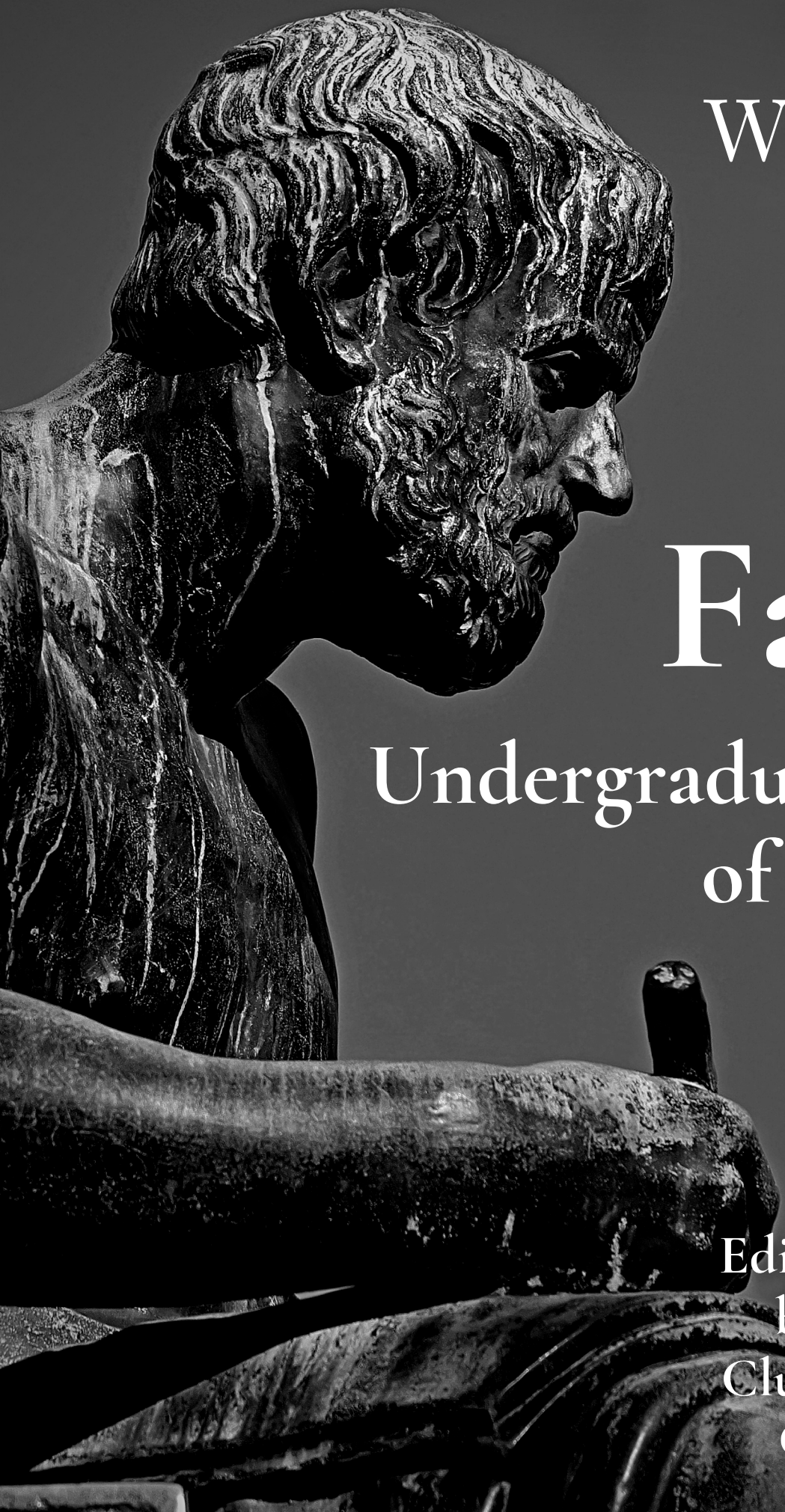

Issue 4
Winter 2021

Falsafa

Undergraduate Journal of Philosophy

Edited and published by The Philosophy Club at University of California, Irvine.

Falsafa

Undergraduate Journal of Philosophy

ISSUE 4
WINTER 2021

The Falsafa Team
Chief Editor: Maggie Lajoie
Deputy Chief Editor: Bradley Holder
Managing Editors: James McDaniel, Meredith Sneed
Designer: Chelsey Taniguchi
Publishers: Jacqueline Markham, Xuehong Zhao
Editors: Alexandros Gkikas, Autumn Gahagan, Emma Kerro, Haley Dyson, Jacqueline Markham, Safari Rugui, Xuehong Zhao

The Philosophy Club at UCI
President: Wade Smith
Vice President: Matthew Evans
Treasurer: Bradley Holder
Chief of Operations: Maggie Lajoie
Secretary: Ian Edwards
Head of Operations: Connor Kuhn
Creative Director: Chelsey Taniguchi
Communications Assistant: Eliud Romano
Content Director: Kaomi Booker
Content Manager: Roy Hyglesias Petit-Frere
Director of Communications: Sunny Zhang
Director of Visual Communications: Cynthia Lopez
Librarian: Michael Mocanu
Library Manager: Miles Chen
Marketing Director: Simone Ledbetter
Social Media Director: Chriselle Badua

Contact Information:
Email: falsafauci@gmail.com
Website: www.falsafaorg.com
Club facebook: www.facebook.com/PhilClubUCI/
Club website: www.humanities.uci.edu/philosophy/undergraduate/club.php

Contents

Letter from the Department Chair

It is once again a great pleasure for me to write a message to acknowledge the terrific job done by our undergraduates on the *Falsafa* team. This year, more than ever in our lives, we have been under strain. The pandemic has affected us all in unprecedented ways. It is truly remarkable how, despite these circumstances, the *Falsafa* team has been able to put together a hugely interesting issue.

In it, you will read about traditional problems in philosophy, such as personal identity, or whether knowledge is justified true belief. Moreover, you will read about whether evolutionary debunking arguments with respect to the justification of our moral beliefs can successfully be countered. Finally, you will also find a very timely paper on the institutional moral failure with respect to vaccine development the pandemic has unveiled. Enjoy!

With best wishes,

Annalisa Coliva
Professor and Chair
Department of Philosophy
UCI

Letter from the Chief and Managing Editors

We are proud to present the fourth issue of *Falsafa,* the undergraduate journal of philosophy, edited and produced by undergraduate students from a range of backgrounds, majors, and schools. Throughout the pandemic, our authors and the *Falsafa* team have overcome the limitations of remote work to put together this issue on the ethics of vaccine challenge trials, justified true belief, self-identity, and moral realism. This year we extended the range of philosophical disciplines for consideration, and these topics are representative of our new, broader guidelines. In previous issues, we focused on publishing topics pertaining to philosophies traditionally marginalized in the West. While these topics remain valuable to us, our new focus is on giving our authors the autonomy to write on the topics that interest them the most. We hope this new direction will interest you too.

This issue would not have been possible without our dedicated team, including editors, publishers and a designer, who made their work for the journal a priority even during these difficult times. The new online workspace enabled us to recruit staff from across the country and select papers submitted from around the world. A special thanks goes to our authors for producing excellent work in these changing times—in one case drawing on the pandemic as inspiration for their paper.

Thank you to the hundreds of schools who supported *Falsafa* by sharing our call for papers and to UC Irvine's Department of Philosophy for their support and guidance. On behalf of the editing team here at *Falsafa*, we sincerely appreciate your continued support and interest in our mission to highlight the work of talented, undergraduate philosophers.

***Falsafa*, the Undergraduate Journal of Philosophy at UC Irvine**

Maggie Lajoie, Bradley Holder
Chief and Deputy Chief Editor

James McDaniel, Meredith Sneed
Managing Editors

Knowledge as Justified Belief Wherein No Contradictory Evidence Is Obtainable for the Holder

Katsuki Ishihara, McGill University

Abstract

In this paper, I defend the theory of knowledge as justified true belief, or the JTB account of knowledge, and reshape this theory to make it accord with the problems raised against it. I argue against the Gettier Problem, which became a major objection to Plato's JTB account. I also falsify the objections made against the claim that deductive inference must be based on true premises, namely Richard Feldman's Havit case, and alternatives to the Gettier case, such as Roderick Chisholm's sheep in the field. I consider the possible objection that because the justification condition is too strict, S cannot know anything if S could not have been aware of the false premise. To respond to this objection, I redefine knowledge as justified belief wherein no contradictory evidence is obtainable for the subject.

Introduction

Knowledge as Justified True Belief, or the JTB account of knowledge, was first proposed in *Theaetetus* written by Plato. His theory of knowledge as justified true belief can be formulated as follows:

S knows that p, if and only if

(1) p is true

(2) S believes that p is true

(3) S is justified in believing that p is true.

According to this theory, there is a difference between knowing and merely believing. Believing can be unwarranted, whereas knowing requires the three conditions above. Hence, it is possible that one believes p and does not know it. This theory had long existed as the most cogent definition of knowledge until, in 1963, Edmund Gettier proposed two cases that disprove it. The first case is the Job Case; Smith forms a belief "Jones is the man who will get the job, and Jones has ten coins in

his pocket."[1] He generalizes the belief as, "The man who will get a job has ten coins in his pocket."[2] However, as it turns out, Jones does not get a job, and instead, Smith does. Moreover, Smith unknowingly has ten coins in his pocket. In the second case, or the Barcelona Case, Smith forms a belief, "Jones owns a Ford."[3] From this argument, Smith introduces a disjunction, "Either Jones owns a Ford, or Brown is in Boston"; "Either Jones owns a Ford, or Brown is in Barcelona"; and "Either Jones owns a Ford, or Brown is in Brest-Litovsk."[4] Now, Jones turns out to be lying and does not own a Ford. Furthermore, Brown happens to be in Barcelona. Gettier's suggestion with these two cases is that even though Smith has justified true beliefs about both cases, Smith does not know the beliefs he formed. Therefore, the JTB account of knowledge is flawed. The question I want to raise is whether his arguments really undermine the JTB account of knowledge.

Analyses

Before the analyses, I want to note the use of the implication in knowledge. Unlike logical implication used in formal logic, the inference in the analysis of knowledge necessarily requires a true antecedent. This is because of the justification condition. For example, Smith tells Thomas that he has a dog. Smith shows Thomas a photo of his dog. Thomas forms a belief that Smith has a dog. From this belief, Thomas forms a belief, "One of my friends has a dog." However, Smith turns out to be lying, and the photo is of his sister's dog. Thomas clearly does not know that Smith has a dog nor has a justified true belief. Because as soon as Smith turns out to be lying about the dog, Thomas's belief, "Smith has a dog," becomes a mere unjustified utterance. Smith's other belief, "One of my friends has a dog," is falsified because if it were not for Thomas, Smith would not think in that way in the first place. If Smith wants to keep the belief that "One of my friends has a dog," he must find another friend who has a dog to satisfy the justification condition. Otherwise, his belief becomes no different than randomly saying, "One of my friends has a horse." Therefore, as soon as Thomas learns that Smith was lying, he must abandon his beliefs as knowledge. It follows if the antecedent's truth condition does not hold, the consequence and the justification condition fail. Therefore, in epistemological arguments, if p is false, $p \rightarrow q$ and q are false. If p is false, S must find another p to infer q.

[1] Edmund L. Gettier, "Is Justified True Belief Knowledge?," Analysis 23, no. 6 (June 1, 1963):pp. 121-123, https://doi.org/10.1093/analys/23.6.121, 122.

[2] Gettier, 122.

[3] Gettier, 122.

[4] Gettier, 122-23.

To analyze the Gettier Problem, I employ predicate logic. The symbolization keys are as follows:

Universal Derivation (UD): the male candidates for the job

Jx: x gets a job

Cx: x has ten coins in his pocket

j: Jones

s: Smith

Now, Smith's first belief, "Jones is the man who will get the job, and Jones has ten coins in his pocket," can be symbolized as (Jj ∧ Cj) [a]. Smith's next belief, "The man who will get a job has ten coins in his pocket," is an existential introduction of [a]. Therefore, the sentence must be ∃x(Jx ∧ Cx) [b]. As I have shown at the beginning of this section, I would like to emphasize that [b] cannot be inferred without [a]. Next, Jones ends up not getting a job, but he still has ten coins. This sentence is (¬ Jj ∧ Cj) [c]. At this point, [b] is invalid because [c] falsifies [a], therefore [b] could not have been inferred. Finally, Smith gets a job, and unknown to Smith himself, he has ten coins in his pocket. Because he does not know he has ten coins in his pocket, the sentence is simply Js [d1]. Let us suppose Smith knew that he has ten coins in his pocket. The sentence would be (Js ∧ Cs) [d2]. Only from [d2], can Smith infer [b] by existential introduction. Smith's belief [b] is justified only based on [d2], which is not a false premise. However, because [d2] is not the case in this situation, Smith cannot have [b]. Therefore, Smith does not have justified true belief or knowledge.

The symbolization keys for the second case are as follows:

UD: Smith's colleagues

Fx: x owns a Ford

Bx: x is in Barcelona

j: Jones

b: Brown

Smith's first belief, "Jones owns a Ford," can be expressed as Fj [a]. Smith uses a disjunction introduction to claim, "Either Jones owns a Ford, or Brown is in Barcelona." The sentence for this claim is (Fj ∨ Bb) [b]. Smith only knows [b] and surely does not know Bb. To know Bb, Smith needs another justification for Brown's location. The same can be said for Boston and Brest-Litovsk. Nonetheless, Jones turns out to have been lying about owning a Ford, and thus it follows ¬ Fj [c]. Again, if the antecedent is false, the consequent and the sentence must be false; hence [b] is false. If Smith wants to infer a sentence regarding Brown's location, Smith can do so by deducing another sentence from [c], which will be (¬ Fj ∨ Bb) [d]. Therefore, Smith is not justified in believing [b] and does not know it either.

To sum up my arguments, the Gettier problem violates the rule, "If the antecedent

is false, the consequent and the sentence must be false"; both cases reuse the justification condition for another inference. Based on the analyses, I conclude that the Gettier Problem does not logically undermine the JTB account of knowledge.

Here, I will analyze the argument made against the claim that logical inference must be based on true premises. I will start with Feldman's Havit case. The case is as follows: Smith forms a belief that both of his colleagues Havit and Nogot own a Ford. From this information, Smith infers that someone in the office owns a Ford. However, Havit turns out to have been lying about owning a Ford. Nevertheless, Smith's belief, "Someone in the office has a Ford," is intact. I will interpret this case like the earlier ones. The symbolization keys are as follows:

UD: Smith's colleagues
Fx: x owns a Ford
h: Havit
n: Nogot

Smith's first belief, "Both Havit and Nogot own a Ford," would be (Fh $\wedge$ Fn) [a]. From this, Smith infers, "Someone in the office owns a Ford," or $\exists$xFx [b]. However, [b] cannot be inferred immediately from [a]. If Smith has [a], and if he wants to infer another sentence immediately, he must infer, "There are at least two different persons who own a Ford," or $\exists x \exists y$(Fx $\wedge$ Fy $\wedge$ x$\neq$y). If Smith wants to infer [b] from [a], there must be a step of conjunction elimination, which would result in either Fh [c1] or Fn [c2]. To make an inference [b], Smith has two routes, either [c1] or [c2]. Both routes lead to the same conclusion [b]. Because of these two routes, even if one of Havit and Nogot is lying, Smith can still make the inference to [b] via the valid route and justify making such an inference. Therefore, this objection does not defeat the JTB.

Similarly, Chisholm's sheep in the field case and Russell's stopped clock cannot deny the JTB account. Because these cases have similar structures, I will only analyze Chisholm's sheep case. The sheep case is as follows:

> A person *takes* there to be a sheep in the field and does so under conditions which are such that, when under those conditions a person takes there to be a sheep in the field, then it is evident for that person that there is a sheep in the field. The person, however, has mistaken a dog for a sheep and so what he sees is not a sheep at all. Nevertheless, it happens that there is a sheep in another part of the field. Hence, the proposition that there is a sheep in the field will be one that is both true and evident and it will also be one that the person accepts. But the situation does not warrant our saying

> that the person knows that there is a sheep in the field.[5]

Let me analyze this case. First, S perceives a dog that looks exactly like a sheep in the field. Only based on this perception, S forms a belief that there is a sheep; S does not have any other source that justifies S's belief. Now, S falsely believes that there is a sheep in the field. Nonetheless, there is a real sheep. However, this claim is not related to S's belief at all; whether or not there is a real sheep in the field, S keeps believing that there is a sheep in the field based only on S's perception of the dog that looks like a sheep. I will explain whether S knows that there is a sheep in the field in the next chapter.

Other cases have similar structures as the sheep case:

P1: S believes that p.

P2: S is not aware that it is, in fact, ¬ p.

P3: Unknown to S, p.

∴ S's belief is true and justified, but it does not count as knowledge.

As I explained above, this type of argument is flawed because the premise does not provide the truth condition to S's belief. The unknown truth does not affect S's belief.

Objection and Replies

Some might say that there are cases wherein S could not have been aware of the false premise. Is S justified in believing the information, even if it is false? For example, in Chisholm's sheep in the field case, if S cannot know what S is seeing is a dog and has sufficient evidence to believe that what she is seeing is a dog, does S know there is a sheep? I want to propose that she does until her belief is radically contradicted by other evidence. If S could not have known that S is being deceived, and if S otherwise satisfies the other criteria for knowledge —the justification and the belief conditions— S can say S's belief is knowledge. What I mean by "radically contradict" is that S's justification for her belief is denied. For example, in the sheep case, S's belief, "There is a sheep," radically contradicts the fact that it is a dog. As long as S is not aware that she is deceived, S can say she knows there is a sheep. However, as soon as S learns that her belief had been based on the false premise, S can no longer claim to know that it is a sheep. But until then, she can claim that it is a sheep because S is justified in believing so. The reason I believe S can claim to know p as long as S is not aware of the fact ¬ p is that S cannot be responsible for not knowing ¬ p. If there is no way for S to know ¬ p and if

[5] Roderick M. Chisholm, Theory of Knowledge (Upper Saddle River, NJ: Prentice-Hall, 1966), 93.

there are sufficient reasons to believe p, it seems to be utterly unreasonable to accuse S of not knowing ¬ p. Because S could not have known ¬ p, S did not have a choice. Accordingly, S cannot be held responsible for not knowing ¬ p.

The concept with which I try to define knowledge is time. There are only a few things of which we can be absolutely certain. Those things are a priori, such as tautology. Those things are timeless and self-sufficient; they are always true and do not require external justification. However, most things we encounter are a posteriori. Those things are transitory and require external justification. And the quality of a posteriori knowledge is dependent on the external justification; if strong evidence that negates justification appears, a posteriori knowledge fails. Because a posteriori knowledge is susceptible to change, it would be appropriately defined with the time condition; S knows p as long as no evidence radically contradicts p.

Some might argue that my proposal is not valid in terms of skeptical arguments. Skeptics argue that knowledge is impossible. For example, we cannot know whether we are brains in vats. Then, the right thing to do is to suspend judgment by saying we do not know whether we exist in reality. Because I argue that we cannot always know what is objectively true, skeptics are likely to argue that I must suspend the judgment on p instead of saying there is no reason that I do not believe p is false. In other words, they might argue that my argument is unwarranted. If S cannot know p objectively, even if p seems to be true to S, S must suspend the judgment. Besides, what we seem to know a priori can be wrong if we are manipulated into believing they are true. We might be mistaken about, "A triangle has three sides." As in classic Cartesian doubt, we might be manipulated by the evil genius. Hence, we cannot know anything for certain.

To this objection, I would argue that it is impossible to suspend all the judgment. I would provide two arguments. First, if we suspend all the judgment and argue that we cannot know anything, the concept of knowledge becomes empty. Therefore, we could not have argued that we cannot know anything in the first place. Second, if we cannot know anything, we cannot know that we cannot know anything. Therefore, we could not have argued so.

Nonetheless, the evil genius doubt still refutes those two arguments. Because we are still deceived by the evil genius to believe that syllogism is infallible. At this point, it becomes impossible to argue against Cartesian skepticism. To develop the argument, we need to define the realms of truths. I would argue that there are three

different realms of truths. From the narrowest to the broadest, they are subjective truth, objective truth, and surobjective truth. In the realm of subjective truth, what is true is a posteriori. This is because this truth derives from one's subjective experience. The objective truth, on the contrary, is a priori. This truth is objectively true; hence, it proves itself. For example, the proposition, "A triangle has three sides," is self-sufficiently true. The surobjective truth goes beyond the objective truth. It explores the metaphysical truth; it tries to answer what is true beyond our conventional understanding of the world by casting doubt. Cartesian skepticism is of this sort because it asks whether the objective truth is true in more metaphysical states. What I want to examine is this surobjective truth.

Let us consider the evil genius again. Suppose we are manipulated by this evil genius. We are manipulated into believing what is false, including logic. Hence, we believe logic is true, even though we might be manipulated by the evil genius. Under this supposition, we know that we cannot know the surobjective reality; we might be manipulated, or we might not. However, we cannot know that we know the surobjective reality because we might be manipulated into knowing it. This leads to the infinite regress of knowing and not knowing of the surobjective reality. This infinite regress tells us that if this evil genius exists, we are completely deceived by it. If we are completely deceived, there is no difference between reality and the state of being deceived. If it does not exist, surobjective reality is false. Either way, surobjective reality amounts to objective reality, and what we believe as the objective reality will never be affected by the existence of the evil genius. Hence, we can safely conclude that what is a priori true is objectively true. I am not arguing whatever seems true for S is true. S still needs justification to claim S knows something. In this sense, the knower has immense responsibility for claiming the knowledge.

This account can also be functional in the legal context. Suppose the jurors are voting whether they should convict the defendant or not. The jurors can rule out any possibilities of far-fetched conspiracy theory because if there is a complete conspiracy scheme, there would not be any means to prove the validity of doubt; in the real legal context, any doubts would not arise in the first place if the scheme is perfect. On the contrary, if there is any doubt, they should not cease to seek the truth for the sake of justice.

Let us move back to the contrast between subjective truth and objective truth. I have proposed that if one has convincing enough evidence about p, then p is true as long as p

is not radically contradicted. The skeptical objection told us to suspend the judgment because one cannot know what is objectively true. And even if one has convincing enough a priori justified evidence about p, skeptics argue one might be deceived. Nevertheless, I argued against this worry about being deceived. It follows that one does not have to suspend the judgment if one is a priori justified in believing p. Despite that, it is not always possible to a priori justify one's a posteriori belief. For example, S cannot be justified in believing the dog as a sheep because S is justified in believing so only because S "perceives" the sheep/dog. Perception is a posteriori. However, we do not want to deny the validity of S's perception because if what a posteriori is false, we cannot know most things, as I wrote earlier. We greatly rely on our perceptions and experience. That is why I proposed "as long as there is no evidence that radically contradicts p." If S is justified in believing an a posteriori belief, S can claim to know it insofar as it is not contradicted by other evidence.

I would also like to note that even according to my proposal, the Gettier Case and the Havit Case are invalid because they do not satisfy the justification condition. However, Chisholm's sheep case would be true if S does not consider the existence of the real sheep. Because S could not have known it was a dog, S knows the dog as a sheep until S learns it is a sheep.

Conclusion

I have defended the JTB account of knowledge by denying the Gettier Problem and the Havit case in terms of predicate logic. The Gettier Problem does not undermine the JTB account because Smith violates the basic epistemic rule, "If the antecedent is false, the consequent and the sentence must be false." The Havit Case also fails because the generalization, "Someone in the office owns a Ford," cannot be inferred immediately from the premise. Regarding the objection, I have argued if S could not have been aware of the false premise and if S can satisfy the justification and the belief conditions, S's claim that p is justified knowledge as long as the belief p is not radically contradicted. This is because S cannot and should not be held responsible for what S could not have known. I have concluded, as far as a posteriori knowledge is concerned, S knows p iff:

(1) S believes p to be true;

(2) S is justified in believing that p is true;

(3) no evidence radically contradicts p for the time being.

Bibliography

Chisholm, Roderick M. In *Theory of Knowledge*, 93. Upper Saddle River, NJ: Prentice-Hall, 1966.

Feldman, Richard. *Epistemology*. Upper Saddle River, NJ: Prentice Hall, 2003.

Gettier, Edmund L. "Is Justified True Belief Knowledge?" *Analysis* 23, no. 6 (June 1, 1963): 121–23. https://doi.org/10.1093/analys/23.6.121.

The Subjectivity of Personal Identity

James Rogers, University of Cambridge

Abstract

Personal identity is an important concept in human life. It is very important to people that they have a sense of identity; however, the metaphysics of personal identity is not something that is commonly discussed. In this paper, I examine how the identity of an object or entity across time requires certain persistence conditions. I argue that numerical identity is always a social construct intended to fulfill certain functions and that no matter how we try to conceptually engineer personal identity we will always end up not being able to fulfill all the functions we want and instead must prioritize certain functions based on our own subjective preferences. I then conclude that if we are to base a given theory on these subjective preferences then it is reasonable to state that the theory is subjective.

Introduction

The main argument of this paper is that there is a wide range of functions—these functions allow us to refer to what we want to call a person as a person—that we want the concept of personal identity to fulfill but no account of personal identity can fulfill all of these functions. An example of one of these functions might be explaining how we can refer to ourselves as the same person when we exist at different points in time. As a result, any theory which is chosen will be based on claiming that certain functions are more important than others, which if we assume our preferences are subjective makes these theories subjective. It is my claim that there is no logical reason for choosing certain functions over others, so we will never reach a consensus on personal identity. Key concepts that will be used are Occam's razor and Hitchens' razor as no reason can appear to be used to explain why certain functions of the concept of personal identity are more important than others (other than our own subjective preferences), assuming that no additional assumptions should be added into the theory, such as

this function is more important because I believe it is (Occam's razor) and any reason asserted without evidence can be dismissed. (Hitchens's razor). Similar reasoning is used to assert that ethics are subjective or a social construct or that god does not exist. This isn't to state that ethics has to be subjective or that god cannot exist but that it is reasonable to take the stance that ethics is subjective and god does not exist until we are shown otherwise. I propose that a similar stance is appropriate for personal identity.

Identity and Numerical Identity being of concepts

It is important to note that there are different kinds of identity, especially identity as it relates to one specific object. The most obvious is the identity of indiscernibles (this is sometimes called Leibniz's law or qualitative identity). This states that if X has the exact same properties as Y, then X is identical to Y. This basically states that no two objects or entities can have the exact same properties. This definition clearly encounters problems with referring to an object across time. Consider a dog. A vet needs to be able to refer to a singular dog when treating it and determine that a dog at two different points in time is the same dog. However, the dog will have a different atomic structure as atoms will leave and join its body, so the dog will not satisfy Leibniz's law and not be qualitatively identical. In order to satisfy this kind of identity we need a new kind of identity, numerical identity.

The dog is clearly not the same at one time as it is at the other; however, we would still like to talk about the dog as one thing. Numerical identity can then be defined. Two objects at two different points in time are numerically identical if all the relevant properties or conditions are the same. For example, in the case of the dog the exact atomic structure of the dog is an irrelevant property. An example of a relevant property might be the physical continuity of the dog from the dog since inception. Physical continuity is the relation of two objects of the same kind in time that at no point completely lose physical form and regain it. For example, if there is a dog X at time t, this dog will lose atoms throughout their existence and have them replaced. At a different point in time, time t*, a dog that is physically continuous with dog X will have a different atomic structure but at no point will dog X at time t* have stopped existing and reappeared, meaning it is physically continuous with dog X. An object A can be said to be numerically identical to object B, then, if and only if both objects share the same relevant properties C (these relevant properties can be called persistence conditions) and the status of the object in relation to all other irrelevant properties D can (but does not necessarily have to) vary.

Numerical identity is always conceptual and a social construction. Unlike Leibniz's law and qualitative identity, numerical identity does not refer to the actual state of every property being identical. For example, a tree will be different at every second of its existence (at least at the atomic level). We decide on the persistence conditions of the tree and then from this decision, which is self-evidently a social construct as it is constructed by society, a shared conception of what it means to be one tree across time. It refers to just the persistence conditions being the same. The persistence conditions are selected by society in order to fulfill the functions we want the concept to fulfill. For example, when considering the persistence conditions of the dog we choose something like the physical continuity of the dog and not the condition that the dog has four legs. This is because everyone wants to be able to refer to a dog, which loses a leg, as the same dog before and after it loses its leg. It is important to note that there are objects that can have consistent persistence conditions; this means that all the functions that we want an account of the numerical identity of the object to fulfill can be fulfilled by one account. I will later argue that personal identity is not one of these concepts.

What it is that Makes an Account of Numerical Identity Subjective

Subjectivity can be a loaded term with multiple meanings. When I say that the numerical identity of an object or entity is subjective, such as personal identity, then what is meant is that all the functions we want the concept to fulfill cannot be fulfilled, and, other than people's subjective preferences (about which we disagree), there is no logical way of determining which functions should be prioritized. There are objects that can be given persistence conditions in order to fulfill all the functions we want; examples will be given later. In these cases, as they do not depend on any subjective preferences, as there is unanimous agreement, we will not refer to these as subjective. This paper is concerned with conceptual engineering, the idea that concepts should be changed in order to fulfill the functions we need them to fulfill. An example of conceptual engineering in physics is that with the discovery of general relativity we changed our conception of gravity to better fit with this theory. The question is whether the concept of personal identity (numerical identity for personhood) can either fulfill all the functions we want it to fulfill, or whether some concepts can be said to be more important than others. If the latter is true, then personal identity must be subjective.

It can be difficult to state when an account of numerical identity for a given object or

entity can be given. Numerical identity can be easily seen in objects or entities that are qualitatively identical (they obey Leibniz's law). However, for things that do not obey Leibniz's law, such as a dog or a bridge (which will have different atomic structures at different points in time), it can be difficult to determine the persistence conditions. The main test is: When given the persistence conditions X of object Y, any object Z that also has the persistence conditions X will be said to be Y. Demonstrating things through this test is more problematic. It requires that a certain number of people N must accept that the persistence conditions fulfill the functions desired relative to the total population P that can have desires of what functions the persistence conditions must fulfill. This ratio of N/P is a difficult thing to determine. It is unreasonable to propose it should be one. For example, if everyone has an understanding of what the persistence conditions of a particular bridge are and a person on deliriant drugs claims the numerical identity of dog X has to have a persistence condition that it can fly, so that when they become sober, they say X is a different dog because they no longer see it flying. Nevertheless, while one is unattainable, the closer N/P is to one the more consensus on the persistence conditions there are. The main point that this paper requires is simply that the ratio must be close to one but not necessarily one because the smaller the ratio is the less confidence we can have in the assertion of an object's numerical identity.

If we talk of one tree, we do not refer to one state of a tree with its exact properties. We refer to a set of exact trees (a tree at a specific moment in time with all its precise properties) throughout time, connected by a kind of physical continuity. The physical continuity and the tree being alive are the persistence conditions of the tree C. These persistence conditions are then found through what we want our concept of a tree to do, and we determine what we want our concept of a tree to do through our shared experience of what we want our concept to do. No one or extremely few people would propose that the persistence conditions of a tree are whether it is eight o'clock or not because it does not fit our intuitions. In many cases this will be physical continuity but not all. For example, a statue of a person could be melted down into a sphere of metal. This sphere would have physical continuity with the statue, but it would not be the same statue. This is because the persistence conditions of its being the same statue require the statue to retain a certain macroscopic appearance. The important point here is that consistent persistence conditions that we can all agree on can be provided for many objects and entities.

Trying to Find a Consistent Account of Personal Identity

Personal identity is then the numerical identity of a person. The goal of an account of personal identity is to find the persistence conditions of the person. A standard definition of personal identity is "necessarily, for any x, if x is a person at t and something y exists at t*, x=y if and only if x at t and y at t* stand in relation R, where R is the relation preferred by the particular theory."[1] A main account of personal identity comes from some kind of psychological continuity (this is sometimes called the psychological approach), for example through memory or as Locke put it "sameness of rational being."[2] This "sameness of rational being" is the R in the account of personal identity. The other common view comes from the idea that personal identity comes from some kind of physical continuity, usually stating that we are human animals. This view is commonly referred to as animalism.

Thought experiments about brain swapping in Bernard William's "The Self and the Future" show that human intuition about what is the self is not obviously consistent. Williams proposes that person A and person B have their minds swapped so that person A's mind will be in person B's body and vice versa. Then they will be told one body will be subjected to a pleasant experience (receiving money) and the other will be subjected to an unpleasant experience (torture). He then states that both A and B should choose which body is receiving the pleasant and unpleasant experience; for example, A chooses that B's body receives the money and A's body receives the torture, while B chooses the same, that B's body receives the money and A's body receives the torture. He then concludes that the one who chose the pleasant experience for the other person's body on the assumption that their mind will be transferred into it will be happy about it regardless of which body they are in, while the other person's choice "was unwise."[3] This thought experiment shows that we identify ourselves with our mind when considering what will benefit us, over our physical body. This thought experiment provides a lot of support for the view of psychological continuity, as it seems when we are considering how we should act in order to benefit ourselves in the future, we are concerned with benefiting our mind rather than our body. This obviously presents problems for animalism as we

[1] Radim Belohrad, "Subjective Theory of Personal Identity and Practical Concerns," *Organon F* 22, no. 3 (2015): 283

[2] Jessica Gordon-Ruth, "Locke on Personal Identity," Stanford Encyclopedia of Philosophy, December 16, 2020, https://plato.stanford.edu/entries/locke-personal-identity.

[3] Bernard Williams, "The Self and the Future," The Philosophical Review 70, no. 2 (1970): 166.

don't identify with the human animal when thinking about our future selves.

However, the works of animalists, for example Olson, also show issues with the approach of psychological continuity. Locke defines a person through psychological characteristics: A person is "an intelligent thinking being" that "can consider itself as itself, the same thinking thing at different times and places."[4] Olson points out that the psychological approach relies on the assumption that "every person is essentially a person."[5] However, this isn't necessarily obvious. If we consider person A, who goes into a vegetative state, then person A is no longer a person, at least by Locke's account. We would still refer to them as the same person though. If person A went into a vegetative state, people would still refer to the thing (thing is being used here because if we use Locke's definition of a person, A is not a person because it does not have the sameness of rational being) as A. People would still not talk about A as if he had ceased to exist, so the psychological approach struggles to reconcile these intuitions. The psychological approach then cannot fulfill the function of being able to consistently refer to people in vegetative states as the same person. Olson concludes that being a person isn't essential to identity then. He then further concludes that the persistence conditions of being human are being the human animal. While they both have merit, the important point here is that both animalist accounts and the psychological approach both cannot fulfill all the functions we want our concept of personal identity to fulfill.

Crucially, the animalist approach and the psychological approach are both incompatible. This is partly because when we refer to a person X, we will often either be referring to them as the human animal or them in a more abstract sense such as their mind. Another important point to make is that any attempt to make personal identity either a conjunction or disjunction of the psychological and animalistic approach will fail. The conjunction would fail in the thought experiments that involve these extreme conditions such as vegetative states or brain swapping. The disjunction also fails in the mind swapping experiment, as in that instance, a choice must be made which is more important the human animal or the psychology of the person. The main point of this section is simply to establish that no account that has been thought of so

[4] Jessica Gordon-Ruth, "Locke on Personal Identity", Stanford Encyclopedia of Philosophy, December 16, 2020, https://plato.stanford.edu/entries/locke-personal-identity.

[5] Eric Olson, "Personal Identity," in The Blackwell Guide to Philosophy of the Mind, ed. Stephen P. Stitch and Ted A. Warfield (Oxford: Blackwell, 2009): 359.

far can fulfill all the functions we want the concept of personal identity to fulfill, and it appears that animalism and the psychological approach are incompatible, something which Olson himself argued.[6]

Subjective Personal Identity

Firstly, subjective in this context means simply that there is no reason to state that some functions that we want personal identity to fulfill are more important than others. It means that there does not exist a consistent concept of personal identity. We clearly have some kind of concept of personal identity, so we must wonder what and why it is. The first thing that comes to mind is the phenomenology of personal identity. By this I am referring to the fact that through life we seem to experience it as if we were one person. We only ever have one perspective at a given time. This does not necessarily mean that we only have one perspective through our entire lives though. While we only ever experience life as one thing, we don't necessarily exist as one uniform thing.

The issue then becomes why should we conclude that personal identity is subjective? Why should we conclude that no account of personal identity can ever fulfill all the functions we currently want it to fulfill? While both animalism and the psychological approach can seem to be counterintuitive at times, we accept many theories even though they can be counterintuitive. This includes theories in mathematics that many people say are necessarily true. Unlike mathematics or science, accounts of numerical identity simply try to fulfill all the functions that we want that object or entity to fulfill rather than being empirically observed or logically deduced. The issue for personal identity as a specific form of numerical identity is finding any account of personal identity that does not seemingly arbitrarily say which functions are more important than others. I am talking here of functions that almost everyone agrees the concept should fulfill, such as the thought experiments previously stated. If someone were to come up with a ridiculous function, such as the time being eight o'clock, then as no one would agree with it then it would be less valid than functions that everyone agrees with. It is important to note that the functions that seem incompatible are the functions almost everyone wants the concept of personal identity to fulfill. Everyone wants their mind to experience happiness, rather than some other mind in their body. Everyone also wants to refer to a person as the same person even if they go into a vegetative state.

The main point to be stressed is there doesn't appear to be any method, other than our own subjective preferences, that can

[6] Ibid., 358.

validly show why some functions should be chosen while other functions should be ignored. Therefore, until such a method is found we will have to conclude any account of personal identity is subjective, as it is based on these subjective preferences. Any one person can admittedly have a preferred account if they find that they prefer certain functions to be fulfilled rather than others, but this isn't a good reason for why other people should choose their preferred account.

It is obviously necessary to refer to people as one thing over time. It seems reasonable that for us to have this convenience and to reconcile our experience as one thing we should produce the concept of personal identity. Philosophers have stated alternatives to having just one concept of self. However, the inconvenience of having to use the concept of an "earlier self" and "future selves" as Parfit suggests,[7] make it unlikely these concepts would ever catch on, although they can be incredibly useful for the metaphysics of personal identity. An important point to be drawn here is that concepts based on our subjective experience, without any universal agreement on them, can still be very important and people can still have very deeply held beliefs about them. The law, while just a social construct, still has a major significance in people's lives, and people can have very strong opinions about what the law should be. Similarly with ethics, Mackie, who pronounced that all ethics are subjective, maintained that one can still have deeply held beliefs about these values.

The social construct of personal identity is most likely influenced by things that can be analyzed as objects that have consistent persistence conditions, for example people's psychology and human biology. While the fields of psychology and biology are very effective at analyzing how people think and feel and how living systems function, it is far from obvious these fields could give an account of personal identity. Personal identity shares functions we want it to fulfill with things like human psychology and biology; however, it is unclear how these could be combined in some way to fulfill them all. It seems unlikely that an account could fulfill all these functions. It is important to note finally that this paper is assuming a materialist outlook of the world. If each person has a soul that is always the same and obeys Leibniz's law and it can be said that the soul is what a person truly is, then there is no need to talk of numerical identity of personhood. This is because there would be an unchanging identical thing ontologically. However, just as I have assumed there is no method for determining which functions are more

[7] Derek Parfit, "Personal Identity," The Philosophical Review 80, no. 1 (1971): 25

important than others because there is little to no evidence for doing so, I am assuming there is no soul and therefore little to no evidence of it.

Conclusion

Numerical identity is the property of two objects being the same across time. The numerical identity of an object occurs when persistence conditions can be ascribed to an object, meaning that an object has some properties which, if left unchanged across a given period of time, could be said to be the same object. We choose persistence conditions to fulfill the functions that we think the object or entity must fulfill in order for that object or entity to be itself. In the case of numerical identity of personhood, no account of the persistence conditions can fulfill all the functions we want them to fulfill therefore, all attempts at accounts so far require prioritizing some functions as more important than others. As there does not seem to be any method for explaining why some functions are more important than others, other than our own subjective preferences, we must conclude that any account of personal identity is subjective.

Bibliography

Belohrad, Radim. “Subjective Theory of Personal Identity and Practical Concerns.” *Organon F* 22, no. 3 (2015): 282-301.

Olson, Eric. “Personal Identity.” In *The Blackwell Guide to Philosophy of Mind*, edited by Stephen P. Stitch and Ted A. Warfield, 352-368. Oxford: Blackwell, 2002.

Parfit, Derek. “Personal Identity.” *The Philosophical Review* 80, no.1 (1971): 3-27.

Stanford Encyclopedia of Philosophy. “Locke on Personal Identity” Accessed December 16, 2020. https://plato.stanford.edu/entries/locke-personal-identity.

Williams, Bernard. “The Self and the Future.” *The Philosophical Review* 79, no. 2 (1970): 161-180.

No Predicative-Quantificational Ambiguity: On Hanson's Problem for Evolutionary Debunking

Isaac Hadfield, Trinity College, Oxford

Abstract

Evolutionary debunking arguments (EDAs) aim to undermine the justifications for our beliefs by highlighting the distortive effect that evolutionary processes have had on them. In particular, the influence of evolution undermines our moral beliefs, because natural selection selects based on fitness rather than being responsive to objective moral truth. In response, some suggest that because natural selection affects populations rather than individuals, it is a mistake to think that natural selection undermines our individual moral knowledge. Louise Hanson words this objection more precisely, claiming that EDAs equivocate between quantificational and predicative evolutionary explanations. This paper argues that this response to EDAs is mistaken, because it overgeneralizes from one kind of epistemically impotent quantificational explanation to other kinds of quantificational explanations.

Introduction

In *The Descent of Man*, Darwin considers what it would be like if bees had the capacity to make moral judgements. Darwin says that under bee-morality, "there can hardly be a doubt that our unmarried females would, like the worker-bees, think it a sacred duty to kill their brothers, and mothers would strive to kill their fertile daughters; and no one would think of interfering."[1] Moral realism holds that there are moral facts, and that these facts are objective, independent of our subjective evidence or attitudes. Evolutionary debunking arguments (EDAs) suggest that because natural selection affects our moral judgements and beliefs, and natural selection selects for fitness, irrespective of

[1] Darwin, Charles. *The Descent of Man, and Selection in Relation to Sex.* (London: John Murray, 1871), 73.

objective moral truths, we are not justified in thinking that our moral beliefs align with objective truth. Under different environmental conditions, perhaps natural selection would have caused human morality to resemble Darwin's hypothetical bee-morality more closely. Therefore, the worrying thing about evolution is that it casts doubt on an objective and attitude-independent account of moral facts. EDAs are strictly epistemological: they do not bear on the nature of moral facts by directly casting doubt on their existence, but rather indicate that we could not know them if they did, because the process of natural selection acts as a defeater for our justifications.

A recent strand of argument against EDAs suggests that natural selection does not undermine our moral knowledge, because it affects populations rather than individuals, and so does not affect individuals epistemically. This suggestion has most notably been made by Mogensen[2] and in even stronger terms by Hanson, who sees EDAs as equivocating between quantificational and predicative evolutionary explanations.[3] Section 2 outlines the structure of EDAs against moral realism, before section 3 explains Hanson's predicative-quantificational distinction and the problem this poses for EDAs. Section 4 discusses the importance of the negative view of evolution's explanatory power to Hanson's argument, and generates a case undermining Hanson's argument against EDAs. Section 5 shows that this case represents a broader failure of Hanson's predicative-quantificational distinction and its application to evolutionary explanations.

Evolutionary Debunking Arguments

Evolutionary debunking arguments start with an evolutionary claim: that the process of natural selection has significantly affected our moral beliefs and judgements. Given that natural selection is indifferent to the truth of our moral beliefs and judgements, selecting instead for fitness in an environment, it follows that our moral beliefs and judgements have been significantly affected by a process that is indifferent to their truth.

1. Evolutionary claim: Our moral beliefs and judgements have been significantly affected by the process of natural selection.
2. Natural selection is indifferent to the truth of our moral beliefs and judgements.
3. Our moral beliefs and judgements have been significantly affected by a

[2] Mogensen, Andreas L. "Do Evolutionary Debunking Arguments Rest on a Mistake about Evolutionary Explanations?", *Philosophical Studies*, 173 (2015): 1799–817.

[3] Hanson, Louise. "The Real Problem with Evolutionary Debunking Arguments". *The Philosophical Quarterly* (2016): 1–26.

process that is indifferent to their truth (from 1, 2).[4]

The route from (3) to the conclusion that we are not justified in thinking that our moral beliefs and judgements align with attitude-independent moral truths may take a number of different paths. Sharon Street's evolutionary debunking argument suggests that there are two explanations of the connection between evolution's influence and moral truth open to the moral realist. First, the realist can claim that there is no connection between evolution's influence on our moral attitudes and independent moral truths.[5] As such, unless we are right by sheer luck, most of our evaluative judgements would be wrong, having been distorted by evolutionary pressures. Second, the realist could claim that there is a connection between evolution's influence and independent moral truths, but this is simply scientifically untenable. We might then complete our debunking argument as follows:

4. If our moral beliefs and judgements have been significantly affected by a process that is indifferent to their truth, it would be a sheer coincidence if they aligned with attitude-independent moral truths.
5. If it would be a sheer coincidence that our moral beliefs and judgements align with attitude-independent moral truths, then we are not justified in thinking that they do.
6. We are not justified in thinking that our moral beliefs and judgements align with attitude-independent moral truths (from 3, 4, 5).

Richard Joyce illustrates a similar kind of argument with the example of the Napoleon pill. Suppose an imaginary belief pill could dispose you to form beliefs about a concept which you might not otherwise have had beliefs about. For example, there might be a Napoleon pill that makes you form random beliefs about Napoleon. If you discovered that you had been slipped such a Napoleon pill, this would naturally undermine all your Napoleon-related beliefs. It would not necessarily prove your beliefs about Napoleon to be false, but in the absence of reliable and independent evidence, your justification for holding your Napoleon beliefs would be lost.[6] If our moral beliefs and judgements have been shaped by natural selection, indifferent to their truth, Joyce says we are in the position of having

[4] Adapted from Hanson, Louise. "The Real Problem with Evolutionary Debunking Arguments". *The Philosophical Quarterly* (2016): 3.

[5] Street, Sharon. "A Darwinian Dilemma for Realist Theories of Value", *Philosophical Studies*, 127 (2006): 109–66.

[6] Joyce, Richard. *The Evolution of Morality (Life and Mind: Philosophical Issues in Biology and Psychology)*. MIT Press, 2005.

taken a Napoleon pill with regards to our moral beliefs, and so all the beliefs we have concerning morality are undermined. If we accept this argument, then it is clear that facts about evolution are important for an account about the nature of moral facts. Namely, if we are not justified in thinking that our moral beliefs and judgements align with attitude-independent moral truths, then it appears untenable to maintain a realist account of our moral knowledge.

It is worth pointing out that one move open to the realist is simply to deny the first premise. For Street, all our powers of reasoning and judgement have been affected by evolution to increase our chances of survival, and this is supposed to obviously include our moral judgements and beliefs. Nagel simply takes the evolutionary debunking argument to be a *reductio* of the empirical premise.

> "Since moral realism is true, a Darwinian account of the motives underlying moral judgment must be false, in spite of the scientific consensus in its favor."[7]

This being said, I agree with Katia Vavova that we should proceed on the assumption that the first premise is true, so as not to "shrug off" this philosophically interesting argument.[8] A more common approach to defending against EDAs is to deny the fourth premise. On this move, it is not a sheer coincidence that our moral beliefs and judgements align with attitude-independent moral truth, because there is some third factor that we have developed as evolutionarily advantageous, and yet also allows us to know attitude-independent moral truths. Parfit, for instance, suggests that we have evolved to be able to respond to reasons and rationality.[9] This, in turn, is what grounds our moral capacities. Erik Wielenberg makes a similar move, suggesting that our cognitive capacities, which have evolved due to natural selection, have given us the ability to understand the concept of rights.[10] There is an expansive literature proposing third-factor accounts and responses to them. The focus of this paper, however, is on a more recent strategy against EDAs, which suggests that natural selection affects populations rather than individuals, and therefore cannot act as a defeater to the moral beliefs of individuals.

Hanson's Response to EDAs: Predicative/Quantificational Equivocation

[7] Nagel, Thomas. *Mind and Cosmos*. Oxford: Oxford University Press (2012), 105.

[8] Vavova, Katia. Evolutionary Debunking of Moral Realism. *Philosophy Compass*, *10*(2) (2015), 109.

[9] Parfit, Derek. *On What Matters. Volumes 1 & 2*. Oxford: Oxford University Press, (2011). Vol. 2, 531.

[10] Wielenberg, Erik. "On the Evolutionary Debunking of Morality", *Ethics*, 120 (2010): 441–64.

Hanson has offered a recent response to evolutionary debunking arguments.[11] She suggests that evolutionary debunking arguments rest on an equivocation between predicative readings and quantificational readings of their premises. Recall the evolutionary claim: Our moral beliefs and judgements have been significantly affected by the process of natural selection. In predicating we say something about a subject, and so on a predicative reading of the evolutionary claim, the premise is that "having been shaped by natural selection" predicates on individuals, such that I, as an individual, have had my moral beliefs and judgements shaped by natural selection; the same is true of you and every other individual. In contrast, according to Hanson, on a quantificational reading the evolutionary claim quantifies over people in general, so the premise is that natural selection explains the fact that individuals with these similar kinds of moral beliefs exist.[12] Hanson calls this a quantificational reading because it takes the explanation as quantifying over the population, explaining why all, most, or many people have the beliefs that they do. The result is that we can generate two different versions of the evolutionary debunking argument based on whether the evolutionary claim is read predicatively or quantificationally.

Predicative reading of the evolutionary debunking argument:

> 1a. Predicative claim: you, I, and everybody else have each individually been caused to accept the moral beliefs we accept by the forces of natural selection.
> 2. Natural selection is indifferent to the truth of these beliefs.
> 3a. Forces that are indifferent to moral truth have made it the case that I have these moral beliefs, and you do, and that other individuals do, etc.[13]

Quantificational reading of the evolutionary debunking argument:

> 1b. Quantificational claim: natural selection has made it the case that individuals with these moral beliefs exist, rather than individuals with different moral beliefs.
> 2. Natural selection is indifferent to the truth of these beliefs.
> 3b. Forces that are indifferent to moral truth have made it the case that individuals with these moral beliefs exist, rather than individuals with different moral beliefs.[14]

[11] Hanson, Louise. "The Real Problem with Evolutionary Debunking Arguments". *The Philosophical Quarterly* (2016): 1–26.
[12] Ibid, 1-26.
[13] Hanson, "The Real Problem with Evolutionary Debunking Arguments", 5–6.
[14] Ibid, 6.

Hanson argues that the evolutionary claim is only plausible on the quantificational reading. For Hanson, when talking about evolution and natural selection with regards to traits and tendencies in individuals, we are saying, "that individuals who had that trait or tendencies towards it had more offspring than individuals who didn't, resulting, ultimately, in a situation where individuals with that trait existed rather than individuals who lack this trait."[15] So evolutionary theory supports a quantificational evolutionary claim but not a predicative one: as Hanson puts it, "what would it even be for the forces of natural selection to push you, or me, or any other individual, towards certain beliefs?"[16]

However, Hanson suggests that only the conclusion of the predicative reading of the evolutionary debunking argument is epistemically worrying. What matters for us is whether the backstory of an individual's belief-formation is compromised as in 3a. On the other hand, she suggests that it is not clear why the quantificational conclusion in 3b should be particularly troubling; just the fact that the "world came to contain" people with certain moral beliefs is not relevant to the particular epistemic backstory of a given individual.

We can see that in the Napoleon pill case, all the beliefs you have concerning Napoleon are undermined because you have personally taken the Napoleon pill. To be truly analogous then, the evolutionary claim would have to be read in the same predicative way, directly affecting the beliefs we have as individuals. But for Hanson, the evolutionary claim is not plausibly predicative in this way: "the evolutionary claim says nothing about the causal process by which individuals formed their moral beliefs—it is just a claim about the causal process that led to *there being lots of individuals* who have the beliefs in question."[17] The problem Hanson raises for EDAs then is that in appearing to have a devastating impact on the justification of our beliefs, EDAs are in fact equivocating between the plausible quantificational evolutionary claim and the epistemically worrying predicative conclusion.

We might initially think that the quantificational conclusion in 3b would in fact be epistemically troubling. After all, it is intuitive to think that evolution "making it the case that individuals with certain moral beliefs exist" should indeed make us reconsider our moral beliefs as individuals. It looks like the quantificational conclusion is exactly the type of conclusion we take to be epistemically worrying in cases like Darwin's bee-morality example; natural selection could make it the case that there

[15] Hanson, "The Real Problem with Evolutionary Debunking Arguments", 7.

[16] Ibid, 7.

[17] Ibid, 9.

exist female bees with the moral judgement that it is a sacred duty to kill their brothers. However, the reason why Hanson takes this quantificational conclusion not to be epistemically concerning for individuals is demonstrated by the following case paraphrased from Roger White:[18]

> *Napoleon Party*: At the door of a party, partygoers are asked whether Napoleon invaded Egypt. You know that Napoleon did invade Egypt, and so respond "Yes." Later, you discover that after asking this question, the host killed everyone who answered "No", while letting everyone who answered "Yes" into the party. Further, the host decided which side to kill and which to let in on the basis of a coin flip.

White points out that when you later discover what has happened, the process, though disturbing, does not challenge your belief that Napoleon invaded Egypt. Even though the group of party goers all believe that Napoleon invaded Egypt, and the explanation for why the group all believe that Napoleon invaded Egypt is independent of its truth, this fact about the group does not entail anything about why you have this belief rather than another belief. Similarly, just by the fact that evolution, a force that is indifferent to moral truth, has made it the case that individuals with certain moral beliefs exist, it does not follow that evolution challenges our beliefs as individuals. Only a predicative conclusion, then, would be epistemically problematic, but only a quantificational evolutionary claim is plausible. According to Hanson, evolutionary debunking arguments gain their plausibility from equivocating between predicative and quantificational premises.

Natural Selection and Explanation

Hanson's argument implicitly relies on the negative view concerning whether natural selection can explain the traits of individuals. According to the negative view, natural selection can only explain population level properties.[19] This is one side of the debate over whether natural selection can in fact explain the traits that individuals have. One contrary view, developed by Karen Neander[20] and Mohan Matthen[21] is that "selection might

[18] White, Roger. You just believe that because…. *Philosophical Perspectives*, 24 (2010), 573–615.

[19] Gildenhuys, Peter. "Natural Selection", The Stanford Encyclopedia of Philosophy (Winter 2019 Edition), Edward N. Zalta (ed.), Retrieved 9 April 2020.

[20] Neander, Karen. "What Does Natural Selection Explain? Correction to Sober", *Philosophy of Science*, 55 (1988): 422–6.

[21] Matthen, Mohan. "Evolution, Wisconsin Style: Selection and the Explanation of Individual Traits", *British Journal for the Philosophy of Science*, 50 (1999): 43–150.

determine an organism's traits by affecting the frequency with which certain alleles occur in the ancestral gene pool from which the organism must inherit its genotype."[22] In this way, we might think that although natural selection does operate at the group level, it has an individual impact such that we should still accept the predicative evolutionary claim.

This idea that evolution, while having generally group level quantificational explanatory implications, also has individual predicative ones, is considered by Hanson. However, she argues that because the fundamental explanatory power is quantificational, the process of evolution still cannot generate epistemically relevant worries. The reason for this is that Hanson takes it to be the case that even those who think that natural selection can explain traits in individuals accept that the explanatory work that natural selection of traits can do (whether in the predicative or the quantificational sense) is done by explaining the frequency of traits in a population. For Hanson however, it is still the case that explanations of the *frequency* of our moral beliefs cannot have any bearing on the epistemic status of these beliefs, even if we have to accept that natural selection can in some way explain the traits of individuals. Because this explanation is still in terms of frequency, Hanson believes it is still more analogous to White's party case rather than the Napoleon pills.

Although this kind of explanation might not be the same as the definite discovery that you were slipped one of the Napoleon pills, it is not clear that relying on frequency undermines the bearing of these explanations on the epistemic status of moral beliefs. In White's party case, the causal process that led to group members having particular beliefs about Napoleon is completely unconnected to the backstory of how they came to acquire those beliefs. However, in the case of evolution, if Neander and Matthen are right that natural selection explains individual traits by affecting the frequency of particular alleles in our ancestral gene pool, the causal process is linked to the backstory of how individuals within the group came to acquire those beliefs.

Consider by way of analogy an alternative version of White's party case in which there are two entrances to the party. Admission through the first entrance is just as before; based on an initial coin flip, individuals with certain beliefs about Napoleon are either let into the party or killed. Those joining the party through the second entrance are all individuals with no beliefs

[22] Mogensen, Andreas L. "Do Evolutionary Debunking Arguments Rest on a Mistake about Evolutionary Explanations?", *Philosophical Studies*, 173 (2015): 1804.

about Napoleon at all. When the people in the party mix, individuals from the first entrance mention their beliefs about Napoleon to those from the second entrance, who come to share those beliefs on the basis of this testimony. When those from the second entrance come to realize that the people they have been conversing with were selected for their beliefs about Napoleon on the basis of sheer chance, it seems clear that this fact should affect their confidence in their beliefs about Napoleon. Yet this is not a predicative case like Joyce's original Napoleon pill example. The causal chain from the testimony of the partygoers, and from the individual partygoers to the truth about Napoleon, may still be perfectly sound. The epistemically worrying event is, in this case, still an event which affects the probability of individuals with certain beliefs existing compared to not existing, not one that explains the origins of particular individuals' beliefs.

Quantificational Explanation and Epistemic Defeat

It may be tempting, at this point, to conclude that whether evolutionary debunking arguments against moral realism are successful depends on whether the negative view about natural selection is true or false. This is, in effect, the concession that Mogensen makes: if natural selection can explain the traits that individuals have, then it would act as a defeater for our moral beliefs, but current accounts of this process such as those posited by Neander and Matthen are inconclusive. However, entering into this debate is not necessary here: a process may affect an individual's acquisition of a trait without being part of the direct causal history which explains that trait's acquisition. For this reason, the problem that the two-entrance-party example raises for Hanson is worrying for her analysis independently of whether the negative view is true or not.

Consider Elliot Sober's example of a classroom in which all the children are able to read at the Grade Three level.[23] The explanation for why the children are able to read at the Grade Three level may similarly suffer from this kind of quantificational-predicative ambiguity. One explanation, in the quantificational sense, would be that being at a Grade Three level is a requirement to join the class. A contrasting explanation, in the predicative sense, would involve giving an explanation for why each individual child is able to read at the Grade Three level. Hanson takes this example to be illustrative of quantificational-predicative ambiguity as a fundamental divide in kinds of explanations

[23] Sober, E. *The Nature of Selection: Evolutionary Theory in Philosophical Focus.* Cambridge, MA: MIT Press (1984), 149.

for traits in members of groups.[24] We can see now that this divide is confused. Consider the potential quantificational and predicative readings that Hanson's analysis might apply to those with newfound Napoleon beliefs in the two-entrance-party case:

> *Predicative claim*: you, I, and everybody else have each individually been caused to accept the Napoleon beliefs we accept by the random selection of a coin flip.
> *Quantificational claim*: the random selection of a coin flip has made it the case that group members have these Napoleon beliefs rather than other Napoleon beliefs.

The predicative claim here seems implausible. The individuals came to hold their beliefs about Napoleon because of the testimony of the other partygoers, whose beliefs have an origin which is independent of the coin flip. As such, it appears very odd to say that this group level selection mechanism of the coin flip has caused each individual to have the Napoleon beliefs that they do. The quantificational claim is the only plausible reading of the two-entrance-party case, yet as we have seen, it does appear that the knowledge of this selection mechanism is epistemically worrying for those individuals from the second entrance. The reason why this seems so epistemically compromising is because of our anti-luck intuition about knowledge: a belief that we hold cannot count as knowledge if it is only by sheer chance that our belief actually corresponds to the truth. If the other side of the coin was selected by the coin toss, then people with different Napoleon beliefs would have entered the party. As such, the people from the second entrance would have come to hold entirely different beliefs.

This is not problematic for Hanson only because the case is intuitively epistemically problematic for those from the second entrance: this would beg the question by taking the defeat of the individuals' justifications as a premise. Rather, it demonstrates that what Hanson classifies as quantificational overgeneralizes. Examples of quantificational explanations such as "being at a Grade Three level is required for children to join the class" and "answering 'Yes' is required to join the party" are particular kinds of quantificational explanations: they are statements about a group's membership conditions, and do not influence the acquisition of the properties in question by individual group members. In White's party case, those admitted to the party already have beliefs about Napoleon, and the murderous policy of the host does not affect

[24] Hanson, Louise. "The Real Problem with Evolutionary Debunking Arguments". *The Philosophical Quarterly* (2016), 5.

this. While this kind of quantificational explanation may indeed not be epistemically problematic, this does not mean all group level explanations do not affect the trait-acquisition of individual group members, and it is in these cases that there may be epistemic compromise. In the case of evolution, our moral beliefs and judgements have already been selected for. Therefore, when considering our own moral beliefs and judgements, we are in the position of those who gain their beliefs about Napoleon from the party host's policy at the first entrance; learning about this selection process should similarly undermine our confidence about these beliefs. As such, we should maintain a quantificational reading of the evolutionary claim but take the conclusion to be epistemically worrying for individuals.

Conclusion

Facts about evolution matter because they cast doubt on our justifications for thinking that our moral beliefs and judgements align with attitude-independent moral truths. The suggestion that quantificational readings of the conclusion of EDAs are not epistemically problematic rests on a mistake in defining quantificational explanations only with examples which do not affect the trait-acquisition of individuals. This effect on individuals does not entail that the explanation is in fact predicative, as it does not necessarily explain the causal history for the specific traits of individual group members. This is demonstrated with the two-entrance-party case, in which a group level process has an epistemically damaging effect on the acquisition of group members' beliefs without explaining this acquisition at the individual level. As such, we should maintain a quantificational reading of the evolutionary debunking argument but take the conclusion to be epistemically worrying for individuals. In conclusion, the strength of evolutionary debunking arguments does not rest on predicative-quantificational ambiguity.

Bibliography

Darwin, Charles. *The Descent of Man, and Selection in Relation to Sex.* London: John Murray. Volume 1. 1st edition.

Gildenhuys, Peter. "Natural Selection", The Stanford Encyclopedia of Philosophy (Winter 2019 Edition), Edward N. Zalta (ed.), Retrieved 9 April 2020.

Hanson, Louise. "The Real Problem with Evolutionary Debunking Arguments". *The Philosophical Quarterly* (2016): 1–26.

Joyce, Richard. *The Evolution of Morality (Life and Mind: Philosophical Issues in Biology and Psychology)*. MIT Press, 2005.

Matthen, Mohan. "Evolution, Wisconsin Style: Selection and the Explanation of Individual Traits", *British Journal for the Philosophy of Science*, 50 (1999): 43–150.

Mogensen, Andreas L. "Do Evolutionary Debunking Arguments Rest on a Mistake about Evolutionary Explanations?", *Philosophical Studies*, 173 (2015): 1799–817.

Nagel, Thomas. *Mind and Cosmos*. Oxford: Oxford University Press, 2012.

Neander, Karen. "What Does Natural Selection Explain? Correction to Sober", *Philosophy of Science*, 55 (1988): 422–6

Parfit, Derek. *On What Matters. Volumes 1 & 2.* Oxford: Oxford University Press, 2011.

Sober, E. *The Nature of Selection: Evolutionary Theory in Philosophical Focus.* Cambridge, MA: MIT Press, 1984

Street, Sharon. "A Darwinian Dilemma for Realist Theories of Value", *Philosophical Studies*, 127 (2006): 109–66.

Vavova, Katia. Evolutionary Debunking of Moral Realism. *Philosophy Compass*, *10*(2) (2015), 104–116.

White, Roger. You just believe that because…. *Philosophical Perspectives*, 24 (2010), 573–615.

Wielenberg, Erik. "On the Evolutionary Debunking of Morality", *Ethics*, 120 (2010): 441–64.

Challenge Trials in Vaccine Development: An Institutional Moral Failure

Avery Caulfield, Tufts University

Abstract

Expediting vaccine development processes while maintaining their safety and efficacy is a primary objective of the pharmaceutical industry, irrespective of exceptional public health circumstances. One tool that has historically been used to accelerate this process is the *challenge trial*. For reasons explored in this essay, there has been a near-universal rejection of employing this method in the development of a vaccine for SARS-CoV-2. To probe the ethical legitimacy of this rejection, I posit that the issue can be reformulated in terms of an altered version of the familiar trolley problem. Using this reformulation, I argue that the repudiation of challenge trials during the 2020 pandemic is based on illegitimate ethical grounds, such as the *primum non nocere* (first, do no harm) clause of the Hippocratic Oath. Finally, I call for a reexamination and reformation of the widespread principles that led to this institutional moral failure.

The Trolley Problem

Imagine you are spending an afternoon walking along a set of trolley tracks. As you turn a bend, you are met with a dire crisis: at least 10,000 people were kidnapped, blindfolded, dragged kicking and screaming, and tied down to the tracks before you. An out-of-control trolley is moving towards them and will kill 500 of them if no action is taken. There is a lever before you that, when pulled, diverts the trolley to another track. On this second track, there are 100 people, 5 of whom the trolley will kill. This second group of people voluntarily stepped onto the track, and have posted signs beside themselves, saying things like "please divert the trolley," "I know I might die, but I accept that risk, in hopes that I will save more people on track 1," etc. Do you pull the lever?

The above thought experiment is an altered version of the classic dilemma, the trolley problem, which can provide insight into a crucial moral issue facing our medicinal institutions: the decision of whether or not to conduct challenge trials in the development of vaccines for SARS-CoV-2. In this essay, I'll argue that the above problem faithfully represents the relevant moral considerations at play on the question of challenge trials. My use of the trolley problem will be twofold: it will (1) orient the reader's moral considerations, such that one's answer to the problem ought to inform one's position on challenge trials[1] and (2) facilitate an argument from democracy by providing insight into the public's related moral intuitions. In the course of developing the comparison, I'll undercut the oft-cited ethical basis for rejecting challenge trials.

Vaccine Development

Vaccine development is, more often than not, a tedious and lengthy process, usually spanning 10-15 years.[2] Such a duration is often necessary to ensure the safety and efficacy of a substance administered to large numbers of people. Because of the current circumstances, many readers will be familiar with the three-phase method usually employed in vaccine development, in which a candidate vaccine or a placebo is administered to an increasing number of people in each stage. After developers administer the vaccine, these individuals are released back into the population, and their rates of infection are examined relative to those of a placebo group. Many tactics have been employed to accelerate the current process of SARS-CoV-2 vaccine development, like overlapping phases and beginning vaccine manufacturing before the completion of human trials. Despite these efforts, it remains the case that thousands of people *per day* die from COVID-19, and the majority of the world is still many months, if not years, away from universal access to a vaccine. Needless to say, it is imperative that we develop effective vaccines as quickly as possible.

Challenge Trials

In the past, the pharmaceutical industry has taken advantage of challenge trials to fast-track vaccine development. Rather than releasing the participants back into the world, challenge trials entail intentionally infecting consenting participants with the pathogen to quickly test the efficacy of a candidate vaccine.

In March 2016, the Bill and Melinda Gates Foundation funded a challenge trial in the U.K. on about 100 willing and informed participants in hopes of accelerating the development of a typhoid vaccine.

[1] See the pond example in Singer (1972) for a textbook example of this strategy.

[2] "Vaccine Development, Testing, and Regulation."

According to Dr. Anita Zaidi, Director of the Diarrheal Diseases Program at the Foundation, this trial allowed them to forgo a process of developing their vaccine that would otherwise have taken 3-4 years.[3] So, what is holding us back from conducting a similar trial to hasten the development of a SARS-CoV-2 vaccine? The most common answer to this question is that for each of the challenge trials that have occurred in the past, there has been an effective treatment for the pathogen to which we have exposed participants; if participants developed serious complications, the treatment could have been administered, thereby evading any potentially fatal consequences of conducting a challenge trial. Because there is not yet an effective treatment for COVID-19, opinions vary greatly about how to move forward with a challenge trial, if at all. Some, like AIDS Vaccine Advocacy Coalition (AVAC) and Treatment Action Group (TAG), reject that a challenge trial should be conducted at all until there is a safe and effective treatment.[4] Others, like Peter Smith in his March 2020 paper on the subject, postulate that lack of an effective treatment entails that:

> their net risk could be acceptable if participants comprise healthy young adults, who are at relatively low risk of serious disease following natural infection, they have a high baseline risk of natural infection, and during the trial they receive frequent monitoring and, following any infection, the best available care.[5]

While it would involve less risk to the participants if they comprised young, healthy adults, such a trial would be limited in its ability to demonstrate that a vaccine is safe and effective for all members of the population, as its sample would not include participants representative of vulnerable communities.

Argument

Here, I contend that the utilization of challenge trials on a group of informed, consenting participants from a randomized sector of a volunteer population (of any age) would be significantly more ethical than failure to conduct challenge trials at all. The choice, then, that burdens vaccine developers should not be whether to conduct challenge trials, but rather on whom to conduct these trials.

To begin, let's compare a challenge trial volunteer to a member of the population who contracts COVID. In a challenge trial,

[3] McNeil, Donald G., Jr., "They Swallowed Live Typhoid Bacteria — On Purpose."

[4] Treatment Action Group, *AVAC and TAG Statement on Ethical Conduct of SARS-CoV-2 Vaccine Challenge Studies.*

[5] Eyal, Lipsitch, and Smith, "Human Challenge Studies to Accelerate Coronavirus Vaccine Licensure."

medical researchers would inform volunteers of the risks of contracting coronavirus, obtain their consent despite these risks, and supervise them for the duration of their illness. A member of the population who contracts COVID, on the other hand, is unlikely to be similarly informed, certainly does not consent to their infection, and oftentimes does not have access to sufficient medical care. Consider each of the two above individuals; which of them would it be more ethical to infect with Coronavirus? From my perspective, there is a clear answer to this question that aligns with bioethical orthodoxy: if someone is going to be infected with an illness, it would be better to have their informed consent. I suspect the reader strongly agrees without my providing justification.[6]

If conducting challenge trials expedited vaccine development (and thus vaccine distribution) at all, then doing so would reduce the number of nonconsenting, uninformed people who are naturally infected and killed. In their place would be consenting and informed individuals infected and killed in a lab setting. On this analysis, even if it were the case that the number of victims in a challenge trial was *equivalent* to the number of lives in the population that the trial saved, it would still be morally preferable to conduct the trial. In other words, conducting challenge trials would be morally preferable to failure to do so *even if we ignored* the extraordinary utilitarian benefit that would likely result from our conducting them.

In actuality, there are weighty utilitarian considerations at play that further strengthen the case for conducting challenge trials in the development of a SARS-CoV-2 vaccine. In all likelihood, doing so would save many more lives than it risks. Peter Smith's paper estimates that about 100 volunteers would be required to undertake a single challenge trial. Taking a mortality rate estimate of 5%, about five people would die in a COVID challenge trial.[7] Let's take an extremely conservative estimate of the degree to which a challenge trial would speed up vaccine development: one month. Once more conservatively estimating, let's say that this month of additional time would save just 10,000 people from the general population from being infected, thus preventing 500 deaths (taking the same mortality rate).

[6] The question of why we consider these factors morally relevant is distinct from the one at hand, but standard answers appeal to self-protection and bodily autonomy, in the tradition of John Stuart Mill's "Over himself, over his own body and mind, the individual is sovereign" (*On Liberty*, 22).

[7] It is irrelevant to my argument which mortality rate estimate we choose here, as the same estimate is applied to both the naturally infected population and the challenge trial volunteer population.

To reiterate, the assumptions in the above estimates are *extremely* conservative. The far more likely outcome of conducting challenge trials is a *significantly* expedited vaccine development process, saving many more lives than 500. Despite this conservative skew, the moral consequence remains clear: conducting challenge trials is vastly more ethical than ignoring them.

Notice that this argument does not take into account potential underlying conditions of volunteers; instead, it simply compares a scenario in which we conduct a challenge trial to one in which we do not. Assuming the soundness of the above argument, the remaining decision facing the pharmaceutical industry involves the degree of restrictions on volunteer selection. Whether excluding groups of people more at-risk from the potential volunteer pool would greatly reduce vaccine efficacy is an empirical question. Obviously, given the choice, it would be preferable to infect someone less likely to face complications with the virus; however, if excluding these groups would compromise vaccine efficacy, the above argument stands: it would be morally preferable to infect an informed and consenting at-risk member of the population than one who is uninformed and nonconsenting.

Importantly, this line of reasoning could not be used to justify, say, harvesting one nonconsenting person's organs to save five others. This would be the case if I had argued on purely utilitarian grounds, but because the principal reasoning is based on discrepancies in knowledge and consent, no such comparison can be conducted validly. In the organ-harvesting case, the sacrificed individual is neither informed nor consenting. Contrast this with challenge trials, in which the trial volunteers are consenting, and the general population is not. The above argument for the morality of challenge trials stems precisely from the vast differences between these two circumstances.

From Numbers to Trolley Problem

This conclusion, while already clear, is even more starkly elucidated by putting all of these considerations in the form of the trolley problem. Accurately doing so necessitates representation of the relevant moral factors in the problem's structure, namely the distinctions in preparedness, knowledge, consent, and number between the two groups. Accordingly, the members of the general population are blindfolded (unaware of the risks they encounter upon contracting the illness), dragged to the tracks kicking and screaming (resisting contraction of the illness), and subsequently tied down to the tracks (given no choice in the matter). On the other hand, the

challenge trial volunteers step onto the trolley's tracks with the knowledge that they are risking their lives by doing so. Like many challenge trial volunteers–for example, some members of the group 1Day Sooner[8]–they are actively advocating that the lever be pulled.

Limitations of Using a Trolley Problem

Both the structure of the trolley problem and the assumptions I have made have their drawbacks, distorting relevant considerations; in this section, I will explore a few. I assumed the same mortality rate for the members of the general population and the challenge trial participants. In fact, a member of the general population would face far greater risk than a challenge trial participant, for several reasons. Firstly, the challenge trial participant would receive better medical care than anyone in the population. Instead of self-monitoring for dangerous complications, like the general population, a volunteer would be regularly monitored by medical professionals overseeing the trial. In addition, a volunteer would be screened beforehand for vulnerability to significant illness, a luxury not afforded to the population at large. Finally, volunteers would be kept in a setting where they were not at risk of contracting other illnesses during the time their body was combatting COVID.

Another disadvantage lies in the inability of the trolley problem to adequately characterize the extent of suffering that would be prevented by conducting a challenge trial (i.e., pulling the lever). In the problem as presented, pulling the lever would prevent the deaths of 500 of the 10,000 people tied to the track. This ignores the fact that conducting a challenge trial would either prevent their becoming sick at all or greatly reduce the severity of their illness. It would more adequately represent the benefits of conducting a challenge trial if, upon pulling the lever, these people were never kidnapped and tied to the track in the first place. Crucially, these distortions, if accounted for, would only strengthen my case.

Argument from Democracy: An Advantage to Using a Trolley Problem

One benefit of reformulating the issue in terms of a trolley problem is the ability to observe where the general population stands on the matter. If an overwhelming majority of the population would pull the lever in the first section's problem, then conducting challenge trials is also the democratic decision. In the traditional formulation of the trolley problem, there are just five people on the initial track and a single person on the track that can be

[8] Bult, "The risky way to speed up a coronavirus vaccine."

switched to. The modified problem offers greater benefits for pulling the lever than in the traditional problem, with a [track 1]:[track 2] death toll ratio of 100:1 instead of the original 5:1. That pulling the lever is *more beneficial* in our problem means we can extrapolate the following: respondents to our problem would be *even more likely to pull the lever* than those responding to the traditional problem.

Data on the traditional problem is decisive. When the general population is surveyed, around 90% of people choose to pull the lever.[9] Misgivings about the legitimacy of the general population's moral intuitions are no concern, as philosophers are fairly decisive, as well, with around 70% of them pulling the lever.[10] If conducting a challenge trial is morally equivalent to pulling the lever in the aforementioned trolley problem, it's abundantly clear where both the general population and the philosophical community stand on the matter.

Despite the reasoning in the fourth section and the above data demonstrating the support of the population, the default position has consistently been to reject the use of challenge trials. To ascertain why this could be the case, I'll explore some potential counterarguments and establish their invalidity.

False Equivalency Argument

This argument usually takes the following form:

One cannot equate the pharmaceutical community's intentionally infecting someone with a virus with a naturally occurring infection in the general population, as doing so ignores the possibility that public trust in medical institutions might erode if we conducted a challenge trial that fatally harmed a volunteer.

It is correct that there are significant differences between a naturally occurring infection and one intentionally brought about in a lab setting. Importantly, many of these differences would lead us to *prefer* that a consenting volunteer in a lab setting be infected, rather than a member of the population, as discussed in the fourth section. This argument points out a potential negative consequence of intentionally infecting a volunteer: if they were to die in a challenge trial, public trust in medical institutions might crumble at a time when this trust is essential.

To begin, let's examine the data on public support for challenge trials: a 2020 study on the matter observed that a majority of respondents across every examined subgroup considered "challenge trials to be more likely to be ethical than standard

[9] Cloud, "Would You Kill One Person to Save Five?"
[10] Bourget and Chalmers, "What Do Philosophers Believe?"

trials."[11] Assuming the data on trolley problems (and thus relevant moral intuitions) hold, the evidence seems to be in favor of the opposite conclusion. Institutions that align themselves with public moral intuitions are more likely to garner trust, meaning challenge trials should be the way forward if one is concerned with maximizing public trust.

Having said that, the effects challenge trials have on public opinion is an empirical question. The data I've just reviewed indicates that challenge trials are unlikely to jeopardize public trust. Accordingly, the burden of proof falls on those advocating the opposite position, yet there has been no substantial demonstration that challenge trials have an effect on public opinion, let alone a significant one.

Finally, suppose conducting a challenge trial did erode public trust to some degree. What reason do we have to believe that this erosion would be more substantial than the one that results from our ushering in thousands of preventable COVID-related deaths by our failure to utilize a pharmaceutical tool at our disposal? Even if public trust in medicinal institutions was the only metric we were concerned about (which it isn't and shouldn't be), there would *still* be reason to believe we should conduct challenge trials.

Consent Argument

In order to give valid informed consent, a person must be appraised of the involved risks in their entirety. Because our knowledge of COVID complications is incomplete, we are unable to obtain bona fide informed consent.

This is a standard, widespread application of the notion of informed consent. Whether our present state of knowledge supports informed consent by challenge trial volunteers is a matter of widespread debate. Lisa Tambornino and Dirk Lanzerath posit in their July 2020 paper that, to provide informed consent, a volunteer would need to be briefed on the following four points:

1. The long-term effects of a COVID-19 infection remain unclear.
2. COVID-19 infection can be fatal.
3. Research participants need to fully disclose their medical history to determine their risk exposure.
4. Research participants may not be able to withdraw immediately from a study that is set in an inpatient setting.[12]

Whether the standard interpretation of informed consent can be met in the case

[11] Broockman, Kalla, Guerrero, Budolfson, Eyal, Jewell, Magalhaes, and Sekhon, "Broad cross-national public support for accelerated COVID-19 vaccine trial designs."

[12] Tambornino and Lanzerath, "COVID-19 Human Challenge Trials – What Research Ethics Committees Need to Consider."

of SARS-CoV-2 is an open question. However, it is uncontroversial that infecting a volunteer briefed on the above points and consenting nonetheless is significantly morally preferable to infecting someone nonconsenting and uninformed. Conducting challenge trials would directly cause some occurrences of the former, and failure to do so would indirectly cause many more occurrences of the latter, as was established in the fourth section. In my view, whether harm is brought about directly or indirectly by the pharmaceutical industry is morally negligible compared to the vast differences in number, knowledge, and consent between the two groups.

The "trials will not necessarily save lives" Argument

There is no guarantee that a challenge trial would save any lives; therefore, conducting a challenge trial could needlessly fatally harm volunteers, an unacceptable risk.

This argument is based on the false premise that a challenge trial demonstrating the inadequacy of a vaccine would mean the volunteer died in vain. Part of the process of successful vaccine development is the empirical elimination of faulty vaccines. Each demonstration that an ineffective vaccine is ineffective is a step towards success. Additionally, this argument neglects the possibility of conducting multiple challenge trials for all available vaccine candidates, in which case the argument in the fourth section stands by application to all conducted trials at once.

Argument from the Hippocratic Oath

Medical professionals have a responsibility to do no harm, per the primum non nocere *clause of the Hippocratic Oath. Despite our most vigilant precautions, we cannot guarantee that participants in a challenge trial could be protected from permanent harm; therefore, we cannot conduct a challenge trial.*

To begin, it remains the case that we can never guarantee complete protection from harm in many medical trials we do conduct. Absolutist applications of this clause like the one outlined above would prevent almost any drug from being administered to any person, as the potential for unexpected complications is never null.

This argument embodies a common application of the Hippocratic Oath, one that I often observe in philosophical discussions about the aforementioned organ-harvesting counterexample to utilitarianism. Its implications are strange and counterintuitive, namely that *any* harm done by inaction (like failure to conduct a challenge trial), *no matter the scale*, is not a morally relevant consideration compared to harm done by action. Justification for this principle often takes the form of points already addressed, like maintaining public

trust in medical institutions. Vitally, this is exactly the moral consideration implicated in the trolley problem. Based on this widespread application of the Hippocratic Oath, the medical community would *never* pull the lever. Consider how strongly this diverges from your own moral intuitions, the moral intuitions of the public, and the moral intuitions of the philosophical community. For what purpose have we adopted this strange notion of harm? In my view, it is acceptable to afford *some* additional ethical weight to harm brought about by action, as opposed to inaction; but to reject *consideration* of potential action-based harm to informed and consenting individuals, even if it would prevent substantially more inaction-based harm to the uninformed and nonconsenting, is entirely foreign to even the most fundamental moral intuitions.

Rather than accepting this application of the Hippocratic Oath as dogma, let's ask ourselves: Who would a challenge trial harm? The consenting, informed people who decided, of their own volition, to take the risk of contracting a deadly disease for the betterment of humanity. Who does failure to conduct such a trial harm? Nonconsenting, uninformed members of the population. It might be *preferable* to avoid harm brought about at the hands of medical professionals for some of the reasons I have already covered, even significantly preferable, but it is by no means the only factor at play. As I argued in the ninth section, when considered alongside the above disparities in number of infected individuals, knowledge, and consent, the consideration of whether harm is brought about by medical professionals directly or indirectly is ethically negligible.

Argument from Hindsight

We've already approved several vaccines that are already being distributed to the population, so what's the point of conducting challenge trials?

At the time of writing this paper, there are several approved SARS-CoV-2 vaccines in the process of being distributed to the American population. That being said, vast numbers of Americans continue to go about their lives unvaccinated and at high risk for contracting COVID. Approving more vaccines more quickly will undoubtedly expedite this distribution process in the way I outline above. Additionally, this argument is tinged with privilege rooted in membership in a country rich enough to secure enough of the few approved vaccines for its population. The notion that further vaccine development is no longer an urgent issue now that the rich countries have secured their shares neglects the great majority of people whose vaccinations depend on an expedited approval process.

Conclusion

Even if the COVID-19 pandemic were entirely behind us, that does not alter the significance of this argument. There has potentially been a *catastrophic institutional failure* responsible for hundreds of thousands of *preventable* deaths, and this is by no means the last pandemic. In fact, future pandemics are only becoming more likely due to factors like climate change and increased human-animal contact.[13] Failure to exercise the most efficient method of vaccine development during this pandemic should not serve as precedent for similar inaction during the next one. What must follow are immediate large-scale examinations and subsequent reformations of our institutional medical ethos. Only upon doing so will we be equipped to avoid future reiterations of present circumstances.

Acknowledgements

Thank you to Jackson Carter, Evan Caulfield, David Denby, Isaac Jacobowitz, and Arno Min for their keen comments on a draft of this essay, and to Jody Azzouni and Cassidy Prejean for their always stimulating discussions of these issues and others.

[13] "5 Reasons Why Pandemics like COVID-19 Are Becoming More Likely."

Bibliography

Bourget, David, and David J. Chalmers. "What Do Philosophers Believe?" *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition* 170, no. 3 (2014): 465–500.

Broockman, David, Joshua Kalla, Alexander Guerrero, Mark Budolfson, Nir Eyal, Nicholas P. Jewell, Monica Magalhaes, et al. 2020. "Broad Cross-national Public Support for Accelerated COVID-19 Vaccine Trial Designs." OSF Preprints. July 15. doi:10.31219/osf.io/au3fq.

Bult, Laura. "The risky way to speed up a coronavirus vaccine." Vox Media Studios, August 12, 2020, Youtube video, 9:12, https://www.youtube.com/watch?v=r2fxJI_cP58.

Cloud, John. "Would You Kill One Person to Save Five? New Research on a Classic Debate." *Time*, December 5, 2011. https://healthland.time.com/2011/12/05/would-you-kill-one-person-to-save-five-new-research-on-a-classic-debate/.

Eyal, Nir, Marc Lipsitch, and Peter G. Smith. "Human Challenge Studies to Accelerate Coronavirus Vaccine Licensure." *The Journal of Infectious Diseases* 221, no. 11 (March 2020): 1752–56. https://doi.org/10.1093/infdis/jiaa152.

Gavi, the Vaccine Alliance. "5 Reasons Why Pandemics like COVID-19 Are Becoming More Likely," June 10, 2020. https://www.gavi.org/vaccineswork/5-reasons-why-pandemics-like-covid-19-are-becoming-more-likely.

History of Vaccines. "Vaccine Development, Testing, and Regulation." The College of Physicians of Philadelphia, January 17, 2018. https://www.historyofvaccines.org/content/articles/vaccine-development-testing-and-regulation.

McNeil, Donald G., Jr. "They Swallowed Live Typhoid Bacteria — On Purpose." *The New York Times*, September 28, 2017, sec. Health. https://www.nytimes.com/2017/09/28/health/typhoid-vaccine-trial.html.

Mill, John Stuart. *On Liberty*. Cambridge Library Collection - Philosophy. Cambridge: Cambridge University Press, 2011. doi:10.1017/CBO9781139149785.

Singer, Peter. "Famine, Affluence, and Morality." *Philosophy & Public Affairs* 1, no. 3 (1972): 229-43. Accessed March 17, 2021. http://www.jstor.org/stable/2265052.

Tambornino, Lisa, and Dirk Lanzerath. "COVID-19 Human Challenge Trials – What Research Ethics Committees Need to Consider." Research Ethics 16, no. 3–4 (July 2020): 1–11. https://doi.org/10.1177/1747016120943635.

Treatment Action Group. *AVAC and TAG Statement on Ethical Conduct of SARS-CoV-2 Vaccine Challenge Studies*, May 7, 2020. https://www.treatmentactiongroup.org/statement/avac-and-tag-statement-on-ethical-conduct-of-sars-cov-2-vaccine-challenge-studies/.

ISSUE 4
WINTER 2021