pp. 156–158. ISSN: 01609327. DOI: [10.1016/S0160-9327\(98\)01154-5](https://doi.org/10.1016/S0160-9327(98)01154-5). (Visited on 04/24/2025).

- [127] Joshua Spencer. "Book Review: Could There Have Been Nothing? Against Metaphysical Nihilism." In: *Mind* 125.500 (2016), pp. 1255–1259.
- [128] Giuseppe Spolaore and Giuliano Torrenco. "The Moving Spotlight(s)". In: *Inquiry* 64.7 (Aug. 2021), pp. 754–771. ISSN: 0020-174X, 1502-3923. DOI: [10.1080/0020174X.2019.1610046](https://doi.org/10.1080/0020174X.2019.1610046). (Visited on 10/11/2022).
- [129] Joan Stiles and Terry L. Jernigan. "The Basics of Brain Development". In: *Neuropsychology Review* 20.4 (Dec. 2010), pp. 327–348. ISSN: 1040-7308, 1573-6660. DOI: [10.1007/s11065-010-9148-4](https://doi.org/10.1007/s11065-010-9148-4). (Visited on 06/19/2024).
- [130] J. Tallant. "Ontological Cheats Might Just Prosper". In: *Analysis* 69.3 (July 2009), pp. 422–430. ISSN: 0003-2638, 1467-8284. DOI: [10.1093/analysis/anp077](https://doi.org/10.1093/analysis/anp077). (Visited on 03/25/2025).
- [131] J. Tallant. "Still Cheating, Still Prospering". In: *Analysis* 70.3 (July 2010), pp. 502–506. ISSN: 0003-2638, 1467-8284. DOI: [10.1093/analysis/anq025](https://doi.org/10.1093/analysis/anq025). (Visited on 03/25/2025).
- [132] Richard Taylor. "Fatalism". In: *The Philosophical Review* 71.1 (1962), pp. 56–66.
- [133] Judith Jarvis Thomson. "Parthood and Identity Across Time". In: *The Journal of Philosophy* 80.4 (1983), pp. 201–220.
- [134] Donald Trump. "National Day of Prayer, 2025". In: *whitehouse.gov* (May 2025).
- [135] Jason Turner. "Why Special Relativity Is a Problem for the A-Theory". In: *The Philosophical Quarterly* 70.279 (Apr. 2020), pp. 385–406. ISSN: 0031-8094, 1467-9213. DOI: [10.1093/pq/pqz051](https://doi.org/10.1093/pq/pqz051). (Visited on 10/11/2022).
- [136] Amos Tversky and Daniel Kahneman. "Availability: A Heuristic for Judging Frequency and Probability". In: *Cognitive Psychology* 5 (1973), pp. 207–232. DOI: [10.1016/0010-0285\(73\)90033-9](https://doi.org/10.1016/0010-0285(73)90033-9).
- [137] Amos Tversky and Daniel Kahneman. "Extensional Versus Intuitive Reasoning: The Conjunction Fallacy in Probability Judgment". In: *Psychological Review* 90.4 (1983), pp. 293–315. DOI: [10.1037/0033-295X.90.4.293](https://doi.org/10.1037/0033-295X.90.4.293).
- [138] Edward Tylor. *Primitive Culture: Researches into the Development of Mythology, Philosophy, Religion, Art, and Custom*. Vol. 1. London: John Murray, 1871.
- [139] Peter Van Inwagen. *An Essay on Free Will*. Oxford: Clarendon press, 1983. ISBN: 978-0-19-824624-4.

- [140] Peter Van Inwagen. "Russell's China Teapot". In: *Right to Believe: Perspectives in Religious Epistemology*. Ed. by Patrick Grim and Dariusz Łukasiewicz. Berlin/Boston: De Gruyter, Inc, 2013, pp. 11–26. ISBN: 978-3-11-031997-2 978-3-11-032016-9.
- [141] Peter van Inwagen. *Material Beings*. Cornell Paperbacks. Ithaca and London: Cornell University Press, 1990. ISBN: 978-0-8014-8306-6 978-0-8014-1969-0.
- [142] Peter van Inwagen. *Metaphysics*. 2nd. Boulder, CO: Westview Press, 2002.
- [143] Peter van Inwagen. "The Possibility of Resurrection". In: *The International Journal for the Philosophy of Religion* 9.2 (1978), pp. 114–121.
- [144] Peter van Inwagen. "When Is the Will Free?" In: *Philosophical Perspectives* 3 (1989), pp. 399–422.
- [145] Peter van Inwagen. "Why Is There Anything at All?" In: *Proceedings of the Aristotelian Society, Supplementary Volumes* 70 (1996), pp. 95–110.
- [146] Kadri Vihvelin. "Compatibilism, Incompatibilism, and Impossibilism". In: *Contemporary Debates in Metaphysics*. Ed. by Theodore Sider, John Hawthorne, and Dean W. Zimmerman. Contemporary Debates in Philosophy 10. Malden, MA: Blackwell Pub, 2008, pp. 303–318. ISBN: 978-1-4051-1228-4 978-1-4051-1229-1.
- [147] David Walker. "A Lesson on Gambling with Pascal". In: *Teaching Philosophy* 5.4 (1982), pp. 311–312.
- [148] Ryan Wasserman. *Paradoxes of Time Travel*. Vol. 1. Oxford University Press, Jan. 2018. ISBN: 978-0-19-879333-5. DOI: [10.1093/oso/9780198793335.001.0001](https://doi.org/10.1093/oso/9780198793335.001.0001). (Visited on 04/23/2025).
- [149] J.W.N. Watkins. "Farewell to the Paradigm-Case Argument". In: *Analysis* 18.2 (1957), pp. 25–33.
- [150] Elie Wiesel and Elie Wiesel. *Night*. Ed. by Marion Wiesel. 1. ed. of this translation. New York: Hill and Wang, 2006. ISBN: 978-0-374-50001-6 978-0-8090-7355-9.
- [151] B. A. O. Williams. "Personal Identity and Individuation". In: *Proceedings of the Aristotelian Society* 57.1 (1957), pp. 229–252. ISSN: 0066-7374, 1467-9264. DOI: [10.1093/aristotelian/57.1.229](https://doi.org/10.1093/aristotelian/57.1.229). (Visited on 10/20/2024).
- [152] Bernard Williams. "The Self and the Future". In: *The Philosophical Review* 79.2 (1970), pp. 161–180.

- [153] T L Wilson. "The Search for Extraterrestrial Intelligence". In: *Nature* 409 (2001), pp. 1110–1114.
- [154] J. Jeremy Wisniewski. *Understanding Torture*. Contemporary Ethical Debates. Edinburgh: Edinburgh university press, 2011. ISBN: 978-0-7486-3538-2.
- [155] Eric T Yang and Stephen T Davis. "Composition and the Will of God: Reconsidering Resurrection by Reassembly". In: *Paradise Understood: New Philosophical Essays about Heaven*. Ed. by T. Ryan Byerly and Eric J. Silverman. Oxford University Press, Mar. 2017, pp. 213–227. ISBN: 978-0-19-879430-1. DOI: [10.1093/acprof:oso/9780198794301.001.0001](https://doi.org/10.1093/acprof:oso/9780198794301.001.0001). (Visited on 08/13/2026).
- [156] Dean W Zimmerman. "The Privileged Present: Defending an "A-theory" of Time". In: *Contemporary Debates in Metaphysics*. Ed. by Theodore Sider, John Hawthorne, and Dean W. Zimmerman. Contemporary Debates in Philosophy 10. Malden, MA: Blackwell Pub, 2008, pp. 211–225. ISBN: 978-1-4051-1228-4 978-1-4051-1229-1.