The *Proceedings of the Southeast Philosophy Congress* is a publication of the Philosophy program at Clayton State University, collecting original undergraduate, graduate and postgraduate research presented at our annual Congress, hosted by Clayton State University. Authors retain full rights of publication to their materials.

The 2009 keynote speaker was Jack Zupko from Emory University.
# TABLE OF CONTENTS

## KEYNOTE ADDRESS

Jack Zupko (Emory University)  
**Ten Myths about Medieval Philosophy** .................................................. 1-15

## PAPERS

Jonathan Arthur (Virginia Commonwealth University)  
**External Sufficient Conditions and Moral Responsibility** ......................... 16-23

Charles Bauch (University of West Georgia)  
**A Phenomenological Understanding of the Rosary**  
**Considering Heidegger's Conception of Worldhood** ............................. 24-31

Jesse Christiansen (Georgia State University)  
**A Defense for Apriority in Naturalized Epistemology** ............................. 32-45

Eli Cohn (Lewis and Clark College)  
**Externalists on Closure** .......................................................................... 46-58

Daniel Crescenzo (University of Georgia)  
**Which Motivation is Valuable?: A Critique of Zagzebski's**  
**Theory of the Value of Knowledge** .......................................................... 59-68

Michael Granado (Georgia Southern University)  
**G. E. Moore Against the Skeptics** ............................................................. 69-73

Chad Hale (University of South Florida)  
**Doing the Will of Beliefs: A Critical Assessment of William**  
**James' Defense of Religious Faith in “The Will to Believe”** .................... 74-88
Shehnaz Haqqani (Clayton State University)
Our Lack of Free Will Due to God's Omniscience.................................89-94

Joshua Horn (University of Kentucky)
Kant's Political Philosophy: An Evolution of Social
Contract Theory..........................................................................................95-103

Christa Johnson (Ursinus College)
A Fresh Take on the Non-Identity Problem...........................................104-113

Maggie Ann Labinski (Loyola University Chicago)
Illuminating Education: Augustine’s De magistro..................................114-126

Sarah McCulley (Ohio University)
Davidson’s Semantical Primitives and Why They
Cannot Work..................................................................................................127-134

Chelsea Ruxor (University of Evansdale)
Externalist Foundationalism........................................................................135-140

Nathan Sheck (Liberty University)
The Varieties of Jamesian Interpretations:
Understanding “The Will to Believe” in Context......................................141-153

Matt Schneider (University of Georgia)
Self-Ownership, Freedom and Responsibility:
An Argument for an Enforceable Obligation to Help............................154-161

Scott Syzmanski (Georgia Southern University)
Justice: Glaucon’s View and Possible Objections.................................162-166

Michael Uhall (University of Georgia)
A Principle Objection: H. P. Grice and the Failure of
Cooperation....................................................................................................167-171
Stijn Van Impe (Ghent University)
The Importance of “Communication” in Kant Reconsidered........172-179

GENERAL BIBLIOGRAPHY
It is surely a sign of age when a professor feels inclined to list the most common myths or misconceptions about his own field of study. But since I am in fact ‘getting up there,’ and thought it would be fun to do, I decided to give it a try.

I was first exposed to medieval philosophy about thirty years ago, in a class I took as a college sophomore from the great historian of post-medieval logic, E. Jennifer Ashworth – though I no idea of her reputation then. I ended up taking several more courses from her, including what amounted to one-on-one tutorials, and writing an Honors Thesis on the logic of William of Ockham. Jenny sent me on to graduate school at Cornell, where I studied with Norman Kretzmann and wrote a dissertation on the philosophical psychology of the fourteenth-century Parisian philosopher, John Buridan. I started my teaching career twenty years ago, first at San Diego State University, and then at Emory University, where I have been a member of the philosophy department since 1995.

Fairly often during my teaching career, I have had to disabuse people of mistaken ideas they had formed about medieval philosophy or the Middle Ages more generally. Where they got these ideas is not always easy to say, but the culprit is often popular culture, especially movies, which caricature the Middle Ages as a period of soot and ignorance between the glories of ancient Rome and the sparkling humanist culture of the Renaissance – with a few characters like Robin Hood (who probably did not exist, by the way) thrown in to spice things up. Unfortunately, our own histories of philosophy are also to blame, tilted as they are toward certain paradigm figures or texts, and to questions that interest us today that turn out to have been hardly discussed at all by medieval philosophers.

So, in the spirit of the Discovery Channel, here is my top-ten list of myths of medieval philosophy, or lies our philosophical forebears told us.

1. **Medieval philosophers were more interested in preserving the wisdom of ancient thinkers than in advancing their own ideas.**

Like most myths, this one has an element of truth. It is true that medieval philosophers spent (what seems to us) an inordinate amount of time thinking and writing about old texts they viewed as authoritative (Lat. = auctoritates). But their notion of authority needs to be properly understood. These were not texts they thought one must agree with no matter what, or authors whose words were cited to shut down all objections and silence further discussion. Rather, authorities were seen as a source of philosophical questions and problems unto themselves, as well as offering the basic parameters of their solution.

Aristotle is a good example. Aristotle was commonly referred to as ‘the Philosopher’ and his works on logic, metaphysics, natural philosophy, and ethics were the foundation of the
undergraduate curriculum at medieval universities. But even though his books were studied by undergraduates and commented upon by scholars, everyone who knew them also knew that there were serious problems in his thought. Usually they tried to solve them in Aristotelian terms, but they were not always successful, and when they weren’t they would try alternative approaches. Some of these difficulties are well known to us – like the phenomenon of *akrasia* or weakness of will in Aristotle’s ethics or the problem of self-knowledge in his psychology – whereas others are not much discussed anymore – like the incompleteness of his account of modal syllogisms or his handling of extra-categorial or transcendentals concepts like being, truth, and goodness. I can discuss only one example here, but it represents nicely, I think, the attitude medieval philosophers had toward Aristotle’s authority.

In Book VIII of the *Physics*, Aristotle rolls out the theory of antiperistasis or ‘mutual replacement’ to explain projectile motion. According to this theory, the reason a thrown object like a javelin continues to move after it has left the thrower’s hand is that the air being pushed by the front of the projectile rushes back to fill the void space created by its movement, this rush of air being sufficient to move the javelin along (nature abhors a vacuum, of course). Now this has more than a whiff of implausibility, and Aristotle, I think, is far from happy with it, but it is the solution he offers, after rejecting several others.¹ So did medieval thinkers accept it just because Aristotle said it? Some did, although not for that reason. Roger Bacon and Thomas Aquinas, for example, both defended the Aristotelian idea that the medium is what imparts movement because only the medium (the air, in this case) is in contact with the projectile, and we can’t have action-at-a-distance (the thrower cannot continue to force the javelin along when he or she is no longer touching it). But like Aristotle, they did so only after finding alternative explanations even more implausible.

John Buridan (c. 1300-61) was among those who rejected Aristotle’s theory. You can tell he was dealing with an authority because his opening remarks on the problem understate Aristotle’s shortcomings: “The question [of projectile motion] I judge to be very difficult because Aristotle, as it seems to me, has not solved it well.”² He then offers a set of quaint but devastating counterexamples based on common sense and ordinary experience: (i) ‘mutual replacement’ might account for the rectilinear motion of an arrow or javelin, but it cannot explain the spinning motion of a top because a spinning top remains in one place – its motion creates no empty space that would cause the necessary rush of air; (ii) we observe that a javelin with points on both ends travels just as swiftly as a javelin with only one pointy end, but this shouldn’t happen because, Buridan says, “the air would be easily divided by the sharpness” at the back end of the two-pointed javelin, and the air pushing it would do so less efficiently than if it were flat or fletched with feathers; and (iii) why

---


wouldn’t a cargo ship piled high with grain or straw have its cargo blown forward by the
same rush of air that comes up behind the ship, pushing it along?²

To solve the problem, Buridan famously offers the alternative theory of impetus or
‘impressed force,’ a quality imparted to the matter of the projectile by the thrower, the
strength of which determines its velocity. Historians of science have shown that the
theory of impetus was crucial for breaking the hold of the Aristotelian axiom that
inanimate objects, i.e., bodies without souls, cannot move themselves, thereby paving the
way for Galileo and the modern science of mechanics. Buridan, it must be emphasized, did
not anticipate Galileo. But he did remove certain key obstacles to Galileo, as well as
cleverly exploiting some of the consequences of his new theory.³ Thus, he noted,
cosmologists could get rid of the clumsy assumption that celestial bodies are moved by
intelligences which somehow inhabit them (again, Aristotle assumed that only living
bodies, or bodies with souls, can move themselves) – something, Buridan hastens to add,
that is not stated in the Bible. Instead, he wonders if the movements of celestial bodies
might be due to God’s having given each of them their own impetus at creation,⁴ a more
parsimonious explanation that also better expresses the simplicity of divine agency.

One can multiply examples here, but the point is that everyone saw there were serious
problems with what Aristotle had said. So, despite the place of Aristotle (and Cicero and
Seneca and Boethius and numerous other ancient authors) as authorities in medieval
philosophy, their authority was hardly absolute. There were of course encyclopedists
whose job it was to collect, distill, and arrange the wisdom of ancient authors into
textbooks that were studied in school, but the encyclopedists were not philosophers.
Philosophers saw ancient authorities as raising serious questions demanding serious
answers, which they were only too happy to give, even – or perhaps especially, as we have
just seen – when their answers diverged from what the authorities themselves said.

2. Medieval philosophy is basically just rational theology.

I sometimes call this the ‘equivalence myth.’ It is also partly true, but the sense in which it
is true matters crucially. The classic picture is that medieval philosophy is an amalgam of
ancient Greek philosophy and religious doctrine, whether Christian, Islamic, or Jewish.
Philosophy was used to rationalize faith insofar as that is possible, but it remained an

³John Buridan, Ioannis Buridani subtilissime quaestiones super octo physicorum libros Aristotelis
(Paris: 1509), rpr. as Kommentar zur Aristotelischen Physik (Frankfurt a. M.: Minerva, 1964); Book VIII,
Question 12.

⁴The theory itself had important antecedents in the sixth-century Aristotelian commentator
Philoponus and the eleventh-century Islamic philosopher Avicenna or Ibn Sina, as well as in the thirteenth-
century mathematician Jordanus de Nemore and the fourteenth-century theologian Francis of Marchia. But
the exact chain of transmission is unknown because the key portions of earlier texts were apparently
unavailable to later authors. For the details, see Clagett 1959: 505-525.

⁵Clagett 1959: 536.
ancillary or handmaiden science, always at the service of theology and never free to find truth on its own. On the positive side, it is often said that what is good and distinctive about medieval philosophy emerged from this partnership. Thus, as Frederick Copleston wrote in his magisterial history of philosophy, published just over fifty years ago and still widely read today, “the great achievement of the thirteenth century in the intellectual field was the realization of a synthesis of reason and faith, philosophy and theology.”

Copleston was surely thinking about Thomas Aquinas when he wrote this (more about Thomas in a moment), and it is true that if your goal in the intellectual field is to synthesize faith and reason, you can’t do much better than the Angelic Doctor. But there were many, many other medieval thinkers who didn’t contribute to this project either because they were interested in other questions or because they didn’t think it was possible. An enormous quantity of philosophical material was produced in medieval universities by the ‘Artists [Artistae]’. These were scholars who worked in the undergraduate faculty of arts, and who were forbidden by statute from teaching theology, which was considered a graduate subject. The primary philosophical genre for these scholars was the commentary, which they developed mostly from lectures on the books of Aristotle that were required reading in the arts curriculum. To modern ears, ‘commentary’ suggests something derivative and unoriginal. But commentaries came in several forms. In addition to literal commentaries (expositiones), where the master would subdivide the text being studied and give line-by-line explanations of the author’s meaning, there were also question-commentaries (quaestiones), where the master was free to riff on philosophical questions he took to be raised by the text being studied. Many commentators on Aristotle’s De anima, for example, were prompted by his remark that knowing the accidents or properties of a thing contributes to knowing its essence (De anima I.1.402b22) to wonder how this is possible if substances and accidents are different kinds of thing. The question is clearly epistemological because it raises skeptical doubt about the quality of our inferences in the science of psychology. One influential commentator, Nicole Oresme (1320-82), points out that it must be due to a naturally necessary (rather than logically necessary) connection between the soul and its accidents that we can, “from a cognition of those accidents, arrive a posteriori at a determinate and quidditative, i.e., definitive, conception of its substance.” The discussion of this problem goes beyond anything in Aristotle and in some ways anticipates the interest in knowledge and certainty typical of early modern philosophers such as Descartes. And it was all discussed independently of theology.

There are untapped riches in the non-theological genres of medieval philosophy, and by this I mean texts that are still unedited, untranslated, and otherwise unknown to the general philosophical public. As these become more available, the equivalence myth will

---


continue to dissolve and be replaced by a much more diverse picture of what medieval philosophy was all about.

3. Medieval philosophy is obscure and overly-technical.

To my mind, this is one of the worst libels of medieval philosophy, but it is hard to expose because technicality is somewhat in the eye of the beholder. But more often than not, it is a judgment based on ignorance, so let me try at least to expose the ignorance.

This is the sort of judgment expressed by William James (who should have known better) when he said, “Scholastic philosophy ... is only common sense grown pedantic.”8 Now I think we can all understand the concern here. Who wants to be pedantic, if by that we mean being doctrinaire about technical knowledge, or engaging in formalism for its own sake – or (worse) for the sake of showing off one’s knowledge to others? This is the very opposite of philosophical. The question is whether medieval philosophers – or the 'scholastics', as James calls them – are guilty of it.

They are not. But I need to say first that this libel has a long history, beginning with the Renaissance humanists, who wanted to move philosophy away from the kind of formal disputation then in vogue in the schools to a new foundation based on authors and texts from classical antiquity, many of which – like Plato's dialogues – had been newly translated from the original Greek. Its perpetrators were many, and the attacks became much worse after the Reformation, when scholastic disputation and especially the study of Aristotelian logic were seen as activities engaged in by those sympathetic to Rome. Thus, we see Francis Bacon (1561-1626) calling his treatise on method the 'New Organon,' a rather unsubtle swipe at the old Organon, whose author, Aristotle, is then described as someone who “utterly enslaved his natural philosophy to his logic, and made it a matter of disputation and almost useless.” Aristotle, in Bacon’s estimation, “spoils his natural philosophy with his dialectic.”9

The first thing to say to this is that despite surface appearances, medieval logic was never more than an organon, i.e., a set of tools used to solve philosophical problems. Now medieval philosophers did think about logic, which typically meant that they pursued logical theory in certain specialized genres designed to test their 'analytical mettle,' if you will. But anyone who reads deeply enough into these genres – such as the tradition of sophismata, where we find detailed analyses of sentences presenting certain logical and semantic conundrums – can see that the point was to use one’s logical toolkit to solve problems in metaphysics, natural philosophy, and ethics, not to play with the tools

---


themselves. If you want good solutions, attention to technique is both natural and inevitable.

As a result, medieval logic is filled with sophisticated argumentation, fine-grained analyses, and genuine innovations that are recognizable even to philosophy students today. But Renaissance humanists hated scholasticism so much that they threw the baby out with the bathwater. That is, in their haste to break with the preceding epoch (and return to an earlier one), they forgot to see if there was anything worth keeping from the schoolmen. This led to some terrible losses for western philosophy.

The most glaring example here is the theory of logical consequence, or what it is that makes one proposition follow from another. I won't get into the details, but it was noticed in Greek antiquity that Aristotle’s system of logic based on the syllogism did not work very well as a method of proof, especially in connection with the analysis of propositions containing modal concepts. So the Stoics came along and made Aristotelian logic obsolete by devising a new system, propositional logic, which seemed both theoretically prior to syllogistic logic and capable of proving more complex positions. Unfortunately, the key Stoic texts were lost with the demise of the ancient world, where 'lost' means that they were not copied and preserved, or they were destroyed during the barbarian invasions, or simply that they were in Greek and nobody could read them. So medieval philosophers did their best with Aristotle, whose works survived and were translated into Latin. Logic was done with syllogisms until the early fourteenth century, when cracks began to emerge in the Aristotelian system in same places where the Stoics found them, i.e., in the theories of demonstration and modality. So a new theory, the theory of consequences, was developed by fourteenth-century logicians such as William of Ockham (1285-1347), Walter Burley (1275-1344), and John Buridan. These philosophers were called terminists because they thought logic should be based on the properties of terms, which are the subjects and predicates of propositions. The result was a much more powerful system for proving propositions and conceptualizing truth and validity.

---

10Modal propositions, as contrasted with assertoric propositions, are those which qualify what they say along certain temporal and/or natural lines: e.g., ‘Socrates is possibly running’; ‘Dogs are necessarily carnivorous’.

11The technical notion of consequence is first found in texts associated with the Stoic-Megarian School, in their treatment of conditionals, or ‘if … then’ propositions. Aristotle comes very close to the non-technical sense of consequence in his discussion of syllogistic inference in Prior Analytics II.2-4, but he takes it all back in Posterior Analytics II.11: “the inference from antecedent to consequent does not hold if only one premise is assumed – two is the minimum” (94a25-26). We might think of him as being on the brink of the discovery of consequences, held back only by his determination to evaluate arguments by reducing them to their tripartite syllogistic forms.

12Etymologically, ‘terms (termini)’ are the ‘ends’ or ‘limits’ of propositions, so that a simple assertoric proposition, ‘Socrates is running’, would have two terms, i.e., its subject, ‘Socrates’, and predicate ‘is running’. Later terminist logicians usually tried to show how the old system of syllogistic logic could be subsumed by the theory of consequences, a move which enabled them to keep Aristotle as a logical authority.
But when the study of logic fell out of fashion during the Renaissance, the medieval theory of consequences suffered the same fate as the Stoic theory. It had to be reinvented a second time, by twentieth-century logicians such as C. I. Lewis (1883-1964), who needed it to help refine the concept of implication.\textsuperscript{13}

The final irony here is that the Middle Ages was long enough and diverse enough to contain periods intellectual historians have dubbed ‘humanist’. I refer of course to the revival of learning we now know as the twelfth-century renaissance, when poetry and the study of literature thrived. Thankfully, unlike the next renaissance three hundred years later, this was not at the expense of logic.

4. Medieval philosophers held mistaken beliefs about the natural world.

We might call this ‘the Copernicus criticism.’ Yes, it is true that medieval philosophers had many false ideas about the way the world works, foremost among them the notion that the universe is geocentric, with the earth at its center. They got this idea from Ptolemy and ancient Greek cosmology. But here it needs to be said that this was hardly a case of culpable ignorance. From an evidential point of view, the geocentric and heliocentric models of the universe are tied if one relies on naked eye observations, and medieval philosophers lacked a key scientific instrument, the telescope, which allowed Galileo to break the tie. We might even say that medieval philosophers were justified in believing that the earth is the center of the universe, although they did not know it, of course, because you cannot know what is false.

But it would be a mistake to take any comfort in this, especially if it takes the form, ‘... but now we know better.’ Empirical science requires that we test our assumptions about the way the world works and to revise or abandon them should they be disconfirmed. And it is not as if this has been happening less often over time. The history of science is littered with paradigms and theories that were once dominant, but which we now know to be wrong, from phlogiston and phrenology to vitalism and the independence of space/time. The fact is that we do not know any better than medieval philosophers which of our deeply-held scientific beliefs is false, and that is because justification does not entail truth. Some radically empiricist philosophers of science have argued that our epistemic position now is not any better than it was in earlier times, just different. I would not want to take it this far, but it does suggest restraint when we evaluate belief-systems different from our own.

Here again, ignorance is our own worst enemy. According to a popular misconception, people in medieval times believed the earth is flat. Let me set the record straight by telling you that no one in the Middle Ages believed the earth is flat – certainly no educated person, and probably no one else either. The ancient Greeks knew that the earth is spherical: Eratosthenes of Alexandria (c. 276-192 B.C.) had actually calculated its circumference and came surprisingly close to the actual figure of just over 40,000 kilometres. Christian

theologians and the Church Fathers – Origen, Athanasius, and Augustine – all knew that the earth is spherical. There are numerous allusions in the Bible to the earth being spherical. But there is another, even more powerful reason: the roundness of the earth can be seen with the naked eye. I spent a lot of time growing up on a farm on the Canadian prairies, and I remember how on certain days you could see the clouds taper off before finally meeting the horizon (they would never meet the horizon, of course, if the earth was flat). The poet Kathleen Norris once described the undulating cirrus clouds in the South Dakota sky as if she were looking up at the ribs of a leviathan. This is surely what we mean when we talk about the 'vault of heaven'.

So where did the myth come from? One of the main sources was the nineteenth-century American writer, Washington Irving (1783-1859), who in his fictional biography, *The Life and Voyages of Christopher Columbus* (1828), depicted Columbus as trying to convince the Spanish court that if he sailed too far to the west he would not fall off the edge of the earth (in actual fact, the king’s advisors knew the earth was spherical but also very big, which would have made an expedition of the sort Columbus was proposing very expensive). Nevertheless, Irving’s biography was hugely popular, and a myth was born. Some decades later, the flat-earth myth was borrowed by Darwinists who wanted to tar their Christian opponents with it. It is a myth that from the beginning was used to libel one’s opponents as ignorant or stupid.

Unfortunately, the flat-earth myth tends to overshadow some of the really interesting false beliefs medieval philosophers had about the natural world. William of Conches (c. 1090-1154), for example, was not alone in arguing that the earth (which he believed to be spherical) had two temperate or habitable zones, one in the northern hemisphere and one in the southern hemisphere, separated by a torrid zone along the equator through which no living thing could pass. This meant that there might be human inhabitants in the southern hemisphere, though we could never meet them because crossing the torrid zone would mean certain death. William himself believed that humans lived only in the northern temperate zone. The idea of climatic zones came from Stoic and Neo-Platonic cosmologists of late antiquity, and was picked up by the fifth-century Roman grammarian Macrobius, who transmitted it to Latin posterity in his widely-read commentary on Cicero’s *Dream of Scipio*. It was the dominant theory of climate until the early thirteenth century, when Aristotle’s books of natural philosophy – *Physics, On the Heavens and Earth, Meteorology*, etc. – replaced Plato’s *Timaeus* (of which Latin-speakers had only the first half) as authoritative in natural science. The *Timaeus* then enjoyed a second career as an inspirational source-text in poetry and literature. That is another story, but it illustrates the sometimes wholly accidental (and entirely interesting) way in which medieval thought was shaped by which texts from Greek and Roman antiquity were read and happened to survive.

---

14 See Job 26:7, 26:10; Isaiah 40:21-22; Proverbs 8:27.

5. Medieval philosophers were tortured or burnt at the stake if they said the wrong thing.

This myth persists because it is part of our own mythology of academic freedom. Scholars and scientists had it tough in centuries past. Think of what Galileo endured for daring to pronounce that the earth is not the center of the universe. In our own, more enlightened, age, we have seen fit to give professors tenure because a just society values the pursuit of truth and the ability of those who find it to pronounce it, without fear of reprisal.

The only problem with this is that medieval philosophers did not live in fear of being tortured or burnt at the stake if they said the wrong thing. They might have feared being publicly refuted at one of the disputations or formal debates that were held at the universities, but that is surely evidence of the existence of free inquiry, not a strike against it. Occasionally someone (usually a theologian) was investigated on suspicion of heresy, but this was a formal procedure in which the views of the accused were carefully examined by a panel of experts, who advised the authorities on the proper course of action. I stress here that it was always the views, not the person, which were investigated. An author could always testify in his defense, providing an *apologia* or explanation of his views, but there are many cases where the investigation was posthumous, in which case the views were examined on their own merits. If any views needed to be condemned, they would be clearly set out in an official document with the reason(s) for each error fully explained. The author would then be asked (not forced) to recant. There were consequences if he did not, such as losing his teaching position in the university, which might lead to self-imposed exile as he looked for other means of supporting himself. But no one was tortured or threatened with death if he didn’t recant. If the error was especially pernicious, such as the attack on Aristotelianism by the Paris theologian Nicholas of Autrecourt in the 1340s, an author might be required to burn his works in public. Nicholas did so, and was made Dean of Metz Cathedral three years later – not a bad landing for a disgraced academic.\footnote{For background on this and other condemnation episodes, see J. M. M. H. Thijssen, *Censure and Heresy at the University of Paris, 1200-1400* (Philadelphia: University of Pennsylvania Press, 1998).}

It would be wrong to say that torture and burning at the stake never happened. But these were exceptional events. Typically (and in a way that should resonate with our post-911 world), tolerance for disagreement diminished when the Church felt it was under serious threat – a kind of medieval *Code Orange*. Heresy was the main culprit here, although uprooting heresy was never just a matter of getting people to change their beliefs to the correct ones. Heretical sects had their own bishops, churches, and material wealth, so much so that they stood as a kind of parallel authority to the pope. By the late twelfth century, for example, the Cathar heresy had infested most of southern France, where it became part of the local culture, pitting secular rulers and armies against each other. Thousands were tortured and killed in the subsequent Albigensian crusade, during which Pope Gregory IX brought the office of inquisition under the control of papacy itself and entrusted its operations to the Dominicans. Those accused of heresy were given the chance
to recant; if they did not, they were usually handed over to civil authorities and either hanged or burned at the stake. Not pleasant, but then not uniquely medieval either.

I know of no medieval philosopher who was tortured or killed by civil authorities for his views, although this did happen in the Counter-Reformation, which is often confused with the Middle Ages. Everyone thinks of Giordano Bruno being burned at the stake for his heterodox theological beliefs and stubborn defense of Copernicanism, but note that this happened in 1600, in the lifetime of Descartes. It is during the Counter-Reformation that we find the first examples of hysteria over witches (those accused were usually hanged rather than burned) and the use of technology such as the rack to extract confessions from one’s religious opponents. I believe the confusion here is partly due to the tendency of certain reformers to lump together everything Catholic as evil, a move that indiscriminately swept up the whole of the Middle Ages with the Renaissance Church as idle and corrupt. We saw above how this tendency resulted in tragic losses of philosophical learning.

6. Medieval philosophy reached its zenith in the thought of Thomas Aquinas.

This is a myth that can still be found in some older histories of philosophy. The basic picture is of medieval philosophy climbing a long and difficult road from virtually nothing at the Fall of Rome to the Golden Age of Scholasticism in the thirteenth century, finally reaching its summit in the brilliant theological synthesis of Thomas Aquinas. Thereafter, it went into decline as the Church was beset by internal division as well as struggles with secular rulers, while philosophers were afflicted by nominalism and skepticism and their own divisions into rival schools of thought. This led to the disaster of modernism and the Reformation.

I caricature here – but only slightly. In terms of the historiography of medieval philosophy, its advent corresponds suspiciously to the 1879 publication of Aeterni Patris, the encyclical of Pope Leo XIII which enlists Thomas Aquinas in the cause of ‘restoring’ Christian philosophy to serve as a bulwark against the deadly influence of modernism. Here is what Pope Leo says about Thomas:

Among the Scholastic Doctors, the chief and master of all towers Thomas Aquinas, who, as Cajetan observes, because ‘he most venerated the ancient Doctors of the Church, in a certain way seems to have inherited the intellect of all’ [34]. The doctrines of those illustrious men, like the scattered members of a body, Thomas collected together and cemented, distributed in wonderful order, and so increased with important additions that he is rightly and deservedly esteemed the special bulwark and glory of the Catholic faith. With his spirit at once humble and swift, his memory ready and tenacious, his life spotless throughout, a lover of truth for its own sake, richly endowed with human and divine science, like the sun he heated the world with the warmth of his virtues and filled it with the splendor of his teaching.
There is no doubt that Thomas Aquinas was a brilliant philosopher and that he is most deserving of his place in the pantheon of great western thinkers. But that is not quite what Pope Leo is saying here. His real purpose is revealed in the next section of the encyclical:

Moreover, the Angelic Doctor pushed his philosophic inquiry into the reasons and principles of things ... and as he also used this philosophic method in the refutation of error, he won this title to distinction for himself: that, single-handed, he victoriously combated the errors of former times, and supplied invincible arms to put those to rout which might in after-times spring up.

Thomas Aquinas was a Dominican friar who did his fair share of refuting heretics and combating error. This was especially true early in his career, when he took very seriously the Dominican charge to drive the Cathars from France and northern Italy. But he did not do so single-handedly, and it is certainly not all that he did. Pope Leo, however, saw in Thomas just the kind of intellectual ally the Church needed to combat modernism and its associated heresies: liberalism, which threatened the political order; secularism, which deprived the Church of any real influence in worldly affairs; individualism, which made faith a matter of conscience rather than obedience to Church teaching and tradition; and (later), Darwinism, which stripped humans of their special place in creation, and indeed, removed God from creation entirely. Catholic intellectual history since Pope Leo has been dominated by the effort to make good on this medieval allegiance.

What was the actual authority of Thomas Aquinas during the later Middle Ages? Certainly he was an influential figure, but unsurprisingly, his reputation waxed and waned just like any other major philosopher or theologian. As my former colleague Mark Jordan once put it, Thomas had not been dead five years before people began disagreeing over what he had said. The Condemnation of 1277 by the Bishop of Paris defines as erroneous more than a dozen positions that had been defended by Thomas himself, and Thomas was routinely criticized in the theological writings of members of the rival Franciscan order. The most famous of these was the Correctorium fratis Thomae (The Correction of Brother Thomas) by the Franciscan, Walter de la Mare. The 1277 condemnation of Thomas's positions was revoked before he was canonized in 1325, but even then, his authority was felt mostly in Dominican churches and religious houses. This is the origin of the Thomist school of philosophy (as opposed to the Neo-Thomist, which arose after Pope Leo’s encyclical). But what needs to be remembered is that during the Age of the Schoolmen, the Thomist school was not the only school. There were also Scotists, Ockhamists, Nominalists, Terminists, Albertists, and others, each with their own idea of how philosophy should be done – hence the period known as the ‘Wegestreit,’ or battle of the ways. Schoolmen deliberately took up opposing positions. A standard school exercise was to have a student defend a position in the manner of a Thomist, or Scotist, or Nominalist, and so on.

What we see in the later Middle Ages, then, is a crazy quilt of rival ideas and people who passionately defended them. What we do not see is philosophical life being dominated by any particular group or thinker.
7. Medieval philosophy is identical with the thought of the Catholic Church.

This is a hybrid of the two previous myths. Following Pope Leo's encyclical elevating Thomas to his preeminent position among Doctors of the Church, Catholics became less inclined to see beyond the Angelic Doctor when they studied philosophy whereas Protestants were more inclined to see Thomas as the lone spokesman for the sorry epoch of human history between the Fall of Rome and the Reformation. Such sectarian motives are less evident today, but we can see some of the fallout in the way the canon of western philosophy has developed. Just look at any introductory text in philosophy or Ph.D. reading list. Thomas Aquinas is the paradigm medieval philosopher and the paradigm work of medieval philosophy is always his *Summa Theologiae*, Thomas’s textbook for Dominican novices that Pope Leo wanted to make the blueprint for Catholic and Christian philosophy. Pope Leo ordered the preparation of the critical edition of Thomas’s works (now known as the Leonine edition, 1884-, and still in progress), which led to a cottage industry of translating, explicating, and anthologizing his writings. For the longest time, the only works of medieval philosophy that existed in English translation were those of Thomas Aquinas; if you wanted to find out about anyone else, you had to read Latin. At the same time, Philosophy departments at Catholic universities (and only Catholic universities) became avowedly Thomist, as the study of other scholastic thinkers and works gradually faded away.

This foreshortening of our historical horizon was a fairly recent phenomenon. But it doesn’t take much reflection to see that medieval philosophy is as different from modern Catholic thought as the pre-Tridentine Mass is from the post-Vatican II Mass. The situation for non-specialists is much improved today thanks to the appearance of editions and translations of works by other eminent medieval philosophers. Here I will mention only the multi-volumed *Cambridge Translations of Medieval Philosophical Texts* (1988- ), which has brought to the attention of philosophy teachers some of the incredible riches of medieval logic, metaphysics, ethics, and political philosophy.

8. Medieval philosophers were interested in proving the existence of God.

This is a common misconception among students of philosophy because of the way medieval philosophy is presented in philosophy textbooks, virtually all of which include Anselm of Canterbury’s so-called ‘ontological’ argument for the existence of God and the Five Ways of Thomas Aquinas. But the myth is fairly easy to debunk. Proofs for the existence of God are in fact unusual in medieval philosophy and theology, and even where we do find them they never occupy center stage. There is good reason for this, if you think about it. The existence of God was not controversial since virtually everyone was a theistic believer. Why would you want to demonstrate the existence of someone whose existence is not in question?

So what was Anselm’s problem? The actual role of the ontological argument in Anselm’s philosophical theology has been much discussed in the past fifty years. But no one thinks that the *Proslogion*, the text from which the argument is taken, is about proving the
existence of things foundationally, or in the face of skeptical doubt, like Descartes and so many other modern philosophers have tried to do. Rather, beginning with Karl Barth, a large scholarly consensus has formed around the idea that Anselm was seeking to express his belief in God rationally, like the way an artist might paint a devotional picture or a musician might write a hymn. This would be in keeping with Anselm’s adoption of the Augustinian credo, ‘Faith seeking understanding [Fides quaerens intellectum],’ and the fact that he was writing for his brothers in the monastery at Bec and not trying to convince pagans or atheists. The rest of the Proslogion makes good on this aim by working through the other divine attributes Anselm thinks can be proved by means of ‘necessary reasons.’

As for the Five Ways, when I teach them I am always struck by how flimsy, if not utterly fallacious, they are. Moreover, even if we grant the initial conclusion of each argument, they all contain a further, highly problematic identification step: ‘and this we call God.’ I would submit that Thomas knew very well that his Five Ways are not by themselves very convincing; at most, he shows how someone possessed of the principles of Aristotelian natural philosophy could use them to gesture towards the God of Abraham in the guise of the first mover, first cause, etc. But Thomas quickly moves on to other, much more important things. The Five Ways occur near the beginning of the First Part of the Summa Theologiae, which is not even its most important part. Rhetorically, this is a work that builds from the truths of philosophical theology presented in the First Part to the crescendo of moral theology in the Second. There is nothing to suggest that Thomas thought his work would stand or fall on his arguments for the existence of God.

We, however, are much more interested in arguments for the existence of God, since this is a live question in the post-modern world. In addition, we are much more likely to think that a philosophical work containing arguments for the existence of God stands or falls on the basis of those arguments. The only problem with this is that the medieval worldview was different, which makes focusing on such arguments a source of anachronism in our reading of medieval philosophical texts.

9. Medieval philosophers were insular and provincial.

This comes out of another generalization I sometimes refer to as the “Dark Ages” syndrome: the old canard that the Middle Ages was that long period of ignorance and superstition between the Fall of Rome and the Renaissance, when people did not have time to think or travel in their hard, subsistence lives between Viking raids, when everything would be destroyed and they would have to start all over again.

The grain of truth here is that there was a time in the Middle Ages when everyday life on the European continent was pretty grim, and it does correspond to the period when the Vikings were in ascendancy, from about the middle of the ninth century to the beginning of the tenth. But it is not true that intellectual life ceased altogether. One of the greatest of the early medieval philosophers, John Scotus Eriugena (c. 815-877), was active during this time, bringing his peculiar brand of Irish Neo-Platonism to the Carolingian court. And monasteries that were not raided continued their work of copying manuscripts and
teaching. Without their labors, very few literary works from classical antiquity would have survived for us to read today.

Furthermore, with few exceptions, medieval philosophers traveled extensively during their lifetimes and frequently wrote about their experiences. Around the beginning of the thirteenth century, there were two developments that greatly increased the portability of a scholarly career: the rise of the universities and the spectacular growth of two mendicant orders, the Dominicans and the Franciscans. The earliest universities at Paris, Oxford, Cambridge, and Bologna were international gathering-places of intellectual talent with a common father-language (Latin) and throngs of students willing to pay for the privilege of being tutored in the arts, law, medicine, and theology. And travel was the modus operandi of Dominican and Franciscan friars, as they preached and taught their way across Europe—monks without a monastery, as it were. Thomas Aquinas is an excellent example: born near Naples, he studied or taught at Rome, Paris, and Cologne, and frequently traveled on diplomatic and other special missions for his order. The career of John Duns Scotus followed a similar trajectory, but southwards: born in the far north of present-day England, he studied or taught at Oxford, Paris, and Cologne, and probably visited other cities as well helping to establish Franciscan custodial schools.

If anything, the medieval Islamic intellectual world was even more cosmopolitan. The fourteenth-century scholar and explorer Ibn Battuta (c. 1304-77) followed Muslim trade routes throughout the Islamic empire, from Marrakech to East Africa and India, possibly even reaching China. The chronicles of his travels were hugely popular, and remained so well into the twentieth century.

10. Medieval philosophers believed in the existence of unicorns.

This is one myth that might actually be true. Unicorns were frequently depicted in bestiaries, which were illustrated popular compendia of animals, but in this medieval authors were simply following their ancient Greek predecessors, who mentioned unicorns not in their mythologies but in their natural histories. The Greeks seem to have thought that unicorns existed, though one can easily surmise that they were dealing with third- or fourth-hand accounts of rhinoceroses. Be that as it may, unicorns were also rare and magical enough to appear frequently in medieval literature and art, which is the form in which they entered our modern cultural imagination.

The stock example of a possible but non-existent animal in medieval logic is the chimera or goat-stag. I had always assumed this was because medieval logicians thought unicorns really existed, though they could have picked up the chimera-talk easily enough from texts of late antiquity, where the chimera was also the paradigmatic non-existent. But it turns out that I was the one guilty of anachronism. Unicorns are paradigm non-existent in twentieth-century logical discourse. I believe the source here is Bertrand Russell, who writes in his Introduction to Mathematical Philosophy (1919): “Logic, I should maintain, must no more admit a unicorn than zoology can; for logic is concerned with the real world just as truly as zoology, though with its more abstract and general features. To say that
unicorns have an existence in heraldry, or in literature, or in imagination, is a most pitiful and paltry evasion.” The target of Russell’s attack is the Austrian logician Alexius Meinong, who wanted to attribute a level of being short of full existence to fictional entities like Hamlet so he could explain how we manage to conceive of and refer to them.

Ironically, Meinong’s ontological views are very close to, and may have been indirectly influenced by, one of the greatest metaphysicians of all time: John Duns Scotus (c. 1265-1308). What Duns Scotus thought of unicorns I cannot say. But I am sure he would have attributed some being to them if he thought about them at all – otherwise, we would be thinking about literally nothing, i.e., no thing. And unicorns are not nothing.

Conclusion

So there it is: ten myths and ten debunkings. Other historians of medieval philosophy might come up with a different list, or debunk them in a different way. But all of us have run into these ten at some time or other. I suppose if there is a lesson here, it is that as students of philosophy we need to read our historical texts carefully and reconstruct their arguments with a keen eye to context. We don’t have to, of course. We could instead use them as jumping-off points for our own musings about the perennial questions of philosophy – my fellow medievalist Paul Spade calls this the ‘Oh, that reminds me’ school of the history of philosophy. But we would miss a lot, and some of the things we think we see won’t really be there. That bothers me. I think it should bother you, too.

Jack Zupko
Emory University
Jonathan Arthur – External Sufficient Conditions and Moral Responsibility

In his essay, “Compatibilism” John Martin Fischer ponders what kind of control we need to have in order to be a source of our actions and thus morally responsible. When discussing sourcehood Fischer states that when those who think that determinism rules out morally responsibility talk about sourcehood, they refer to total-control sourcehood. Total-control sourcehood is where as Fischer describes it, “when the locus of control is ‘within him’ in a certain way”. More specifically, an agent has control over $X$ only if “for any factor $f$ which is a causal contributor to $X$ and which is such that if $f$ were not to occur, then $X$ would (or might) not occur, then $X$ has control over $f$.”\(^1\) This type of sourcehood says Fischer is “metaphysical megalomania”, a “total fantasy.”\(^2\) To see this one need only think about certain necessary conditions (such as the universe existing being a necessary condition for me writing this paper) to see that total control sourcehood is an impossibility. This reply, however, only deals with necessary conditions external to the agent. Fischer realizes that there may be an “important difference between our lack of control over external necessary or enabling conditions, and our lack of control over externally causally sufficient conditions.”\(^3\)

\(^1\)John Martin Fisher et al., *Four Views on Free Will* (Malden: Blackwell Publishing, 2007), 49.

\(^2\)Ibid., 67.

\(^3\)Ibid., 68.
In his discussion of necessary and sufficient conditions, Fischer presents us with the following diagram:

In the following essay I will address these two questions and argue that if we assume the truth of determinism, then we as agents ought to be very concerned with the horizontal line leading from A to BC. There is an important difference between the external necessary conditions and external sufficient conditions. This difference will show that causal determinism poses, pace Fischer, a grave threat to moral responsibility. Section one will examine the differences in the vertical and horizontal lines, Fischer’s two questions, and the implications the differences have on free will and agency (I will not be discussing the problem of activity, because I believe it falls with agency). Section two will examine what these differences mean for sourcehood, and moral responsibility.

I. Differences

First, we need to clarify background assumptions that I will be operating under. If one were to deny any of the following then my argument will not succeed. I will be operating under the assumption that determinism is in fact true. (That is, given a complete description of the history of the universe and any given time (H), and a complete system of physical laws (L), there is but one possible future). I will also assume the falsity of substance dualism.

---

4Ibid., 69.
Let us begin with the first of Fischer's two questions. What is the difference between the vertical and horizontal lines? It is clear that the difference is that the vertical line is the necessary conditions, and the horizontal line is the sufficient conditions. But this in and of itself does not tell us much. While I do not believe that I can highlight the differences in the constituent members of the two sets, I believe that via counterfactual assessment one can show that the two sets of conditions behave very differently. Consider a comprehensive list of causally contributive conditions for event (BC). The conditions that would not have obtained were (BC) not to have obtained are the sufficient conditions. The necessary conditions on the other hand must be in place for (BC) to obtain. Let us say that (BC) is an event, say of me going to a party. (S) is the set of sufficient conditions external to the agent (my self) for me going to the party (BC), and (N) is the set of necessary conditions for me going to the party (BC). Logically, we could re-write this as (S) → (BC), and ¬ (N) → ¬ (BC). Note that ¬ (BC) does not entail ¬(N), but by modus tollens, ¬(BC) entails ¬(S). What can we glean from this? It shows us that whether or not I go to the party (BC), (N) remain unchanged. On the other hand if I do not go to the party ¬(BC), then the sufficient conditions (S) do not obtain. It seems that our necessary conditions (N) are acting as nothing more than a base, or a platform that allows for (BC) to obtain. If (N) has any causal efficacy at all, it is passive. Our sufficient conditions (S) are a different story. If (BC) does not obtain, then we can rest assured that (S) did not obtain either. (S)'s causal efficacy is of an active sort (i.e., (S) is actively and directly responsible for the fact that (BC) obtains).

Now that I have answered Fischer’s first question we can go on to the second. Why indeed should we be concerned with the existence of the horizontal line if we are not concerned with the existence of the vertical? In light of the previous paragraph’s revelations, we can rewrite Fischer’s question in the following way: “why should we be concerned with external conditions (S) that play and active role in bringing it about that the (BC) obtains and not with some other external conditions (N) that do not play such a role?” The sufficient conditions occupy the causal role that we normally attribute to an agent (in this case we shall call the agent (m)). This is evidenced by the way we normally speak. For example, if I were to ask, “Why did Jim do X?” we would expect the answer to be “Well, because Jim decided to.” But if determinism is true, then Jim did X not because he wanted, but because the antecedent external sufficient conditions determined him to do so. So we are facing two different pictures of causality for an event. The first is the standard way of looking at causality with the agent causing an event: (m) → (BC). But in a deterministic universe we also have external antecedent sufficient conditions that can also fully explain the event (S) → (BC). If this is true, then we either have a case of over determination of (BC) where both (S), and (m) each (but not jointly) cause (BC), or we must subsume (m) under (S). Most compatibilists (or semi-compatablists) in general, and Fischer in particular, would be happy to subsume (m) under (S).

So what happens if we subsume the agency line (m) under the external antecedent sufficient condition (S)? The normal way of thinking, (assuming that we have free will) is that I go to the party (BC) if I so choose. What happens if I choose not to go to the party? Well, if I choose not to go to the party, and in-fact do not attend it, then we have ¬(BC) which entails that ¬(S). But this is problematic. Remember that under determinism the history of the universe entails only one future. The history of the universe plus the laws of
nature insures us of (BC). Here one might be inclined to counter that we again have a case of over-determination, with the external antecedent sufficient conditions entailing (BC) and the conjunction of the history of the universe and the laws of nature entailing (BC), but that is not the case. What are the antecedent conditions for an event but a particular group of facts of the universe and the governing physical laws at any time prior to the event? In short the external antecedent sufficient conditions (S) is a subset of the conjunction of the history of the universe and the laws of nature. For the agent to bring it about that ¬(BC) when (S) entails (BC), the agent must be able to change either some fact about the universe or violate its physical laws. An agent cannot do this. It is controversial at best to claim that we have control over things that have happened before we existed\(^5\). This is exactly what we must claim if we want to allow that the agent can choose ¬(BC) when (S) entails (BC). This is where the consequence argument\(^6\) and rule β (2)\(^7\) come into play. If an agent has no control over the history of the universe (H), the laws of nature (L), and the external sufficient conditions (S) are a subset of the conjunction of the history of the universe and the laws of nature (H+L) generally, then the agent has no control over (BC) either. If we have no control over (S) and (S) entails (BC), if (S) obtains then (BC) obtains irrespective of the agent. Now one may claim that the conjunction of the history and the laws of nature cause the agent’s desires, which in turn will cause (BC), and the agent will often get what he desires. But this, as I will address later in paper is not enough for the assignment of moral responsibility. At last we have arrived at the answer to Fischer’s question. We must be concerned with the horizontal line (S). The external sufficient conditions pose a direct threat to at least free will and as I will argue to moral responsibility as well.

---

5Why must we claim that the agent be able to alter the state of the universe at a time before the agent existed? Given determinism, (H+L) dictates one possible outcome. So let us say that an agent was to alter fact (F), which is a part of (S) and (F) occurred during the agent’s life. For this to happen, the antecedent sufficient conditions for (F) would have to have been different than what they actually were. You can see where this regress goes. The antecedent sufficient conditions (S) for event any event were present at the big bang which obviously was prior to the agent’s birth. This is what the agent would have to alter.


Here is a proper rendering of Fischer’s original diagram:

**Sufficient Conditions**

A → B → C

**Necessary Conditions**

\((N)\)

As you can see, there is neither free will, nor agency.

### II. Sourcehood, Guidance Control, and Moral Responsibility

Fischer claims that it is true that the sufficient conditions rob the agent of regulative control “which requires genuine access to alternative possibilities”\(^8\) and therefore free will. But it does not rob the agent of guidance control, and it is this particular type of control which moral responsibility requires. This is not correct. If the antecedent sufficient conditions rob an agent of regulative control and free will it also robs an agent of guidance control and moral responsibility. Even if guidance control were all that was necessary for moral responsibility, we still would not have it. Guidance control, according to Fischer consists of two parts, 1) That “the mechanisms that issues in action must be ‘the agents own’”, and 2) “it must be appropriately reason responsive.”\(^9\) So what is it to be “the agents own”? We can take it to mean two things. First is a weaker phenomenological claim. Ownership here is that agent experiences it as his (stemming from him, or being caused by him) and therefore claims it as his. But as we all know, phenomenal experience is no guarantor of metaphysical reality. We are looking for metaphysical moral responsibility, not phenomenological moral responsibility. Phenomenological ownership will just not do. I do not think that Fischer had this weaker claim in mind. I believe that Fischer had a stronger claim in mind, that the agent be an ultimate or key cause of guidance control. The type of ownership or sourcehood that Fischer is referring to need not be the total-control sourcehood that we discussed earlier; but only such that if we had a list of ultimate causes for this guidance control, the agent would be among them.

Herein lies the problem. Within a causally determined physical universe an agent is wholly bound, caused, and predicated by the history of the universe and laws of nature. Under determinism, guidance control cannot be owned by the agent in any robust sense. Being a physical universe, guidance control is a physical phenomenon, which by the consequence argument the agent has no control over. Or if you like, given the truth of determinism, the

---


\(^9\)Ibid., 78.
agent is not an ultimate source of this guidance control; the agent is not a source of anything. If \((BC)\) now stands for an agent’s will or decision making process at any given time, then once again we have the external antecedent sufficient conditions \((S)\) causing \((BC)\) not the agent. Determinism necessitates that even this guidance control is not determined by the agent but by the antecedent external sufficient conditions. This includes the reasons, desires, and will that guidance control responds to. The agent’s reasons at play are the result of physical processes as well. The agent has no control over which reasons are selected, or which reasons wins out in deliberation. In what way does the agent “own” this mechanism for guidance control? It appears only phenomenologically; under any other interpretation the agent has no real claim to ownership. Guidance control has as its sufficient conditions the same external antecedent sufficient conditions that robbed the agent of his free will.

Let us examine Fischer’s Frankfurt example a little closer.\(^{10}\) Jones is going to vote, and wants to vote democrat. Now a neurosurgeon has implanted a device in Jones’s head such that were Jones to decide to vote republican, the neurosurgeon would intervene and cause Jones to vote democrat. Fischer claims that had Jones freely chosen to vote democrat and freely votes for the democrat, then he has exercised guidance control and is morally responsible for his vote even though he lacks any robust possibility to do otherwise. The problem with this example is the fact that Fischer does not grant the neurosurgeon quite enough power over Jones to make his example a proper analogy for determinism. The neurosurgeon, like determinism, is responsible for Jones lacking alternative responsibilities. But unlike determinism, the neurosurgeon is not responsible for what is going on in Jones’s head while Jones debates on which way he should vote. In fact it seems that Fischer is presupposing the falsity of determinism when he allows that Jones can of his own accord decide which way he wants to vote. We must restate the story in order to make it analogous to a fully deterministic universe. Here we must grant the neurosurgeon control over not only the alternate possibilities, but over Jones’s thought processes as well. How would this story go? Well we can start out in the normal sort of way, Jones wants to vote democrat. The neurosurgeon implants a chip such that were Jones to vote republican, the chip would cause Jones to vote democrat. Moreover, the neurosurgeon fully determines the physical processes in Jones’s brain that are responsible for his deliberation, his reason, and his will. So does Jones have the proper guidance control? The answer is no. The neurosurgeon (in our story), and the deterministic universe (in real life), are fully in control of Jones’s guidance control and fully responsible for the actions Jones performed. Jones is in control of nothing.

The Frankfurt example highlights another problem with Fischer’s guidance control and moral responsibility in a causally determined physical universe. It seems that guidance control would require its own form of regulative control or alternative possibility. Would we hold an agent morally responsible under the guidance control requirements even if given \((S)\) it were impossible for the agent’s guidance control mechanism or his thoughts or will to obtain different than they actually did? I think not. Consider \(P\): Due to forces

---

\(^{10}\)Ibid., 56.
external to the agent \((S)\) the agent’s actions have been fully determined since before the agent existed. Therefore the agent has no free action. The above statement is in keeping with all we have said prior. But let us replace “actions” with “thoughts, wills, desires”. 

Consider \(P^*\): Due to forces external to the agent \((S)\) the agent’s thoughts, wills, desires \((BC)\) have been fully determined since before the agent existed \((S) \rightarrow (BC)\). Therefore, the agent’s thought are not free. If we are to grant the truth of \(P\) then it looks like we should grant the truth of \(P^*\) as well. It is hard to see how we could hold anyone morally responsible under these conditions. In order for an agent to be morally responsible under even guidance control it seems plausible to require that while the agent may not be capable to act freely, the agent be at least able to think, will, and desire freely. To put it another way, it must be possible for the agent to have wished, willed, desired contrary to what he actually did. Seeing that the agent is a physical thing, any change in the obtaining desire or deliberation process or result should correspond to some physical change about some fact in the universe. But this is precisely what Fischer claims the consequence argument rules out. So even if guidance control were all that was necessary for an agent to have moral responsibility, the agent does not have it.

One might argue that this is not an example of guidance control in the way that Fischer describes it. This being the case we may still have claim to moral responsibility. Fischer claims that guidance control must be internal to the agent, owned by the agent, and appropriately reason responsive. It looks like our modified Frankfurt example fulfills all of these requirements. The guidance control is still reason responsive (although technically Jones is not responsible for the reason), the process is located within Jones’s skull (so it is located within the agent), and Jones owns (as much as any determined agent could) the action. This is to say that they both own the phenomenological experience of the action stemming from the agent. This, as I stated before is not the metaphysical ownership that Fischer needs for actual guidance control, but it is the only form of ownership that Fischer could ever hope to achieve. This roughly satisfies the conditions for guidance control. However, due to the deterministic nature of the universe and the antecedent sufficient conditions over which Jones has no control, Jones’ possession of guidance control is not a metaphysical reality, but at most a phenomenological experience. Jones does not have guidance control in reality, nor does he possess the capability of being morally responsible.

### III. Conclusion

We have seen that if we accept the truth of determinism, the vertical line \((N)\) has a limited sort of causal efficacy such that \((N)\) and \(\neg(BC)\) could obtain. The truth of the horizontal line \((S)\) on the other hand completely determines the truth or falsity of \((BC)\). In completely determining the truth of \((BC)\), \((S)\) wholly excludes the agent \((m)\) from any sort of causal power. So it looks like we, as agents must be concerned with the horizontal line of sufficient conditions. Not only does it rob us of alternative possibility and free will, it robs us of agency, sourcehood, and moral responsibility. If determinism is true, then we have no control over any physical aspect of the universe. Since the universe is (or at least our best science tells us that the universe is) wholly physical, then we have no control over anything. No control over the paths we choose, the choices we make, or the thoughts we
think. Alternate possibilities are ruled out, and guidance control is ruled out as well. With the departure of alternative possibility and guidance control, so too goes moral responsibility and free will. If determinism is true we are agents in name only. Agency, moral responsibility and compatibilism disappear, and we become automatons morally responsible for nothing and locked in a hard incompatibilist universe.
Charles Bauch – A Phenomenological Understanding of the Rosary Considering Heidegger’s Conception of Worldhood

In *Being and Time* there is a chapter entitled “The Worldhood of the World” wherein Heidegger asks concerning the worldhood of the world: “By what avenue do we meet this phenomenon?”¹ This question is well placed, because in what follows Heidegger takes the reader down a series of paths in order to finally meet this phenomenon of the world. It is through his analysis of entities within the world that he lets the reader see the phenomenon of the world through the possible ways which entities can be encountered in Dasein’s “handling” of them. I would like to take up his path and follow it but do so in a very particular way. In Heidegger’s writing, he considers entities generally. Occasionally, he has recourse to some particular entity within the world (giving specific examples of the hammer and the car signal) but his overall approach is an understanding of the worldhood of the world with a view of entities which does not set any one aside and focus solely on it in order to arrive at an ontological understanding of the world as a constitutive part of Dasein’s Being-in-the-world. This way of discussing entities generally is fruitful in that it shows a widely-applicable relationship between Dasein and the world in every instance of Dasein’s Being-in-the-world and dealings with entities. In what follows, I will focus on one particular entity within the world: the rosary and follow Heidegger through his process of disclosing the phenomenon of worldhood. I will show how the rosary as an entity within the everyday environment with which Dasein is involved both fits into Heidegger’s conception of equipment and diverges from it, thereby challenging his way of approaching phenomenologically the worldhood of the world through Dasein’s everyday dealings with equipment.

When discussing something like a rosary tendencies immediately arise that call us to interpret the rosary-thing. We may see the religious thing as something “special,” as something carrying with it “sacred” theological attributes or qualities. In like manner, a contrasting view of the rosary can stir one to quickly dismiss the religious thing and the belief structure imbued with it, perhaps providing a reductionistic account of the rosary-thing’s supposed sacredness and theological weightiness. I am concerned with how Dasein approaches the rosary, namely how Dasein’s everyday attitude toward the rosary reveals a worldly part of Dasein’s Being-in-the-world and how our everyday dealings with it show a certain kind of Being which the rosary has. While it will be exceedingly difficult if not impossible to completely do away with some preliminary interpretation of the rosary-thing it is not my intention to get bogged down in the theological disputes which try to defend or debunk the “validity” of certain religious practices and their associated objects. Like any phenomenological investigation we are concerned with what shows itself in the rosary-

¹Heidegger, p. 289.
thing itself and we aim at the Being of the rosary in so far as our dealings with the rosary reveal the worldly Being of Dasein.

When I see the rosary in front of me, I am not immediately thrown into some speculative or theoretical mode of thinking about the rosary; I plainly see the rosary as given before me. But what exactly is in this seeing which occurs in the world of Dasein’s day to day Being? If we are after a phenomenon of the worldhood of the world, we will need to delve deeper into this kind of everyday seeing which does not immediately interpret ontologically the world of Dasein and yet can proceed with a pre-ontological understanding of that world which in every instance accompanies our encounters with the rosary. Our present analysis needs to turn its gaze toward a kind of everyday Being-in-the-world which Heidegger calls the environmental world of Dasein and which Heidegger says is "closest to Dasein." In its intimate closeness to Dasein, this world and the Being of the entities encountered within it are always being left in the dark and receive little attention. Often we live in the interpretations of the entities around us and in our absorption neither pausing to reflect on our concern for the entity which we encounter daily nor the way which our dealings with the entity illuminates the worldly aspect of our Being and that of the entities. However, if we are after a phenomenological understanding of the worldhood of the world we must begin with the everyday world of Dasein and our concern for the entities which we find in this world. After all, this everyday kind of Being-in-the-world comes prior to every other concernful dealing and way of seeing an entity that is within the world.

In order to explicate what everydayness entails and to get back to the entity which is the theme of our analysis let us return to the rosary. According to Heidegger, our everyday dealings with the entities exhibit a concern for them that leads us to manipulate and use them. When we employ this language of use and manipulation with reference to the rosary as the entity which gets used or manipulated, one should pause and think about what Heidegger means by “use” and “manipulation.” Pragmatically, these words are often contextualized by the ways in which tools are used. We say this is useful for that or I can manipulate this to make that. Within this context, the religious person might protest that the rosary cannot be put into the category of mere tools because it is something sacred and demands respect in a way that, for example, a hammer does not. Even the non-religious person could question the use of the rosary since “use” in a positive sense can imply some produced result and for this non-religious person there is no conceivable result or product because the rosary is, in fact, useless.

These interpretations of “use” and “manipulation” miss Heidegger’s meaning. In the case of the religious person’s interpretation, if the rosary "demands" respect it is because it has already valued in a specific way just as the hammer "demands" to be used because it too has been valued in a specific way. In like manner, the non-religious person values the

---

2Heidegger, p. 290.

3Admittedly a vague term but what I mean here is anyone who does not approach the rosary as a sacred, religious object.
rosary in a much different way but nonetheless is directed toward some value of the rosary. We can see that the entity is never just an entity, free from any interpretation and utterly valueless. Pregnant in our phenomenal encounter with the rosary (or any other entity) are the interpretations and colorations of value which conceal the entity as a thing in itself.\(^4\) Heidegger recognizes this “investiture with value”\(^5\) and gives the name “equipment” to those entities which are given value in our everyday concern for them. So “use” and “manipulation” point generally and more primordially to a type of concern for the entity which is already on the way to the equipment structure of the Being of that entity. In “handling” the rosary, both the religious person and the non-religious person see the rosary as something with a value that they can speak of and thus as something already thrown into a nexus of equipmentality.

Before I proceed any further I would like to say a word regarding the religious/non-religious divide and, in addressing this, explain more fully what is meant by “an everyday dealing with the rosary.” For some people, even those who we might have previously labeled religious, the rosary is not something they would encounter on a day to day basis. For this reason, an encounter with the rosary by one of these people will be different than that person’s encounter with a car with respect to the everydayness of our dealings with the entities. In fact for someone who does not encounter the rosary everyday, such an encounter with it will disrupt a kind of everyday Being-in-the-world; so much so that the rosary becomes something extra-ordinary in our interpretation of it. Thus, our encounter may hearken us back to some childhood memory, call us to an awe inspiring religious reflection, or give us pause to consider a religious object that we aren’t familiar with or don’t understand.

These extra-ordinary encounters, during which the entity is not used in the way we discussed above, mark “breaks” from our everyday concernful dealings. The entity is not invested with value but rather has value in its mere presence. Heidegger calls this kind of Being that the entity has presence-at-hand. While presence-at-hand is important and I will return to it again in my considerations of the rosary, these out of the ordinary encounters are fleeting and are not what Heidegger finds important in revealing the worldhood of the world. The everyday Being of Dasein takes primacy in Heidegger’s exposition of the phenomenon of the world because the entities within it reveal a worldly part of Dasein. The priority of Dasein’s everyday Being-in-the-world can be demonstrated in that Dasein’s handling of the entities within the everyday world of concern must first be shed and broken out of in order to arrive at presence-at-hand.

However, a rosary is not always encountered as an exceptional entity present-at-hand. Indeed many people handle a rosary everyday because they are devout or have taken vows to do so. If we consider the religious person for whom the rosary is an entity in their

---

\(^4\)See p. 292 where Heidegger discusses our “interpretative tendencies which keep thrusting themselves upon us and running along with us.”

\(^5\)Heidegger, p. 292.
environmental world, we are poised to resume our Heidegger-guided analysis of the rosary in the context of its being valued in an everyday concern for it.

I mentioned above the nexus of equipmentality into which the rosary was thrown and I would like to return to this concept. For Heidegger equipment is never found to be an isolated, singular thing. It belongs to a “totality of equipment” out of which the individual equipment emerges. “Equipment is essentially something in-order-to” and thereby has what Heidegger calls assignment or reference. If we take the rosary and apply what Heidegger is saying about equipment we can see that the rosary has an equipment totality. It emerges out of other equipment: the crucifix, holy water, kneelers, the chalice etc. These are all encountered in the church setting and, being religious entities, they are equipment to adore God. When we handle the rosary, that is, when we pick it up, hold its beads and pray we engage in the rosary’s assignment, its in-order-to. The act of praying with the rosary discloses most genuinely the equipmentality of the rosary and the way in which it is encountered as something in-order-to. Heidegger says: “The kind of Being which equipment possess...we call ‘readiness-to-hand.’” For Heidegger, this readiness-to-hand is not found in the outer appearance of the rosary but rather it is shown in our involvement with the rosary when we engage ourselves with the rosary as something assigned to commune with a higher power. In our dealings with the rosary, we see the various instances of the rosary’s in-order-to. For instance, we approach the rosary-thing as something in-order-to serve God. This “seeing” around and through the rosary is what Heidegger calls circumspection. When I approach the rosary I do not see the rosary first but rather see the totality of its involvements which assigns it to something for prayer or praise of God.

What do I mean when I say that the rosary has a kind of Being which is ready-to-hand? For Heidegger, this readiness-to-hand of the entity is not something theoretically derived after an analysis of the entity; in fact it must recede along with the entity itself in order to be genuine readiness-to-hand. When we engage with the rosary, that is to say when we pray, we are no longer concerned with the rosary itself as something ready-to-hand but are now absorbed in the “work” the act of praying and while we are still holding the rosary in our hands, our concern is now for the prayers to be said. In this way, the rosary recedes in its readiness-to-hand and what becomes ready-to-hand is the prayer we are praying. The prayers to be said are comparable to “the work to be produced” which Heidegger calls the towards-which. For the devout person who prays the rosary often, the prayers that are said

---

6Heidegger, p. 292.

7Ibid., p. 292.

8Ibid., p. 293.

9Heidegger uses this world and while I realize that it is a crude way of interpreting prayer, I wish to stick to his vocabulary for the time being.

10Heidegger, p. 294.
are not just words without significance; they are used for the glorification of God or the intercession with the Virgin Mary. It is in this “work” that we see the rosary’s usability and the toward-which in the structure of its Being ready-to-hand.

The rosary is always made of some material that we can trace back to nature. The discovery of the natural material which was used to make the rosary would refer to what Heidegger calls the where-of of an entity and in this discovery, nature is seen as that ever productive womb out of which flows materials for our use. The rosary has too a reference to the person who is to use it. Some rosaries are brightly colored with thick beads and are clearly meant for children while others include more beads and are intended for a longer duration of prayer (i.e. for an adult).

The Being ready-to-hand which the rosary (or any entity) has in our everyday concern for it is not a subjective characteristic of the entity that we “give” to it. In our everyday concern for the rosary we do not initially see it as pure presence-at-hand behind which lies the ready-to-hand and the equipmental structures we have discussed. In terms of an average, day to day concern for the rosary a purely theoretical type of seeing is the exception. When we first encounter the rosary we do not see it as a collection of atomic particles fitting into the ontological world of a science and occupying one niche of the ontic “world” that is comprised of other entities present-at-hand. What is closest to Dasein is not the world of theory but the everyday, environmental world. If we are after the worldly part of Dasein’s Being, then we must focus on Dasein’s world of everyday concern which is, in every instance, closest to Dasein and restrain ourselves from going “beyond” this world to a theoretical observation of the rosary which always proceeds out of our original, everyday concern for it. Similar to Heidegger’s discussion found in his essay on the fundamental discoveries of phenomenology, when we walk into a room what we see first is the chair itself, a chair that is, in our initial concern with it, in our way and therefore pushed to the side. Only after this kind of primary concern has played itself out, do we begin to dwell on the constituent parts of the chair and see it as a Being of nature present-at-hand. Therefore, Heidegger writes: “Readiness-to-hand is the way in which entities as they are ‘in themselves’ are defined ontologico-categorically.”

We have seen that in our absorption in the activity of prayer, the Being ready-to-hand of the rosary was veiled or receded in our involvement with the rosary. In his worldhood of the world chapter Heidegger gives us three ways (conspicuousness, obtrusiveness, and obstinacy) in which the readiness-to-hand belonging to the equipment within the world of our environment reveals itself as unready-to-hand and thereby exhibits the manipulability and usability of those entities encountered in the world of our everydayness—our environment. Using conspicuousness as our example, we encounter the rosary and see that it is broken. In its brokenness and unusability we see its Being unready-to-hand as present-at-hand. This is a different kind of present-at-hand than the extraordinary and theoretical presences-at-hand in that what is now present-at-hand is the equipmental structure of the rosary as an entity in the world of our everyday environment. We see the broken rosary and we become acutely aware of its use and assignment precisely because we cannot

\footnote{Ibid., p. 295.}
handle it in prayer, and now, in our everyday concern with it, it shows itself as unready-to-hand. Whereas in the instance when the rosary was not broken we could, through circumspection, be absorbed in its totality of involvements and remain unaware of its Being ready-to-hand thereby passing over the Being of the rosary in the world of our everyday concernful dealings. Now however, upon finding the rosary broken “what is ready-to-hand within-the-world is accessible for circumspective concern,” this being due to a break in our circumspective absorption in the totality of the rosary’s involvements. The broken rosary, as an entity in the world of our everydayness, exhibits the worldly character of usability which accompanies our everyday concernful dealings and reveals the readiness-to-hand of the rosary in the environmental world of Dasein.

Let us recapitulate what we have said up until now. We said that an everyday kind of Being is closest to Dasein and therefore we should consider entities within the world of our everyday concernful dealings if we want to see, phenomenologically, the worldly part of Dasein’s Being. In our everyday concernful dealings, we invest entities with value in so far as we can “use” or “manipulate” them and the entities are encountered as equipment. These entities, like the rosary, have a kind of Being which is ready-to-hand and this Being reveals itself in our involvement with the entity (although the Being ready-to-hand of the rosary in this instance is passed over and recedes). The Being ready-to-hand also reveals itself when we encounter the entity in its modes of detrimentality that is in its being obtrusive, obstinate, or conspicuous.

While we have taken the rosary down the path which Heidegger lays out and we have seen how it can apply to Heidegger’s analysis of equipment one might feel as if the analysis provided has neither adequately described the kind of Being which belongs to the rosary nor given a full account of the religious person’s everyday concernful dealings with the rosary. Is it possible that the rosary, despite being an entity within the world of everyday environment, can diverge from the kind of Being ready-to-hand of the entities that are encountered and handled in Dasein’s everydayness?

Our first indication of the inadequacy of applying Heidegger’s analysis of equipment to the rosary is grasped when we compare the towards-which of the rosary to the towards-which of another entity like the hammer. Heidegger discusses the towards-which as “the work to be produced” and as something with usability so that the entity is used for something. However, if we treat the rosary and the praying as something which is useable for something we lose sight of the faith-centered and sacred activity of praying with a rosary. At the point when the religious person is said to use his/her rosary and the accompanying prayers for something, we have reason to doubt the authenticity of that person’s faith. In fact, the faith centered praying of the rosary is radically contrasted to the towards-which of the hammering to build a fence. Yet, the rosary is, ontically, as much of an entity within the world as the hammer and, as we have seen, in the everyday world of the religious person. Like the hammer, it is made to be handled by the person engaging with it. It can be

---

12Ibid., p. 299.

13Ibid., p. 294.
discovered in terms of other religious entities. However, whenever we try to make concrete the assignment or reference of the rosary we are left describing the rosary and the prayer of the rosary as equivocal to a project, as something for-the-sake-of. If the rosary has the kind of Being which is ready-to-hand then it must bespeak an involvement in something. Heidegger traces back “involvement in” until he reaches a primary towards-which, which he calls Dasein’s for-the-sake-of-which. Given that all the involvements of the entities ready-to-hand relate back to Dasein’s Being Heidegger next discusses how Dasein lets the entities ready-to-hand be involved. When we let something be involved we must already have an understanding of the totality of involvements we are letting the entity be involved in. This totality of involvements is the world which Dasein directs itself toward and it has been disclosed to Dasein beforehand so that Dasein can let something be involved in the first place. The letting be involved in something is always in terms of a towards-this and this “towards-this” must, for Heidegger, be disclosed beforehand with a certain intelligibility.”\textsuperscript{14} However, the world in which the rosary is involved is in essence a spiritual one and, being grounded by a genuine act of faith, the rosary cannot be involved in a world that Dasein knows intelligibly beforehand. The rosary is therefore not genuinely an entity ready-to-hand. In a faith based concern for the rosary, one could be said to let the rosary be involved in a divine world but it is because of the faith-concern, with which the religious approaches the rosary, that the world of the rosary’s involvements is not the same as the world of Dasein’s projects, wherein Dasein assigns itself an in-order-to. Due to faith, the world that the rosary is involved in is present to the religious person as they pray and reveals itself to that person.

What I find most peculiar about the rosary is its hybrid status when viewed in the context of Heidegger’s discussion of worldhood. It presents us with the possibility of being either equipment ready-to-hand or, in a genuine encounter with it, as something sacred and akin to the chalice that Heidegger discusses in The Question Concerning Technology. It is fathomable to conceive of the rosary as a tool or a project of the religious person’s Being-in-the-world. Consider the religious person who, in their everyday concern, encounters the rosary (like any other equipment) as work to be done, as it is involved in the world of religious duties and handled for-the-sake-of completing a daily ritual similar to a chore which one must finish. Yet the rosary presents itself to the religious in a different light and despite it being a part of their everyday dealings, their concern for the rosary is not that of assigning it to a for-the-sake-of-which, which shows the rosary’s involvement in a world that is both familiar and wherein Dasein seizes upon and plays out its potentiality-for-Being through the entities ready-to-hand. The religious person encounters the rosary everyday but in this everyday encounter the rosary is not always something ready-to-hand. The concern of the religious person for the rosary diverges from a concern for entitles that leads us to use or manipulate them. As an everyday, environmental entity, the religious person can discover the rosary in its sacredness, as having the kind of Being which is neither present-at-hand nor ready-to-hand but brings forth out of itself something unconcealed and reveals something new to Dasein’s Being-in-the-world. In faith, the religious person’s dealings with the rosary open up a spiritual world which defies any assignments because the rosary is not genuinely for any kind of use. When one prays it, one

\textsuperscript{14}Ibid., p. 307.
prays *with* the rosary so that the rosary neither recedes in the prayer nor is its presence-at-hand later brought to the fore in its brokenness. It has the kind of Being that is always *with* its handler, thereby accompanying him/her in the phenomenological seeing of the spiritual world. In this kind of Being, it sheds any totality of involvement and stands out in its revelatory Being.
Jesse Christiansen – A Defense for Apriority in Naturalized Epistemology

1. Introduction

Willard Quine’s attempt to naturalize epistemology (1969) launched an epistemic debate that survives to this day. The epistemic feud divides disciplines that should be working harmoniously to solve the problems of modern epistemology. An investigation into justification should not be couched as either exclusively analytic or empirical.

Recently, many naturalists have sought a middle ground in the debate. The result is versions of naturalized epistemology that allow for elements of justification that are not just empirical. Such versions are commonly referred to (e.g., Alvin Goldman 1999) as forms of moderate naturalism. There are many types of moderate naturalism. These moderate accounts locate the epistemic rivalry between philosophy and science within the following question: should apriority be allowed in a naturalized story? In this paper I will frame the debate within this moderate tradition. I will be presupposing primarily a traditional definition of empiricism. The traditional definition of empiricism asserts that all knowledge is based on or derived from experience. The term science will be broadly construed to mean inquiry and method derived solely from empirical means.

Allowing apriority in naturalized epistemology encompasses many issues. For example, how do we accurately define and explain apriority? What should a program of moderate naturalism look like and why should we adopt it?

This paper has two main parts. The first part provides a modern analysis of apriority. In the last part I argue for a place for apriority in naturalized epistemology.
2. Analysis of Apriority

A clear, cogent analysis of apriority is a daunting task. My goal is this section is to analyze features of a priority in the context of a priori warrant. However, it could be argued that there are two views of apriority that are required for a complete and plausible account: a priori warrant and a priori knowledge. In our analysis of apriority, our interest will be best served by treating apriority as a form of warrant and not knowledge (or truth). Goldman puts it quite well:

I shall follow the practice of recent discussions that treat the a priori as a species of warrant or justification. This has several advantages. First, it properly allows for the possibility that a belief might have a priori warrant but fail to be true, and hence fail to be a piece of knowledge. Second, it sidesteps, or at least marginalizes, the question of what else is required for knowledge beyond justified true belief. Third, it highlights the fact that unlike the necessary/contingent distinction, which is a distinction between types of truth, the a priori/a posteriori distinction is fundamentally concerned with sources of warrant or justification, not types of (true) propositions.

Goldman 1999, pg. 1-2

Although it is important, as Goldman points out later in his article, that a theory of a priori warrant be upgradable to a theory of a priori knowledge, it will not be required or directly relevant for me to argue for a priori knowledge here.¹

I believe the most important point in Goldman’s passage above is that the epistemic spirit of the a priori is justificatory and not predicating. Whether a belief is true or false has to do with its content. However, whether a belief is a priori or a posteriori has to do with its source of warrant. This is a crucial distinction I would like to clarify.

There is an important difference between the content of a belief, believing a statement to be true, being warranted in believing a statement to be true, and possessing a priori warrant for a statement. First, only a belief can be warranted, not its content. Second, just because I believe a statement p to be true, this does not mean p is true. After all, p may be false. Third, even if I believe I am warranted in believing p, I may not be warranted. And fourth, even if I am warranted in believing p, it does not follow that I necessarily have a priori warrant for believing p.

¹My focus is on theories of a priori warrant that allow for apriority in naturalized epistemology. A theory of a priori knowledge would require an in-depth discussion of Gettier-type problems, which would take us too far afield.
Goldman (1999) structures his analysis of apriority by listing and then treating various traditional features of the a priori. I find this structure useful and I will use the same approach. However, I will not be treating traditional features that Goldman and other contemporary epistemologists have already rejected (e.g., necessity, infallibility, unrevisability, etc.). Instead, I will focusing my analysis on what I find to be the two most promising features of apriority from a justificatory standpoint: non-experientiality and innateness. Overall, I will be contrasting Goldman and Antony’s views against Kitcher’s views.

2.1 Non-Experientiality

Non-experientiality is arguably the most prominent and traditional feature of apriority. That a priori justification be independent of experience is the characterization most commonly associated with apriority. I will consider the following statement, which I will call (NE) for non-experientiality:

(NE) I have a priori warrant for believing p if and only if the source of my warrant for p is independent of experience.

What is exactly meant by independent of experience? One important way to view experience-independence is identified by Goldman (1999) as non-perceptual. Goldman refers to this view as “a negative characterization of the a priori: a warrant is a priori if it is not perceptual” or as “the absence of an experiential or perceptual basis of belief” (Goldman 1999, pg. 8). This would give the following definition of non-experientiality:

(NE1) I have a priori warrant for believing p if and only if the source of my warrant for p is non-perceptual.

But is this really an accurate characterization of the non-experiential feature of apriority? According to Goldman, it is wrong to “equate a priori warrant with non-perceptual warrant” (Goldman 1999, pg. 8). Goldman argues that there are types of warrant that are neither perceptual nor a priori. He cites, for example, introspection:

(I)ntrospection can give rise to warrant, but its type of warrant is neither perceptual nor a priori. Introspection should not be regarded as a species of perception, especially for present purposes, because it has no distinctive type of sensory experience associated with it. Of course, many objects of introspection—e.g. pains, itches, and tickles—have sensory qualities, but introspection per se does not. One can introspect thoughts without any accompanying sensory quality. So one cannot equate a priori warrant with non-perceptual warrant. (Goldman 1999, pg. 8-9)

In other words, I can have non-perceptual warrant for a belief and still not have a priori warrant. Thus NE1 fails.
Another way of viewing non-experientiality is *partial non-experientiality*. This view asserts that one’s warrant for a belief can be first a posteriori and later a priori. Goldman describes it quite nicely:

This proposal is not meant to imply that whenever an agent uses a process that is an a priori warrantor, any belief-output of the process is *wholly* a priori. On the contrary, if one starts with a set of believed premises that originate in perception and then applies an inferential a priori warrantor to that set of beliefs, the resulting conclusion belief does not have *pure* a priori warrant. Nonetheless it seems instructive to say that such a conclusion belief has an *element* or *component* of a priori warrant, simply because there is one strand of its warrant that is a priori. (Goldman 1999, pg. 12)

This definition of non-experiential apriority would give us the following:

(NE2) I have a priori warrant for believing $p$ if and only if the source of my warrant for $p$ originates in perceptual experience but is later warranted independent of experience.

To further clarify NE2, let’s take the example of $2+2=4$. My original source of warrant could be perceptual. For example, I could perceive two groups of two objects lying next to each other. My source of warrant for the particular appearance of these objects would be perceptual. However, to later arrive at the conclusion that $2+2=4$ I would need to make inferences involving the $+$ sign and the $=$ sign. Such inferences would involve a rational capacity that would provide a source of warrant independent of experience.

Kitcher (2000) argues that his theory of warrant is a version of NE2. His analysis is provided below:

[A statement] is an a priori warrant for X’s belief that $p$ just in case [a statement] is a process such that for any sequence of experiences sufficiently rich for X for $p$

(a) some process of the same type could produce in X a belief that $p$
(b) if a process of the same type were to produce in X a belief that $p$, then it would warrant X in believing that $p$
(c) if a process of the same type were to produce in X a belief that $p$, then $p$. (Kitcher 2000, pg. 67)
Kitcher’s analysis is unusual and difficult to understand. However, there are three main points that I think will help clarify his view.

First, Kitcher’s sequences of experiences and resulting belief-producing processes\(^2\) originate in the experiential world *outside the agent*. This is a key distinction between Kitcher’s account and Goldman and Antony’s account. For Goldman and Antony, belief-forming processes are internal to the agent and independent of experience. For Kitcher, external belief-producing processes produce a belief in the agent.

Second, a sequence of experiences *need not overlap* with any other sequence of experiences, as long as the process that produces the belief in me is *some process of the same type*. For example, I could believe \(p\), an agent in China could believe \(p\), and our sufficiently rich sequences of experiences need not depend on each other whatsoever, as long as our belief-producing process is of the same type.

Third, beliefs may be warranted *whether or not* they rely on any specific sensory input. Kitcher includes the following example:

> The knowledge\(^3\) of contemporary mathematicians may be proximally produced by their reflections of what they have absorbed from the past, reflections that do not depend on any specific sensory input, but are ultimately dependent on the collective experiences of the tradition in which they stand. (Kitcher 2000, pg. 90)

Kitcher refers to the above case as his *tradition dependence* view. Any sequence of experiences can include the collective experience of others. A belief that depends on the collective experiences of others would not be based on any direct sensory input. Hence, beliefs can be based on direct or indirect sensory input, as long as the sequence of experiences is sufficiently rich to produce that belief.

---

\(^2\)I will use belief-*producing* process in order to distinguish it from Goldman and Antony’s belief-*forming* process. I believe this highlights a key difference between their views. For Kitcher a process outside an agent *produces* a belief in the agent. On Goldman and Antony’s view the agent *forms* the belief herself.

\(^3\)Notice Kitcher’s use of the word knowledge here is separate from warrant or a priori warrant. In this passage, it is the reflections of the mathematicians that may lead to warrant or a priori warrant.
What is exactly a priori on Kitcher’s account? Whether a belief depends on direct or indirect sensory input, it would still not be non-experientially-based because such a belief would be ultimately rooted in experience.\(^4\) Hence, the component of Kitcher’s analysis that allows for a priori warrant must have to do with his idea of any sequence of experiences.

Let us apply the example of 2+2=4 again. According to NE2, there is an a posteriori warrantor and then an a priori warrantor. Let any sequence of experiences sufficiently rich to result in the belief-producing process that provides a posteriori warrant be two groups of two objects lying next to each other. What would be the sequence of experiences sufficiently rich to result in the belief-producing process that provides a priori warrant? To become an a priori warrantor, the sequence would have to be independent of experience. This is where Kitcher’s account runs into trouble. It is difficult to see how any sequence of experiences could produce the belief in me that the + sign and the = sign lead me to the conclusion that 2+2=4. How could I be said to experience the + sign and the = sign?

Could the fact that Kitcher’s sequences of experiences need not overlap with any other sequence somehow allow for a priori warrant? There are two problems with this possibility. First, even a non-overlapping sequence of experiences would still be rooted in experience. Second, on Kitcher’s view, a non-overlapping sequence of experiences would have to be based on some belief-producing process of the same type. This type-relation between the two belief-producing processes would itself be an experience-rooted overlap.

Kitcher could object that I do not need to make experience-independent inferences involving the + sign and the = sign to be fully justified in believing that 2+2=4. After all, there is the rich tradition of mathematics. Since the sequence of experiences leading to warrant for my belief need not rely on any direct sensory input, I could simply accept the wisdom of centuries of mathematicians. The problem is that the external process of relying on collective experience would not yield a priori warrant. Furthermore, such an approach would be unreliable when dealing with beliefs other than logical or mathematical truths. For instance, long ago most people believed the earth to be flat. If it were not for individuals like Christopher Columbus who reasoned for himself that this view could be wrong, we might still believe a grievous error.

\(^4\)Another possible interpretation of Kitcher’s theory is that beliefs not relying on direct sensory input could be non-perceptually a priori warranted (NE1). However, we rejected NE1 because non-perceptual warrant and a priori warrant are not necessarily the same thing.
Goldman’s accounts differs from Kitcher’s account in that he views belief-forming processes as providing sources of non-experiential warrant. However, his account does not rule out a posteriori sources of warrant for beliefs or beliefs whose warrants originate in perception and are later a priori warranted. Hence, his view allows for both non-experiential and partially non-experiential apriority and would be represented as follows:

\[(NE3) \text{ I have a priori warrant for believing } p \text{ if and only if the source of my warrant for } p \text{ is based on belief-forming processes that do not require perception.}\]

I believe this is the most plausible view. First, it allows for partially non-experiential apriority. This allows us to flexibly but cogently preserve the most traditional feature of apriority. Secondly, by shedding the non-perception constraint, NE3 opens the door to wholly non-experiential apriority, which brings us to the next sub-section.

2.2 Innateness

Another way that apriority can be non-experiential is via innateness. A formulation is provided below. I will use (I) to refer to innateness.

\[(I) \text{ I have a priori warrant for believing } p \text{ if and only if the source of my warrant for } p \text{ is innate.}\]

One way a source of warrant can be innate is through belief-forming processes. For Goldman and Antony belief-forming processes are part of one’s cognitive architecture (Goldman 1999 and Antony 2004). Antony refers to some belief-forming processes as part of one’s “cognitive machinery” which, to alter, “one presumably would need surgery” (Antony 2004, pg. 9). The central idea here is that these innate belief-forming processes\(^5\) are hard-wired into us from birth and are in a general sense extremely difficult to change.\(^6\) Such processes are wholly non-experiential. There is no way one can experience her own belief-forming process.

\(^5\)Like Goldman and Antony, my goal here is to argue for innate belief forming processes as non-experiential sources of a priori warrant. That such sources of warrant also provide scientific evidence that may warrant psychological hypotheses is not directly relevant to my discussion here.

\(^6\)I am not arguing that an agent’s cognitive machinery would not be radically changed if she suffered a serious head injury, for example. That would be counterintuitive. I am simply arguing that in the ordinary life of an individual cognitive machinery is relatively permanent.
There are multiple examples of innate belief-forming processes. Goldman's list includes: “perceptual processes in the several sense modalities, remembering, introspecting, and (many forms of) reasoning or calculating” (Goldman 1999, pg. 11). Reasoning and calculating involve what Kitcher (2000) refers to as propositional and conceptual preconditions of experience.

Preconditions of experience, broadly construed, are conditions that are present before an agent can have a particular experience. For instance, we have been using the example of 2+2=4. It could be argued that there is an innate conceptual precondition that allows us to cognize numerical concepts.

In the case of propositional preconditions the precondition would take the guise of an innate capacity to perform a particular cognitive task. In the case of a conceptual precondition the precondition would be some relevant innate concept that allows the agent to have a particular experience. What is argued to be non-experiential here and thus a priori is the fact that such preconditions occur before a particular experience.

Both Goldman and Antony invoke examples of each to illustrate innate belief-forming processes. Antony refers to a simple case of modus ponens (a propositional precondition) as demonstrative of “the structure of a reliable cognitive machine: the hypothesized syntactic engine inside my head” (Antony 2004, pg. 6). She also uses the example of universal grammar (a conceptual precondition) as evidence of innateness “in the form of explicit rules (linguistic concepts) represented and stored from birth (or close thereto)” (Antony pg. 2004, pg. 7).

Goldman cites examples in the contemporary cognitive scientific literature that illustrate the possibility of innateness by way of propositional and conceptual preconditions. For example, Goldman looks at deductive logic (a set of propositional preconditions), pointing out that many modern theories suggest, “that ordinary people have something like natural-deduction systems built into their heads, quite possibly innately” (Goldman 1999, pg. 17). Goldman also refers to the innate concept of numerosity (a conceptual precondition) and studies that support this possibility. For example, certain studies have produced evidence of a psychological capacity for numerical cognition even in human infants. “Using the standard technique of gauging surprise by length of looking time, Wynn (1992) found that five-month-old infants can correctly detect elementary arithmetic relationships, such as 1+1=2 and 2-1=1” (Goldman 1999, pg. 16).

However, even if we concede the presence of innateness in such belief-forming processes, how does this innateness give us a priori warrant? An argument is needed to go from one step to the next or we could be in danger of conflating innateness with a priori warrant in which case (I) would fail.

---

7 My purpose here is simply to provide a working definition that allows us to view preconditions as non-experiential. Surely a more precise definition can be formulated within a conceptual analytic framework.
One way that both Goldman and Antony argue for innateness providing a priori warrant is by offering distinctions that explain the difference between belief-forming processes that result in a priori warrant and those that do not. Goldman distinguishes between belief-forming processes and methods.

By a “process” I mean something that is part of a person’s fundamental cognitive architecture. By a “method” I mean something that is not part of one’s fundamental cognitive architecture, but something learned, typically by cultural transmission. (Goldman 1999, pg. 14)

Goldman uses the example of a truth-table procedure in logic to further demonstrate.

In one case Harry learns the truth-table method from Ellen, who simply explains how to use it without explaining why it is (necessarily) reliable. Harry simply accepts its reliability from Ellen on trust; he does not use his prior reasoning powers to “see” that it is reliable. In a second case Harry learns the truth-table method from Ellen, who explains why the method is (necessarily) reliable, an explanation that Harry fully comprehends and appreciates in virtue of his pure reasoning powers. In the first case it seems clear that the truth-table method of forming beliefs about tautologies is not an a priori warrantor. For one thing, the method is acquired in part by perception (of Ellen’s testimony), and that perception is not an incidental or eliminable feature of Harry’s acceptance of the method. In the second case Harry seems to have a priori warrant for his belief that the method is (necessarily) reliable, because he himself determines its reliability by pure reasoning powers. (Goldman 1999, pg. 14)

Antony differentiates between possessing and explaining warrant. The former yields a priori warrant and the latter does not. For example, when Harry above uses his own belief-forming process he possesses a priori warrant. When he simply accepts Ellen’s explanation of how to use the truth-table but does not see for himself how to use the truth-table, he does not possess a priori warrant. Hence, belief-forming processes are determined to be a priori warrantors if they are innate processes that are possessed and employed by the agent forming the particular belief. If I accept a belief-forming process based on cultural transmission, according to Goldman, such would constitute a belief-forming methodology and not a belief-forming process. Kitcher’s tradition-dependence view could provide one example of a belief-forming methodology. If I am basing my warrant for a belief on the collective experiences of others, I am simply standing on their shoulders and not forming the belief on my own.
Goldman and Antony’s distinctions above are meant to address the problem of conflation between innateness and a priori warrant. What makes an innate source of warrant a priori is that it is based on my own internal belief-forming processes. This now gives us the following statement:

(I1) I have a priori warrant for believing $p$ if and only if the source of my warrant for $p$ is based on an innate belief-forming process.

However, there is a problem with I1. I may form a belief by my own reasoning process but my belief could still turn out to be wrong. In this case I would not have warrant for my belief after all, let alone a priori warrant. Thus I1 also fails.

Another way that Goldman and Antony argue for theory of a priori warrant is by adding a reliabilist\(^8\) condition to their account. This would provide an answer to the above problem with I1. It takes more than just possessing and employing my own belief-forming process. An agent also has a priori warrant just in case her a priori warrantor reliably leads to true beliefs. We can now give a final formulation of the views of Goldman and Antony regarding innate apriority:

(I2) I have a priori warrant for believing $p$ if and only if the source of my warrant for $p$ is based on an innate belief-forming process that leads reliably to true beliefs.

The objection could be raised that I2 is not sufficient for a priori warrant because a belief does not have to be based on a reliably true innate belief-forming process to be a priori warranted. For example, my belief that a triangle has three sides could be warranted perceptually and later be warranted a priori based merely on the meaning of a triangle. However, I do not see how such a belief could be \textit{a priori warranted} without my rational apprehension of the analytic justification of a triangle. Regardless, this is still an important problem for I2.\(^9\)

\(^8\)Based on their accounts I will presuppose that Goldman, Kitcher, and Antony are all at least standard reliabilists. Since the focus of my paper here is a priori warrant and not reliabilism, I will not discuss reliabilism beyond the reliabilist condition used here by Goldman and Antony to answer this objection.

\(^9\)I believe a proper answer to this objection would require a discussion of internalist versus externalist views of justification, which would take us too far afield.
Kitcher rejects innateness as a priori:

(T)he relativization to lives sufficiently rich for p already allows for a priori knowledge that isn’t innate. (C)onversely, even though one could know that p on the basis of no experience, it doesn’t follow that one could know p on the basis of any sufficient rich experience...” (Kitcher 2000, pg. 69)

I believe there are two ways to interpret Kitcher’s rejection of innateness as a priori. The first way is to interpret his rejection as a criticism of innate knowledge. The other way is to interpret his rejection as a criticism of innateness as a source of warrant. I believe the interpretation will depend on how we read “know that p” (Kitcher 2000, pg. 69).

If we interpret “know that p” (Kitcher 2000, pg. 69) as to have innate knowledge of p, then Kitcher’s rejection of innateness as a priori fails. Innateness is a source that may provide a priori warrant for a belief. However, innateness itself does not lead to knowledge of a proposition.¹⁰

Furthermore, there are several problems with treating innateness as a source of knowledge instead of as a source of warrant. First of all, even if I could know something innately how would I know that I know it? Innateness is supposed to be non-experiential. Also, even if I could know something innately, why should I have to know that I know it in order to be justified that I know it?²¹ Finally, even if I knew something innately it would not follow that I know it a priori.

If we interpret “know that p” (Kitcher 2000, pg. 69) as to have an innate source of warrant for p, Kitcher’s criticism also fails. The process that produces a belief in an agent on Kitcher’s account depends on a sequence of experiences sufficiently rich to produce that belief. Any sequence of sufficiently rich experiences could never apply to an innate source of warrant because any sequence of experiences would always be warranted against experiences in the world. Innate sources of warrant are non-experiential.

---

¹⁰For example, a belief-forming process may be said to cause a belief. However, a belief-forming process itself could not tell us if a belief was true or false.

¹¹Obviously a fair answer to this problem again involves a treatment of internalist versus externalist views of justification, which would take us too far afield (e.g., an externalist may not see this as a problem).
Goldman and Antony argue that propositional and conceptual preconditions of experience are further examples of innate belief-forming processes\(^\text{12}\) that can be a priori warrantors. Thus,

\[\text{(13) I have a priori warrant for believing } p \text{ if and only if the source of my warrant for } p \text{ is a propositional or conceptual precondition of experience.}\]

However, Kitcher rejects propositional preconditions of experience as a priori:

\[\text{If there are any propositions that we have to believe in order to have experience (knowledge), it's an entirely separate issue whether there are processes that would warrant them given any sufficiently rich experience.} \quad (\text{Kitcher 2000, pg. 71})\]

Kitcher is correct to point out the difference between a propositional precondition of experience and any sequence of experiences against which it may or may not be warranted. However, Kitcher wants to make propositional preconditions a posteriori warrantors instead of a priori warrantors in order to reject them as a priori. On Goldman and Antony's accounts, propositional preconditions of experience are a priori warrantors just in case they are innate belief-forming processes that lead to reliably true beliefs. If any sequence of experiences fails to provide a process to warrant my propositional precondition of experience, I would be said to lack a posteriori warrant, not a priori warrant. In this case my a priori warrant would be overridden.

Take a simple case of an agent applying *modus ponens*:

1. If I graduate, my family is proud of me.
2. I graduate.
3. Therefore, my family is proud of me.

What makes the propositional precondition of experience in the above argument an a priori warrantor is not my belief in the logical validity of *modus ponens*. It is the *deductive machinery* inside my head that allowed me to perform the inference. My belief in the logical validity of *modus ponens* may or may not provide warrant against any sequence of experiences (e.g., I may graduate and my family will be ashamed). However, my deductive machinery that allowed me to form the deductive inference gives me *a priori warrant* as long as it leads to reliably true beliefs.

\(^{12}\)See examples cited above by Goldman 1999 (deductive logic and numerosity) and Antony 2004 (*modus ponens* and universal grammar).
Kitcher also rejects conceptual preconditions of experience as a priori:

\begin{quote}
(E)ven though we might have to deploy a concept in order to have experience (knowledge), it doesn’t follow that our belief that that concept was apt for the description of experience would have to be warranted against the background of any sufficiently rich experience. (Kitcher 2000, pg. 71)
\end{quote}

By *apt* for a description of experience Kitcher means *whether or not a concept is directly relevant to the background of experience for which it is a precondition*. Here we see the same issue that arose with Kitcher’s rejection of propositional preconditions of experience. Kitcher is treating conceptual preconditions as a posteriori warrantors instead of a priori warrantors. Relevance and justifiability against a background of experience do not apply to a priori concepts.

Let’s take the example of universal grammar. Suppose I was born with an innate concept that was directly relevant to recognizing basic principles of grammar. This innate concept allowed me to learn English as a child. Let *my parents teaching me my first words* be any sequence of experiences sufficiently rich for me to learn those words. If I were to listen properly and speak recognizable English words, then my innate concept of universal grammar would be warranted against my parent’s teachings. However, what makes my innate concept of universal grammar itself an *a priori warrantor* is that it is a conceptual precondition of experience (an innate belief-forming process). Should my words continue to turn out to be correct than I have *reliably true* belief-forming process.

Although preconditions of experience require further scientific explanation, they provide additional ways to view a priori as innate and thus non-experiential. Although Kitcher raises important inconsistencies when treating innateness and preconditions of experience as types of knowledge or a posteriori sources of warrant, his arguments do not give us plausible reasons to reject innateness as providing *a priori warrant*. Hence, I find I2 the preferred view.

3. Is There Room for Apriority in Naturalized Epistemology?

Good naturalists should not ignore as scientific evidence the occurrences of innate cognitive processes integral to the formation of beliefs. A naturalized epistemology that rejects or overlooks such processes would be implausible and incomplete. Innate belief-forming processes provide a priori warrant based on their non-experiential status. The place for apriority in naturalized epistemology thus lies here.
However, a general problem with innateness theory (including propositional and conceptual preconditions of experience) is that great investigative study into these mechanisms that confer a priori warrant is still needed. Presumably this is the work of the cognitive sciences. Once new data is revealed, it will be the work of epistemologists to establish new conditions for a priori warrant, giving us further insight into the intricate subject of justification. For example, we may learn that we have been improperly ascribing reliability to certain sources of warrant.

Nonetheless, via non-experientiality, I believe what we do have is something important to hold onto from the rich tradition of apriority. The power of non-experientiality is its *transcendental force*—its generally unchanging quality that lends special reliability to certain belief-forming processes that produce epistemically distinct sources of warrant. With future discoveries in cognitive science, the non-experiential aprioristic tradition may end up being more colorful than we ever imagined.

4. Conclusion

I have argued that the more plausible naturalized story is one that includes a non-empirical discussion of warrant. I have also analyzed several features of apriority. I concluded that non-experientiality (including partial non-experientiality and innateness) offers defendable thus surviving features of apriority. Finally, I have argued for a version of moderate naturalism that embraces non-experiential apriority and motivates future scientific research.

---

13I am taking some creative liberty here. However, the idea I have supported throughout my paper is that non-experientiality has a special, relatively permanent nature that gives non-experiential sources of warrant such as innate belief-forming processes certain distinguishability.
Eli Cohn – Externalists on Closure

I. Introduction

Skepticism remains a prevalent problem in epistemology. The skeptic must now show through argument how global skepticism is possible. The mere possibility that we lack knowledge is not regarded as sufficient to deny our knowledge claims. Skeptics have developed such an argument, known as the skeptical trilemma. The trilemma is designed to show an inconsistency among three of our intuitions. The skeptic claims that this inconsistency implies a lack of knowledge. Central to the trilemma is a premise known as the epistemic Closure Principle, which states that if a person S knows that p, and S knows that p entails q, then S knows that q. According to the Closure Principle knowledge is closed under known implication.

The Closure Principle, however, is controversial. The sides of the debate over closure usually correspond to the sides of the internalist/externalist debate about justification. Internalists—who argue we should uphold closure—claim that the reasons a subject has for justifying a belief must be transparent to that subject. That is, the reasons for a belief to be justified are available to the subject’s consciousness. Externalists—who argue we should deny closure—reject the internalist constraint on justification.1 In this paper I examine the externalist position that closure should be denied and responses to this position. I conclude that the externalist attempt to deny closure fails. In section two, I examine Closure and its place in the skeptical trilemma; section three is a brief discussion of the ingredients necessary for proper formulation of the Closure Principle; section four explains Robert Nozick’s truth tracking theory of knowledge and how he uses this theory to deny closure; section five examines Fred Dretske’s denial of closure based on relevant alternatives, conclusive reasons, and heavyweight implications; in section six I offer concluding remarks.

II. Closure and the Skeptical Trilemma

The Closure Principle is central to the skeptical trilemma. The trilemma is a tool used by skeptics to show that it is possible that we lack knowledge on a global scale. If we lack knowledge on a global scale, then none of our beliefs constitute knowledge. The trilemma is valuable to the skeptic because each of the three premises that constitute it are intuitive. The skeptic uses these intuitions to deny knowledge claims.

---

The skeptical trilemma is as follows:

(1) S knows that p
(2) If S knows that p, and S knows that p entails q, then S knows that q
(3) S does not know that q

It should be noted that (2) is an instance of the Closure Principle. The Closure Principle itself will be explained in more detail in the next section.

The trilemma is an inconsistent set of beliefs. If (1) and (2) are true, then we should have not-(3) as the third premise. That is, (1) and (2) should yield not-(3) via modus ponens. Our beliefs should be consistent. But we have (3) instead of not-(3). We will see shortly why this is the case. To escape the force of the trilemma one of its premises must be ceded, but whichever one we choose will have unwelcome consequences. The skeptic denies (1). Denying (1) means denying all of our knowledge claims. Most people consider this result unacceptable. As I discuss in this paper externalists attempt to deny (2). But when we deny (2) we deny other important logical operators. I show this in my discussion of Hawthorne in section five. Other epistemologists deny (3). This may seem mundane at this point (part of the aim of this paper is to explain why denying (3) is not mundane), but denying (3) amounts to admitting our knowledge is not infallible. In Denying (3) we claim that we can know things even though we could be wrong about them! These are the problems that the trilemma forces us to confront.

In order to evaluate the trilemma, we may examine how it works in an ordinary case of knowledge. In such cases, we will have not-(3) as the third premise because our beliefs will be consistent with each other. The inconsistency in our beliefs (and the resulting trilemma) arises when a skeptical hypothesis is introduced, which we will explain in a moment. For now let's consider the following example. Suppose yesterday Smith went to the beach, where he noticed various people swimming in the ocean and recognized one of them as his friend, Jones. For our purposes "p" will stand for “Smith knows that Jones is swimming.” Jones, who also recognizes Smith, decides to swim ashore to talk to him. We have already said that Smith knows that Jones is swimming. Presumably, Smith also knows certain also things about Jones based on the fact that he is swimming. Such things could be “Jones will be out of breath when he gets here,” or “Jones is wet.” Smith knows these propositions because of (2), the Closure Principle. These claims are entailed by “p” and Smith is aware that Jones swimming entails these claims. Let “q” be “Smith knows that Jones is wet,” which again we believe intuitively based on the fact that Smith knows both that Jones is swimming and that swimming makes one wet. The fact that Smith knows Jones is wet shows how knowledge is closed under known implication. If knowledge is not closed (if (2) is false), then Smith will not know Jones is wet.

The skeptic, by affirming (3), denies (1). To accomplish this, the skeptic introduces a hypothesis such as the Cartesian evil demon or Putnam’s brain in a vat.\(^2\) These are global

skeptical hypotheses used to deny all of our knowledge claims. If we are deceived by an evil
demon then we lack knowledge because we are so deceived. If we are brains in vats we lack
knowledge for a variety of reasons. First, it is possible that none of our beliefs correspond
to anything in the actual world and thus cannot be true. Second, even if our beliefs do
correspond to the way the world actually is we will lack knowledge because we will lack
justification for our beliefs. If either the evil demon or brain in the vat scenario is possible,
then we lack knowledge on a global scale. How does the skeptic insert one of these
hypotheses to generate an inconsistency in our beliefs? To answer this question we should
begin again with the first premise. Instead of using “Smith knows that Jones is swimming,”
for “p,” let’s use the claim that I know that I am currently in Portland, Oregon. If I know that
I am currently in Portland, Oregon, then according to (2), I know that I am not being
deceived by an evil demon. For it is true that I am deceived by an evil demon, then I do
not know that I am in Portland, Oregon. Yet I do know that I am in Portland, Oregon, so I
should know that the skeptical hypothesis is false. But I cannot know that the skeptical
hypothesis is false, which is why we are led to (3) and the resulting inconsistency.
Regardless of how much evidence we have against skepticism we cannot know that an evil
demon does not exist and is not deceiving us. The skeptic combines this fact with the
Closure Principle to deny our knowledge claims. Because I cannot know that the skeptical
hypothesis is false, I cannot know that I am in Portland, Oregon.

The Closure Principle is central to the skeptical argument. Take the first, ordinary situation
involving Smith and Jones. Smith knows Jones is swimming and knows that the fact that
Jones is swimming is sufficient for Jones being wet. Thus, we conclude that Smith knows
that Jones is wet via *modus ponens* from the first two premises of the trilemma. Smith
knows that Jones is swimming is (1). Additionally, Smith knows that swimming makes one
wet. Therefore, we conclude based on (1) and (2) that Smith knows that Jones is wet. But
when a skeptical hypothesis such as the evil demon is introduced, the skeptic performs
*modus tollens* in order to deny knowledge. (3) is a denial of the consequent of the Closure
Principle, so the skeptic is able to deny the antecedent as well. This means that either Smith
does not know that swimming makes one wet or Smith does not know Jones is swimming
(or both). Thus the skeptic is able to deny all of our knowledge claims. In both the ordinary
and skeptical cases, the Closure Principle is central to the inference.\(^4\)

It would be natural to deny Closure if it were an unfounded skeptical assumption, but this
is not the case. The Closure Principle is intuitive as are premises one and three in the
trilemma. We can illustrate the intuitive appeal of the Closure Principle using the Smith and
Jones case. If Closure fails in this case, then Smith will know that Jones is wet only when, for

---

\(^3\)Rene Descartes, *Meditations on First Philosophy*. And also Hilary Putnam, “Brains in

\(^4\)Because of the inclusion of the phrase “knows that” we do not have a direct *modus
ponens* or *modus tollens* inference. Dretske has been pointed this out. Whether or not we
can perform the inference depends of how penetrating “knows that” is in a Dretskean
sense. I discuss this briefly later in the paper.
example, Smith observes that Jones is wet, or hears testimony that Jones is wet. If
knowledge were not closed under known implication, Smith could not know that Jones is
wet despite knowing both that Jones is swimming and that swimming makes one wet. But
this seems absurd. If Smith expected Jones to be dry, for example, we would regard him as
failing some epistemic norm, since Smith knows that swimming makes one wet. No one
denies closure in situations such as this, however, externalist or otherwise. Closure need
not fail in all cases for it to fail as a general principle. Closure need only fail in some
skeptical scenarios for it to be the case that knowledge is not closed under logical
implication. Nevertheless, denying Closure is denying an extremely plausible intuition. The
burden of proof on the externalists will be a heavy one.

It should noted that denying Closure does not solve the skeptical puzzle since it remains
true that we are unable to prove the skeptical hypothesis false. Denying closure will,
however, weaken the skeptic considerably. Without the Closure Principle, the skeptic is
unable to use the trilemma to deny knowledge claims. A skeptic may still claim that it is
possible that we are deceived by an evil demon, but this will be mere assertion rather than
argument. Thus, while denying closure is not sufficient for denying skepticism, it goes a
long way to defeating the skeptical argument.

III. Formulating the Closure Principle

There are two individually necessary and jointly sufficient conditions for the truth of the
Closure Principle. First there must be a logical relationship between the two propositions, p
and q, namely, q must be entailed by p. This first condition yields the proposition: if S
knows that p and p entails q, then S knows that q. But this proposition is false. It is possible
that S knows that p and that p entails q but S does not know q. This can be true because S
can be unaware of the logical relationship between p and q. For example one can know that
I am writing this paper in Portland, Oregon, without knowing that I am writing it at 45°
36’N latitude and 122° 36’W longitude, despite the fact that the former entails the latter.

The second necessary condition is a salience condition. S must know the relationship
between the premises (p, and if p then q); otherwise S will not know q. I can know that I am
writing this in Portland without knowing that I am at 45° 36’N latitude and 122° 36’W
longitude because I can be unaware of the relationship between these two facts (namely,
that they refer to the same place). But once I am aware of the relationship, then it is
impossible for me to not know I am writing this paper at 45° 36’N latitude and 122° 36’W
longitude given that I also know that I am writing this in Portland, Oregon. The two
conditions are jointly sufficient for the formulation of Closure stated above: if S knows that
p and S knows that p entails q, then S knows that q. Nozick has accepted this formulation of
Closure and also pointed out that “we would be ill-advised...to quibble over the
details...Although these details are difficult to get straight, it will continue to appear that something like [Closure] is correct.\textsuperscript{5}

IV. Truth Tracking

Some externalists have denied Closure on the basis of truth tracking theories. Knowledge conditions for truth tracking theories are formulated as counterfactuals. Nozick and Dretske both use counterfactual conditions for knowledge, though they differ significantly. Nozick's theory of knowledge begins with two standard conditions and expands to include two truth tracking conditions. The first two conditions are a truth condition and a belief condition:

(1) p is true
(2) S believes that p

These claims are uncontroversial. Nozick adds to these the following two conditions:

(3) If p were false, S would not believe that p
(4) If p were true, S would believe that p

These conditions imply that S's belief in p must track truth for S to know that p. In the closest possible worlds where p does not hold, S does not believe that p. Similarly, in the closest possible worlds where p does hold, S believes that p. So, we can have true beliefs that do not amount to knowledge because we would not hold those beliefs in the closest possible worlds where they are true, or we would still hold them in the close possible worlds where they are false. Finally, it should be noted that one of the main benefits of truth tracking theories is that skepticism never tracks truth. This is because in the close possible worlds where skepticism is true we would still not believe that it is.

There have been some objections to Nozick's view. One such objection uses a variation of the following example.\textsuperscript{6} Say Jack works on the second floor of an office building. His desk faces the window, so that normally, while working, he is looking down at his desk but every so often looks up to see the street outside. He looks outside one afternoon to see someone


\textsuperscript{6}Nozick anticipates this objection in \textit{Philosophical Explanations}, this example is an adaptation from Nozick.
being robbed in the street. But his belief that the person is being robbed fails to track truth. In the close possible worlds where the robbery takes place Jack does not believe that it does, because he does not look outside the window to witness it. Jack’s method of truth tracking, in this case casually glancing out the window throughout the day, does not track truth. He has a true belief, but the true belief is the result of pure dumb luck. It would be unacceptable for Nozick to call this knowledge.

To solve this problem Nozick amends conditions (3) and (4). S must track truth via a certain Method:\(^7\)

\[
\begin{align*}
(3') & \text{If } p \text{ were false, } S \text{ would not believe that } p \text{ via Method } M \\
(4') & \text{If } p \text{ were true, } S \text{ would believe that } p \text{ via Method } M
\end{align*}
\]

\(M\) is whichever method is being employed by \(S\) to track the truth of \(p\). In the above case, Jack does not know he sees a robbery. This is because \(M\), for Jack, is casually glancing out the window and this is not a consistent method. If, on the other hand, Jack were forced to look outside the window constantly, his method might be one that tracks truth. If the robbery takes place with this new method Jack believes that it does because he is looking out the window and witnesses it. If the robbery does not take place Jack would not believe that it does.

After all, he is vigilant and he would have seen a robbery take place had it done so. Nozick uses his formulation of knowledge to deny Closure.

There is no reason to assume the (closest) not-\(p\) world and the (closest) not-\(q\) world are doxically identical for you,\(^8\) and no reason to assume, even though \(p\) entails \(q\), that your belief in one of these worlds would be a (proper) subset of your beliefs in the other. Consider now the two statements: \(p=I \text{ am awake and sitting on a chair in Jerusalem};\ q=I \text{ am not floating in a tank on Alpha Centauri being stimulated by electrochemical means to believe that } p\). Clearly these are very different situations, leading to great differences in what I would then believe. If \(p\) were false...I would not believe that \(p\). Yet if \(q\) were false...I would believe that \(q\)...in that case, I would still believe that \(p\). According to our account of knowledge, I know that \(p\) yet I do not know that \(q\), even though (I know) \(p\) entails \(q\).\(^9\)

Nozick presents a counterexample to closure that highlights the value of truth tracking theories, namely that global skeptical hypotheses fail to track truth. Even if a skeptical hypothesis were true, we would still believe that it was not. That is, we are not able to tell if we were deceived by the evil demon and so will not believe that we are deceived, even if

---

\(^7\)Ibid., p. 184.

\(^8\)Two worlds are doxically identical just in case you have all the same beliefs in both worlds.

\(^9\)Philosophical Explanations, p. 207.
we are. The evil demon world appears identical to the world as we now perceive it. In Nozick’s example, p entails q, and p tracks truth yet q does not, so closure fails. Nozick knows that he is awake and sitting on a chair in Jerusalem because in the close possible worlds where he is not on a chair in Jerusalem he does not believe that he is on a chair in Jerusalem. He does not know that he is not a brain in a vat on Alpha Centauri because this is a skeptical scenario like the evil demon. In other words, that belief that p varies with the truth of p, but belief that q does not vary with the truth of q. Because of this lack of variation in S’s belief that q, S fails to know q and closure fails.

Kripke developed a powerful objection to Nozick’s account of truth tracking.¹⁰ Let’s say that in Rockland, Maine the entire economy is based on tourism. Specifically, Rockland attracts people from all over the world because their town has barns on the side of the road that are of all different colors. The barns are especially attractive in fall, when they stand out against the rest of the passing landscape. Last year, there was a fire in Rockland at the beginning of summer and many of the barns burned to the ground. Not wanting people to know this, for fear of a significant drop in tourism, the citizens of Rockland created multicolored barn facades, ensuring that when the tourists arrive they will think everything is normal, though they actually are looking at barn facades in most cases. The final condition is that there are no red barn facades. When the tourists arrive in fall all of the objects that appear to be red barns are red barns.

Now say that Jack, traumatized by the robbery he witnessed at work, decides it is time to take a vacation. While driving through Rockland Jack tracks the truth of red barns. When there is a red barn Jack believes there is a red barn. When there is no red barn Jack believes there is no red barn. He is using a consistent method to track the truth of red barns; he is consistently looking out the window at barns and his perceptual system is normal. But while Jack tracks the truth of red barns, he does not track the truth of barns. When Jack sees a blue barn facade he believes it to be a blue barn. When he sees a green barn facade he believes it is a green barn, etc. He is using the same consistent method, but when Jack sees a red barn, while he knows it is a red barn, he does not know it is a barn. He tracks truth with respect to red barns, but does not track truth with respect to barns in general. If this objection holds, then Nozick’s theory of knowledge fails and his denial of closure fails along with it. The closure denial fails because of how closely it is tied to truth tracking. Nozick denies closure because p can track truth even though q does not despite that p entails q. But truth tracking conditions cannot be necessary conditions for knowledge in Nozick’s theory, so whether or not p and q track truth is irrelevant to whether S knows p and q.

¹⁰Kripke has never published this example. It was originally used by Goldman, who got the idea from Ginet. I use Kripke’s formulation because it is the best developed and clearest.
Dretske presents a theory of knowledge based on counterfactuals that differs from Nozick’s and is immune to the Kripke barn case. Dretske explains:

S knows it is a barn if that feature of the evidence causing S to believe it is a barn would not exist if there were not a barn. In the case of perception, if its looking like a red barn is what is causing S to believe it is a barn, then S has conclusive reasons to believe it is a barn: it would not look that way (like a red barn) unless it was a barn, and its looking like a red barn is what is causing S to believe it is a barn. Hence, he knows it is a barn. If, on the other hand, it is merely the building looking like a barn that is causing S to believe it is a barn (its color being irrelevant to the causing of the belief) then S does not know it is a barn…Knowledge is information-caused belief and in this second case the information (that it is a barn) isn’t causing the belief. \(^{11}\)

Jack does not know he sees a red barn or any barn for that matter. This is because his belief that he is seeing barns is not caused by barns. So Dretske does not have to admit that Jack knows he sees a red barn without seeing a barn, and thus he can survive Kripke’s objection.

V. Relevant Alternatives and Heavyweight Implication

Dretske first denies closure based on relevant alternatives in “Epistemic Operators.” \(^{12}\) He argues that adding the phrase “knows that” to the conditional in the Closure principle makes the skeptical argument invalid because “knows that” is a semi-penetrating operator. Some operators, according to Dretske, are fully penetrating. A fully penetrating operator is one for which whatever operates on p, such operator penetrates to every consequence of p. Necessity is a fully penetrating operator. Additionally, there are non-penetrating operators. “It is strange that” is a non-penetrating operator, it can be strange that two things occur on their own, but very strange that they would occur together. \(^{13}\) Growing up I played basketball, but not baseball. So it could be strange that I received a baseball for my birthday but not at all strange that I received a ball, even though receiving a ball is entailed by receiving a baseball. In addition to fully and non-penetrating operators are semi-penetrating operators. Sometimes these operators penetrate and other times they do not. “Knows that” and all other epistemic operators are semi-penetrating according Dretske. In the Smith and Jones case “knows that” penetrates. Smith knows that Jones is swimming so when Jones reaches the beach Smith knows he will be wet. Smith would not know Jones was wet if “knows that” failed to penetrate in this case. If an operator fails to penetrate then

---


\(^{13}\) For a more detailed explanation non-penetrating operators, see Ibid., pp.1012.
we cannot perform a *modus ponens* or *modus tollens* inference on it. Hence, if epistemic operators fail to penetrate in cases of skeptical scenarios then the trilemma is invalid.

The fact that epistemic operators are semi-penetrating allows Dretske to deny closure. The skeptical argument is dependent on “knows that” being fully penetrating. If an operator fails to penetrate in even one situation, then it is not fully penetrating, and therefore not closed under logical implication. So, if "knows that" is only semi-penetrating then it cannot be said that any piece of knowledge is sufficient to know I am not deceived by an evil demon. That is, if “knows that” is semi-penetrating then knowing I am in Portland, Oregon does not necessitate my knowing I am not deceived by an evil demon. The force of the trilemma is lost. But even if Dretske is correct in claiming that epistemic operators are semi-penetrating he must account for their failure to penetrate in the skeptical scenarios. For this, he turns to his account of knowledge.14

“Knows that” does not penetrate in the skeptical scenario because the skeptical hypothesis is not a relevant alternative to our knowledge claims. A relevant alternative is “an alternative that might have been realized in the existing circumstances if the actual state of affairs had not materialized.”15 This is never a skeptical scenario. Let’s return to Jack, who, still distraught over the robbery he witnessed, decides to visit the zoo in an effort to relieve his anxiety. The first exhibit he sees is the zebra exhibit.16 Does Jack know he sees a zebra? It looks as though he does. He has been to the zoo before and seen zebras. He has learned about zebras in school, seen specials about zebras on TV, the zookeepers at this zoo do not regularly disguise their animals, etc. But if he knows that he sees a zebra, then he knows he is not seeing a mule painted to look like a zebra, for the former entails the latter (a fact that Jack is aware of). Does Jack know he is not seeing a mule painted to look like a zebra? Say Jack overhears someone remark that a zookeeper has painted a mule because the zebra was sick today. Jack can no longer conclusively rule out all the relevant alternatives. But he still knows he sees a zebra. So, Jack knows he sees a zebra and fails to know something entailed by the fact that he sees a zebra, namely that he is not seeing a painted mule. Hence, Closure fails.

According to Dretske, knowledge requires having conclusive reasons for a belief. Conclusive reasons are reasons which enable us to rule out relevant alternatives. In the above case, Jack knows he sees a zebra without knowing he does not see a painted mule because he can conclusively rule out the relevant alternatives to the former and not the latter. Jack can rule out relevant alternatives to knowing he sees a zebra because he knows what a zebra looks like, is at a trustworthy zoo, etc. He cannot rule out the relevant

---


15“Epistemic Operators” p.1021.

16Adapted from Dretske, Ibid.
alternatives to knowing he is not seeing a mule painted like a zebra. The conversation he overhears about the possibility of the animal in the cage being a painted mule introduces a new relevant alternative to Jack. Jack cannot conclusively rule out this relevant alternative, hence Jack does not know he is not seeing a painted mule. Closure fails because Jack knows he sees a zebra but does not know that he is not seeing a mule painted to look like a zebra. According to Dretske closure fails in cases of skeptical hypotheses because even if we cannot rule out relevant alternatives to a skeptical hypothesis the hypothesis itself will never be a relevant alternative to the original belief “p.”

In recent years, Dretske has refined his account of why and in what cases closure fails. Dretske introduces a distinction between lightweight and heavyweight implications. “When we perceive that p, there are, so to speak, heavyweight implications of ‘p’ that cannot be perceived to be so. There is transmission to lightweight implications of course.” Elsewhere, Dretske explains “this is not to deny that we know—and, perhaps, necessarily know—most of the implications of what we know. But not all of them. There are some things—heavyweight implications—we needn’t know even though our knowledge depends on their truth.” Q is a heavyweight implication of p just in case knowledge of p does not transmit to knowledge of q. For example, according to Dretske, knowledge transmits in ordinary situations but fails to do so in skeptical situations. Dretske offers an example of a heavyweight implication. Suppose our friend Jack, still not over his anxiety, has turned to eating and snacking in large amounts to feel better. Jack is hungry one evening and goes to the cookie jar. He opens the lid and sees cookies. Jack knows that there are cookies in the jar. But Jack does not know a heavyweight implication of this proposition, namely, that mind-independent material objects exist. Presumably there are other means for Jack to know that mind-independent material objects exist, but Dretske makes it explicit that seeing the cookies is not sufficient.

Hawthorne provides several objections to Dretske’s denial of closure. First, Hawthorne objects on the basis of other logical consequences of denying closure. The arguments here centers on the fact that other logical operators will be lost when we deny closure. Closure itself is an intuitive claim, so if denying closure entails denying other intuitive logical operators then it is perhaps best not to deny closure. First, Hawthorne claims that denying closure forces us to deny the distribution principle. According to Hawthorne, the distribution principle states that “if one knows the conjunction of p and q, then as long as

---

17 For a detailed account of why relevant alternatives alone do not entail Closure denial, see Steven Luper, “Dretske on Knowledge Closure,” *Australasian Journal of Philosophy* 84.3: 379-394.

18 “Is Knowledge Closed Under Knows Entailment?” p.16

19Ibid., p. 17.

one is able to deduce p, one is in a position to know that p.”21 The same holds of q. Hawthorne argues that the proposition “g is full of wine,” is equivalent to “g is full of wine and not-g is full of non-wine that is colored like wine.”22 But, because of the denial of distribution, one fails to know that not-g is full of non-wine that is colored like wine even though one knows g is full of wine.

The second argument that denying closure leads to counterintuitive consequences focuses on the role of assertion in knowledge claims. While we sometimes assert things we do not know, we at least believe that what we are asserting is knowledge. From this, Williamson has been able to develop a norm of assertion, namely that one ought not assert what one does not know.23 But when we consider cases of skepticism the norm generates odd results. We can easily imagine a case in which someone asserts a claim p, such as that she has a hand before her, and also asserts that p entails q, that having a hand before her entails that she is not deceived by an evil demon. But surely this person will not assert that she knows that she is not deceived by an evil demon. This odd result is a direct consequence of Dretske’s denial of closure. It entails that people ought to assert the first two premises of a modus ponens argument yet deny that they know the conclusion. This argument by Hawthorne does not strike me as particularly effective. The externalist who denies closure will not believe this is counterintuitive in all cases (such as those involving skepticism scenarios).

Though Dretske acknowledges that the consequences of denying closure are significant, he accepts them in any case.24 Should we accept the consequences along with Dretske? Intuitions will vary on this point, although it is worth pointing out that some philosophers have considered a theory of knowledge that denies closure as failing the desiderata for any such theory.25 Such people will not be appeased by denying other intuitive logical operations such as distribution. For the purposes of this paper, however, I will grant that the logical consequences denying closure are insufficient for abandoning Dretske’s strategy.

Hawthorne argues that Dretske’s theory of knowledge cannot support denying closure based on heavyweight implications. Hawthorne holds that we can rule out relevant alternatives to, and therefore know, heavyweight implications conclusively. He also argues that we can know propositions without having conclusive reasons for holding them. If either is true then Dretske’s account of closure denial fails, because “knows that” will

21“The Case for Closure,” p. 31

22One knows this by equivalence which is not one of the casualties of Closure denial


24See Dretske’s “Reply to Hawthorne” in Contemporary Debates in Epistemology.

25Dretske offers examples in “Is Knowledge Closed Under Known Entailment?”
penetrate to heavyweight implications, in which case “knows that” is fully penetrating. The first claim is false by definition. If we could know heavyweight implications conclusively then they would not be heavyweight implications. The second claim is based on the use of counterfactuals in Dretske’s theory of knowledge. Hawthorne notes that counterfactuals are difficult to interpret in many ways, and using them as necessary conditions for knowledge will doom any such theory. Hawthorne believes counterfactuals as necessary conditions for knowledge will always lead to counterintuitive results, such as the ones Nozick fasces as a result of Kripke’s barn case. Dretske responds by acknowledging that dealing with counterfactuals as conditions for knowledge is indeed curious, but by formulating the counterfactuals carefully we can avoid the problems that are troubling Hawthorne. Thus, Dretske’s account of knowledge can be sustained and he can meet all of Hawthorne’s challenges.

Should we deny closure, then? No. Rather, Hawthorne has neglected the best objection to Dretske, which is that he has failed to show anything the skeptic hasn’t. Both Dretske and the skeptic agree that Closure holds over all lightweight implications. And, both Dretske and the skeptic agree that we cannot conclusively know Closure holds in the case of heavyweight implications. The only difference between the two is that Dretske believes Closure does not hold in the case of heavyweight implication, whereas the skeptic believes Closure does hold, and therefore denies knowledge. So, Dretske has not in fact refuted the skeptic, he has merely shown the possibility that Closure fails in certain scenarios.

So, Closure may fail if we accept Dretske’s theory of knowledge. But, contrary to having independent reasons to accept Dretske’s theory of knowledge, we actually have good independent reasons not to accept Dretske’s theory of knowledge. All we have to do here is recall Hawthorne’s discussion of both the logical and counter-intuitive consequences of Dretske’s theory of knowledge. Dretske’s theory of knowledge forces us to deny intuitions (besides Closure itself) such as the distribution principle, and knowledge claims based on assertion. Dretske has shown that he is willing to accept these consequences, but he provides no compelling reasons for us to accept these consequences. We can therefore choose either to accept Dretske’s theory of knowledge, which has severe counter-intuitive consequences and does not guarantee Closure denial, or we can uphold our intuitions about Closure and try to find another way to refute the skeptic.

VI. Conclusion

I have explored the many problems facing an externalist denial of closure. While this does not entail the falsity of externalism as a theory of knowledge, it does indicate that some versions of externalism are inadequate. Theories that have truth tracking conditions cannot support our intuitions about knowledge in cases such as the one offered by Kripke. Theories that rely on relevant alternatives, conclusive reasons and heavyweight implications have counterintuitive costs. Dretske is comfortable with the costs. However, Dretske’s use of heavyweight implications is unenlightening and does not entail Closure denial. When we add these facts to the other counter-intuitive results of Dretske’s theory of knowledge, it is easy to see that we should not accept conclusive reasons as a means of Closure denial.
If closure does not fail, then the skeptic wins the round against the externalists. There are, however, other ways to deal with the skeptical trilemma. Fallibilism and contextualism are possibilities. Fallibilists are those who deny (3) in the trilemma. This means that we can know things (such as that we are not being deceived by the evil demon) even though we could be wrong about them.²⁶ It is worth noting that some externalists are fallibilists. Contextualists (who are also fallibilists) believe that “know” is indexical, so the standards for knowledge shift with context.²⁷ We can know things in ordinary circumstances without knowing them in special skeptical circumstances. It looks as though one of these other options must be pursued if we are hoping for a way out of the skeptical trilemma.

²⁶See Lehrer in *Theory of Knowledge*.

Daniel Crescenzo – Which Motivation is Valuable?: A Critique of Zagzebski’s Theory of the Value of Knowledge

The value of propositional knowledge over and above true belief does not, as Zagzebski argues, come from one’s motivation to get the truth. It instead comes from one’s being motivated by the virtues. I first outline Zagzebski’s epistemology. I then argue by means of counterexamples that Zagzebski’s truth-motivational requirement for knowledge is not only unnecessary, but irrelevant to whether one knows. I further argue that while we may often be, as Zagzebski claims, unconsciously motivated to obtain the truth, this motive does not provide the best explanation for our belief forming behavior in every case in which one knows. Finally, I argue that it is plausible that being motivated by virtues imparts the value necessary in order for true belief to be knowledge.

Zagzebski’s Theory of Knowledge

Zagzebski’s theory of knowledge begins with her conception of virtue. A virtue is “a deep and enduring acquired excellence of a person, involving a characteristic motivation to produce a certain desired end and reliable success in bringing about that end.”¹ The excellence acquired by virtue is an excellence of character which disposes its possessor toward choosing to act in such a way that increases her worth.² The acquisition of this excellence proceeds through a series of acts which are often voluntary.³ Zagzebski understands motivation here as “a disposition to have a motive,” that is, to have “an action-guiding emotion with a certain end, either internal or external.”⁴ Intellectual virtues have all of this in common with moral virtues, but they differ in this respect: Intellectual virtues are reliable in bringing about the final epistemic end of cognitive contact with reality, that is, understanding, wisdom, or truth, i.e., what Zagzebski collectively calls knowledge, whereas moral virtues are reliable in bringing about a

²Zagzebski, Virtues of the Mind, 95; 127.
³Zagzebski, Virtues of the Mind, 128.
⁴Zagzebski, Virtues of the Mind, 128.
⁵Zagzebski, Virtues of the Mind, 137.
The intellectual virtues include sensitivity to detail, open-mindedness, intellectual humility, intellectual diligence, and recognition of reliable authority, among others.

I now turn to the primary focus of this essay, viz., Zagzebski’s argument that the value of knowledge comes from the intrinsically valuable motivation for knowledge which underlies our virtuous motivation. It is important to keep in mind as I outline her argument that for Zagzebski, the motive of knowledge just is the motive of truth whenever propositional knowledge is the kind of knowledge under consideration. By showing that the value of this motivation is explicable without reference to other [non-motivational] values, Zagzebski believes that she can safely conclude that the source of the value of the virtues and rules is this motivation. She suggests that this explanatory independence is evident from the fact that “followers or violators of epistemic rules are commended or criticized primarily because of the epistemic motivations such behavior exemplifies.” In support of this suggestion she analyzes why we criticize a lucky guesser for his true belief: “If by picking a number out of the air I believe that there are 564 cars in the parking lot, and it just happens there are 564, it is obvious that I do not know there are.” I can in fact be criticized for making such a guess. If I were motivated by the desire to know, I would have learned which procedures are reliable in my epistemic community and that guessing is not one of them, and therefore I would not have guessed. The fact that I have guessed therefore indicates a lack of the motivation to know, and it is primarily this lack for which I am criticized. That I am primarily criticized for a deficiency in my motivation to know reveals that this motivation is the source of the value of intellectual virtue.

In her essay “Intellectual Motivation and the Good of Truth,” Zagzebski describes another case in support of her claim, this time focusing on the lack of an underlying motivation to get the truth. She states that she wants to buy a scale so that she can form the belief, which she would enjoy having, that she has lost 5lbs. She does not care whether the scale is accurate and therefore yields correct readings, only that it gives her the belief that she has lost the weight. Since most scales are accurate, she will probably form the true belief that she has lost five pounds. But because she is motivated not by the desire for the truth, but by

---

6Zagzebski, Virtues of the Mind, 166. Zagzebski holds that intellectual virtues are “best treated as a subset of moral virtues” (167).

7Zagzebski, Virtues of the Mind, 114.

8Zagzebski, Virtues of the Mind, 203.

9Zagzebski, Virtues of the Mind, 197. She calls the form of agent-based ethics which locates value in one’s motivation in this way a motivation-based theory (Zagzebski, Virtues of the Mind, 202).

10Zagzebski, Virtues of the Mind, 197.

11Zagzebski, Virtues of the Mind, 207.
the desire to believe that she has lost 5lbs, that true belief is not knowledge. Zagzebski says that in such a case, “[her] act of true belief...is not as good as an act of true belief arising out of the right motive,” and that she does deserve praise for getting the truth and hence [she suspects] that she does not know.

In “From Reliabilism to Virtue Epistemology,” Zagzebski adds to her argument for the intrinsic value of the motive of knowledge a further argument that the motivation to know is necessary in order for one to know. She begins by noting that knowledge has a value above and beyond true belief. Process reliabilists argue that this value comes from the true belief’s having been the result of a process which is reliable, that is, a process that reliably leads one to true belief. Zagzebski discredits their argument with a simple but powerful example: “Would we say that the bad-tasting espresso produced by a reliable espresso maker is any better than a bad-tasting espresso produced by an unreliable espresso-maker?” Nobody would. We would say perhaps that the process is valuable because it is reliable, but not that the product gets any added value from this process. Likewise for belief: Neither true nor false beliefs gain any value from being produced by a reliable process. Zagzebski goes on to argue that faculty reliabilists like Sosa face the same kind of problem: “Is the value of the espresso produced by a machine functioning as it was designed any better than it would be if it were produced by a reliable but undesigned machine, much less an unreliable and undesigned machine?” Again, nobody would agree to this, and if one carries the example over to belief, we see that no value is added to belief simply because it results from properly functioning, reliable faculties.

Zagzebski also considers why reliabilism has been so appealing in spite of this seemingly obvious failure to account for the value of knowledge. She argues that the appeal of reliabilism is due to a misunderstanding of Gettier problems. In Gettier cases, one comes to have true belief, but since one does so by accident, that true belief is not knowledge. Proponents of reliabilism, Zagzebski maintains, believe that all that is needed to render these cases unproblematic is to ensure that one gets one’s knowledge non-accidentally. Both process and faculty reliabilism do this by tying knowledge to non-accidentally truth-conducive mechanisms. But this is not good enough: Some non-accidental beliefs may not be knowledge, either.

---


17Zagzebski also considers why reliabilism has been so appealing in spite of this seemingly obvious failure to account for the value of knowledge. She argues that the appeal of reliabilism is due to a misunderstanding of Gettier problems. In Gettier cases, one comes to have true belief, but since one does so by accident, that true belief is not knowledge. Proponents of reliabilism, Zagzebski maintains, believe that all that is needed to render these cases unproblematic is to ensure that one gets one’s knowledge non-accidentally. Both process and faculty reliabilism do this by tying knowledge to non-accidentally truth-conducive mechanisms. But this is not good enough: Some non-accidental beliefs may not be knowledge, either.
concludes that the missing element which confers value onto true belief, thereby changing it into knowledge, is the motive of knowledge.  

Finally, Zagzebski argues that our being motivated to get the truth is the best explanation of our behavior in many cases in which we form beliefs, even if we are not conscious of this motive. She states that “A person may be motivated to get the truth and avoid falsehood even when he does not consciously think of that as his end on each occasion in which that motive is operative.” But we are praise or blameworthy for our motives whether or not we are conscious of them. She does not go into great detail as to why she thinks it is the best explanation, but one can easily come up with examples which support this suggestion. Say I look at a bus schedule to see at what time the bus I wish to catch is supposed to arrive. I see “11:05 AM,” and form the true belief that the bus will arrive at roughly 11:05 AM. In looking at the bus schedule, I am appealing to appropriate authority, a kind of virtue. I am not conscious of the motive for truth when I engage in this act, but it is certainly plausible that I have such a motive for appealing to appropriate authority.

Central to Zagzebski’s epistemology is the act of intellectual virtue, which she defines as follows:  

An act of intellectual virtue A is an act that [1] arises from the motivational component of A [that is, one is motivated as a virtuous person would be motivated], [2] is something a person with that virtue would (probably) do in the circumstances, [3] is successful in achieving the end of the A motivation, and [4] is such that the agent acquires a true belief (cognitive contact with reality) through these features of the act.  

The motivational component of A is composed of a virtuous motivation to satisfy the end of a particular virtue, intellectual courage for instance. As we have seen, Zagzebski holds that this motivation must itself motivated by the desire for knowledge. Furthermore, when Zagzebski says [3] that an act is successful in achieving its end, she means that one obtains true belief, since when she says [1] that an act arises from the motivational component of A, 

---

18Zagzebski, “From Reliabilism to Virtue Epistemology,” 119.

19Zagzebski, Virtues of the Mind, 120.

20Zagzebski, Virtues of the Mind, 168.


22Zagzebski, Virtues of the Mind, 270. Parenthetical numbers added.

23Zagzebski defines virtuous motivation as “a disposition to have a virtuous motive” (270). I accept her definition and will be using “virtuous motivation” in her sense of the word throughout this essay.
she means that one is motivated by the final end of obtaining true belief. She then defines knowledge in terms of acts of virtue: Knowledge is “a state of cognitive contact with reality arising out of acts of intellectual virtue.”\textsuperscript{24} As I am primarily concerned with propositional knowledge, I stipulate that from this point on, unless otherwise noted, whenever I use the term “knowledge” I am referring propositional knowledge.

**Truth Is Not Always the Motive for Propositional Knowledge**

I will now argue that one's underlying motive for being motivated by the intellectual virtues is irrelevant to whether one knows. This means that the truth motivation which is so essential to Zagzebski's account of intellectual virtue is likewise irrelevant, insofar as this motive is construed as a motivation underlying one's immediate motivation by such intellectual virtues as epistemic diligence or open-mindedness. As a first step in my argument, I draw attention to two cases which defy Zagzebski's claim that being motivated to know [the truth] imparts the additional value on true belief which makes it knowledge. In each case, I will distinguish between underlying and immediate motivations and their corresponding final and immediate ends, since making this distinction is necessary in order to see why Zagzebski is wrong about the source of the value of knowledge over and above that of true belief.

The Praise-Loving Scientist. I cultivate my intellectual virtues so that my colleagues will hold my beliefs in high esteem, so that they will praise me and adore me. I am reliably open-minded, intellectually fair, humble, diligent, adaptable, and so on, whether or not I think others will notice, since I believe that this is the best way to gain praise and adoration. I have cultivated my intellectual virtues for long enough that my motivation for praise and adoration is actually unconscious much of the time. In short, I am intellectually virtuous in the way in which we usually think of a person as being intellectually virtuous. With each act of virtue $Z$, \textsuperscript{[1]}\textsuperscript{25} I am motivated by the desire for praise or adoration to be motivated by the intellectual virtues, \textsuperscript{[2]} I act as a virtuous person [probably] would in the circumstances, \textsuperscript{[3]} I achieve the immediate and final ends of my action, that is, I act virtuously and obtain praise and adoration, and \textsuperscript{[4]} I acquire true belief as the result of 1–3. Are the true beliefs I come to possess through acts of virtue $Z$ knowledge?

The motivational component of $Z$ is composed of a virtuous motivation to satisfy the end of a particular virtue, open-mindedness, for instance, which is itself motivated by the desire for praise and adoration to satisfy the end of praise or adoration. Zagzebski takes the final motive and final end as primary in her analysis of the motivational and success components, as the source of the value of knowledge over and above true belief. Because my final motivation is not the desire for truth, nor is its end truth, she must maintain that I

\textsuperscript{24}Zagzebski, *Virtues of the Mind*, 270.

\textsuperscript{25}I have removed reference to “as a virtuous person would be” in \textsuperscript{[1]} so as to avoid begging the question of which part of the motivational component is necessary in order for one to be intellectually virtuous. Likewise for the case of the skeptical philosopher on page 8.
do not come to know through acts of intellectual virtue Z. Yet I do seem know. One might be
suspicious in light of my underlying motive that I have not been as motivated by the virtues
as I appear to have been. Perhaps this suspicion results from our intuition that in using
intellectual virtue as a means to ends other than the truth I am likely to put less effort into
being virtuous than I would if I used them as a means to the end of truth. But in the event
that I really have been motivated by the intellectual virtues, our intuition tells us that my
true beliefs count as knowledge in spite of the fact that my desire for praise and adoration
underlies my virtuous motivation.

The Skeptical Philosopher (Skeptic about knowledge). I cultivate my intellectual virtues
because I believe them to bring me not true beliefs, but only well-justified beliefs [although
they usually do in fact bring me true beliefs]. I am reliably open-minded, intellectually
fair, humble, diligent, adaptable, and so on, because I believe that this is the best way to find
strong justification for my beliefs. As with the praise-loving scientist, I have cultivated my
intellectual virtues for long enough that my motive of forming well-justified beliefs is
actually unconscious much of the time. Hence, I am intellectually virtuous in the way in
which we usually think of a person as being intellectually virtuous. With each act of virtue
Q, [1] I am motivated by the desire to obtain strongly justified belief to be motivated by the
intellectual virtues, [2] I act as a virtuous person [probably] would in the circumstances, [3]
I achieve the immediate and final ends of my action, that is, I act virtuously and obtain
strongly justified belief, and [4] I acquire true belief as the result of 1 – 3. Are the true
beliefs I come to possess through acts of virtue Q knowledge?

In this case, as with the praise-loving scientist, because my final motivation is not the desire
for truth, nor is its end truth, Zagzebski must maintain that I do not come to know through
acts of intellectual virtue Q. But in this case, even more than that of the praise-loving
scientist, it seems odd to say that I do not have knowledge. This might be because the
object of my underlying motive is traditionally thought of as a necessary component of
knowledge, whereas the underlying motive of the praise-loving scientist is not. But without
assuming that truth is the more valuable or only valuable underlying motive, we should be
reluctant to conclude that my true beliefs are any more knowledge than those of the praise-
loving scientist. Again, there is no intuition to deny that the true beliefs I have gathered
through the acts of intellectual virtue Q are knowledge. Quite the contrary, our intuition is
to say that I do indeed have knowledge, in spite of my belief that I do not.27

26The skeptical philosopher in this example does not withhold assent to strongly
justified belief. He instead believes beliefs which are strongly justified, but without holding
them to be true.

27I recently spoke with Zagzebski about my concern that one’s motivation for truth
need not be the final motivation in order for one to know. She agreed that this seemed
plausible and directed me to a 2009 publication of hers entitled On Epistemology. In this
text she departs notably from her position in Virtues of Mind that one must have an
underlying motive for truth. She now suggests that while one deserves greater praise for
coming to know as a result of being motivated by the truth as an end in itself, one can also
come to know in cases where truth is not the final motivating end. The motive of truth need
I conclude from these counterexamples that the motivation for truth need not underlie one’s virtuous motivation in order to know. But I think a stronger claim is also warranted. If such disparate underlying motives as the desire for praise and adoration and the desire for well-justified beliefs can be sufficient for knowledge [in conjunction with 2 – 4 in acts of intellectual virtue Z and Q], it seems plausible that which underlying motive one has is of little or no relevance to whether or not one knows. Hence, Zagzebski’s definition of an act of virtue ought to be revised so as to remove any reference to a particular motivation underlying one’s being motivated by the intellectual virtues. Moreover, the removal of a final motivation for being motivated by the intellectual virtues from the motivation component of acts of intellectual virtues necessitates a corresponding removal of a final end for acting virtuously in the success component (viz., obtaining truth).

Zagzebski has a possible response to the kinds of counterexamples I have given. Recall that for Zagzebski we are often unconsciously motivated to obtain the truth and that in many cases our belief-forming behavior is best explicable in terms of a motive to get the truth. She might then argue that there is a deeper, unconscious motive to get the truth at work in my counterexample cases, that the best explanation of the praise-loving scientist and the skeptical philosopher’s underlying motive is the further underlying, unconscious motive to get the truth. In so doing she could render the counterexamples I have given innocuous to her theory.

It seems plausible enough as I noted in my exposition of this point in the first section of this essay that the best explanation of our belief-forming behavior in many cases is an underlying, unconscious motive to get the truth. However, such a motive is not, as Zagzebski must hold given her definition of knowledge, the best explanation for every instance of virtuous belief-forming behavior which leads to knowledge. Consider the praise-loving scientist and the skeptical philosopher. A conscious motive already provides a good explanation of their virtuous belief-forming behavior: It is because the praise-loving scientist desires praise that he is motivated by the intellectual virtues, and hence acts virtuously in his belief-forming behavior, and it is because the skeptical philosopher desires strongly justified beliefs that he is motivated by the intellectual virtues, and hence acts virtuously in his belief-forming behavior.

only be present somewhere in one’s chain of motivation in order for the motivational component of an act of intellectual virtue to be satisfied (Linda Zagzebski, On Epistemology (Canada: Wadsworth. 2009), 126 – 27). Although I have not had time to fully digest the implications of her newly modified position on my critique, I will make a preliminary assessment. While I think Zagzebski’s modified motivational requirement is an improvement over the requirement as stated in Virtues of Mind and that it effectively diffuses the threat of my praise-loving scientist counterexample, I also think the threat from my skeptical philosopher counterexample remains. The skeptical philosopher is not motivated by truth at all; he is instead motivated by the desire for strongly justified belief. It seems that he comes to know without having any truth motivation whatsoever. Zagzebski’s modified motivational component also fails to explain why we blame the truth-motivated lucky guesser for guessing (See pages 11 – 12 in this essay).
It is not clear how the addition of a deeper, unconscious motive to get the truth could improve my explanation of the scientist or the philosopher’s virtuous belief-forming behavior if one does not already presume that motive must be present in order for them to know the true belief they obtain from that behavior. Zagzebski needs to argue why the addition of this unconscious motive makes the explanation of their belief-forming behavior already available via examination of conscious motives even more satisfying without presupposing its intrinsic value in order to demonstrate that such a motive provides the best explanation. Furthermore, since the link between conscious motives and other motives or actions is much easier to establish than the link between unconscious motives and other motives or actions, the former seem less speculative a basis for explaining virtuous belief-forming behavior. Hence, if Zagzebski were to argue convincingly that the underlying motive for truth provides a better explanation, the speculative nature of her claim nevertheless gives us reason to doubt that there is such an underlying motive in every case where one’s belief-forming behavior results in true belief which counts as knowledge.

Where does the Value of Propositional Knowledge Come From?

If the underlying motive driving one’s virtuous motivation is, as I have argued, of no relevance to whether one knows a proposition, then the value of knowledge must come from elsewhere. I now argue that the value of knowledge above and beyond true belief issues one’s immediate virtuous motivation, that is, from one’s being motivated by the intellectual virtues.

The Truth-motivated Lucky Guesser: Jim cultivates what he takes to be intellectual virtues, among them, guessing, from an underlying motivation to get at the truth. But guessing is not a virtue, nor are any of the other so-called virtues which he cultivates. Since the so-called virtues he cultivates do not reliably lead to true beliefs, he is usually unable to obtain true belief. Furthermore, most of his epistemic community is likewise motivated to get the truth but thoroughly intellectually vicious. When confronted with the same parking lot Zagzebski’s lucky guesser is confronted with, Jim correctly guesses that there are 564 cars in the lot. His true belief that there are 564 cars in the lot does not come from an act of virtue on Zagzebski’s or my own account, and therefore, he does not know that true belief. We would certainly blame him for guessing in this case. But our reason for blaming him cannot be, as Zagzebski argues it is in her lucky guesser case, that he lacks the underlying motivation for truth since he in fact has this motivation.

In this case his motivation cannot lead him to learn which procedures are reliable in his epistemic community since such procedures are missing from that community. Nor is it plausible that the source of value is some other underlying motivation, since modified lucky-guesser cases of this kind could be constructed for any other underlying motive and yet we would still blame him for guessing. Furthermore, as Zagzebski has already argued, the primary source of our blame cannot be his failure to adopt reliable procedures or to instantiate proper function, since a true belief which results from reliable procedure or proper function does not obtain its value from that procedure or function. Nor do we blame him because he does not have knowledge: We would not blame someone who satisfies 1 – 3 in Zagzebski’s definition of act of virtue and yet fails to obtain true belief. What is left then
is that we blame Jim for not being motivated by the individual virtues. The fact that we blame Jim for not being motivated by the individual virtues reveals that being so motivated is the source of the value added to knowledge over and above true belief.\(^{28}\)

I conclude that it is plausible that the motivational component of A need only consist in one’s being motivated by intellectual virtue(s) at some point in one’s chain of motivation in order to satisfy the motivational requirement for an act of intellectual virtue. Since one must obtain the end of one’s virtuous motivation if one is to act virtuously and virtuous motivation consists in one’s being immediately motivated by the intellectual virtues, satisfying the success component is a matter of fulfilling the end(s) of one’s immediate virtuous motivation, being open-minded or appealing to proper authority, for instance. Importantly, the success condition understood in this way does not reduce to doing what a virtuous person would (probably) do under the circumstances, since one may act as a virtuous person (probably) would in a given situation and yet not satisfy the end of her virtuous motivation. There are situations, for instance, in which a virtuous person who is motivated to appeal to proper authority would (probably), in spite of exercising great diligence in determining what is the proper authority, nevertheless appeal to improper authority.

My account also provides a plausible explanation for why Zagzebski’s bathroom scale buyer does not know that she has lost 5lbs when she obtains the true belief that she has. It is not that she lacks the underlying desire to get the truth, as Zagzebski argues, but rather, because she is not motivated by the intellectual virtues. The virtue in question here would be appeal to proper authority. When Zagzebski looks to the meter on her scale to see how much weight it says she lost, her motivation for doing so is not the desire to fulfill the end of appealing to proper authority, but rather, the end of appealing to any source of information which tells her she has lost 5lbs, since this will satisfy her underlying motive of

\(^{28}\)Montmarquet’s example of the truth-motivated dogmatist also supports my suggestion that our reason for blaming an individual such as Jim (see next paragraph) in his lacking immediate virtuous motivation. Montmarquet notes that “we can imagine an “epistemic fanatic” who is not epistemically virtuous at all (e.g. an extreme dogmatist, absolutely convinced of his possession of the truth, absolutely convinced that his methods of study of some sacred text are every day bringing powerful new truths)” (James A. Montmarquet, *Epistemic Virtue and Doxastic Responsibility*, ed. Paul K. Moser (Landham, MD.: Rowan & Littlefield Publishers, Inc., 1993), 22). Such a person is motivated by the desire for truth, what Montmarquet calls the virtue of epistemic conscientiousness, but because he possesses none of the other intellectual virtues, he is not epistemically virtuous. Hence, the motivation for truth is not a sufficient motivation in order for one to be epistemically virtuous. Furthermore, we blame such a person because he lacks the proper ways of being conscientious, such as open-mindedness and intellectual courage, what I have been referring to as immediate intellectual virtues (Montmarquet, 23). Without these virtues, the central virtue of epistemic conscientiousness does not guarantee that one is doxastically responsible, that is, that one is free from blame for not having obtained beliefs through intellectual virtue. Nor does it guarantee, in the context of Zagzebski’s definition of acts of intellectual virtue, that the true beliefs resulting from such acts are knowledge.
getting the belief that she has lost the weight. Hence, she is does not have the virtuous motivation necessary in order to act virtuously and does not know according to my definition of knowledge. If all other details of Zagzebski’s bathroom scale buyer example were held constant save the addition of an immediate virtuous motivation, that is, if she were motivated by the virtue of appealing to proper authority to buy an accurate scale, she would know that she had lost 5lbs regardless of what motive might underlie that virtuous motivation.

Locating the value from which the value of other components of acts of virtue and knowledge are derivative in one’s being motivated by the intellectual virtues has some noteworthy advantages. First, doing so turns my counterexamples into supporting examples. As demonstrated by the cases of the praise-loving scientist and the skeptical philosopher, any number of motives may motivate one’s virtuous motivation and yet one still come to know if one is immediately motivated by the virtues, and as demonstrated by the truth-motivated lucky guesser case, without an immediate virtuous motivation, the seemingly virtuous underlying motivation of desire for truth cannot lead to knowledge. Second, this relocation of value resists Zagzebski’s argument against process and faculty reliabilism since it too identifies motivation as the source of value. Finally, it seems to comport with our intuitions about what counts as knowledge better than does Zagzebski’s account of intellectual virtue, as my counterexamples also demonstrate.

Conclusion

I have argued that the motive of truth need not, as Zagzebski claims, underlie one’s virtuous motivation in order for true beliefs obtained through acts of virtue to count as knowledge and that motives underlying one’s virtuous motivation are in fact entirely irrelevant to whether one knows. In lieu of these arguments, I have proposed a plausible alternative source for the value of knowledge over and above true belief, namely, one’s immediate virtuous motivation, and argued that there is good reason to believe that it is in fact the source of this value. Much work needs to be done in order to fully defend this proposal. But it should at least be clear that one’s immediate virtuous motivation is a better candidate for the source of the value of knowledge over and above true belief than is the underlying motive of truth.

29The question of whether one’s immediate virtuous motivation needs to be conscious in order for one to know is an important one. It seems likely that in many cases we do indeed appeal to proper authority without a conscious virtuous motivation. When I look at the bus schedule to see when my bus will arrive, for instance, I may do so out of habit, and not because I think that in doing so I am appealing to proper authority. But if I wish to maintain that I come to have knowledge in such cases, I will need to defend my own account against the objection I raised earlier against Zagzebski’s insistence on an unconscious motive for truth. I am not prepared to give this defense just yet, but I do not think that this is an impossible task.
Michael Granado – G. E. Moore Against the Skeptics

At this very moment I am sitting on a couch writing an essay. Most people would agree that I can know the former sentence with certainty…but can I really? Skeptical arguments, such as the ones presented by Descartes and Hume, seek to show that there are reasons to doubt even the most basic beliefs and truth claims. But are these skeptical arguments powerful enough to call into question even common sense beliefs, such as the one that was made in the first sentence? In his article, *A Defense of Common Sense*, G.E. Moore argues that common sense beliefs decisively refute skeptical arguments. Are Moore’s arguments valid, thereby defeating skepticism, or is Moore guilty of some logical error, namely that of begging the question?

Philosophical skepticism insists that one may never be certain that either the beliefs that one holds or truth claims are in fact true. The skeptic claims that there is reason to doubt all of our truth claims, and various arguments are given to support this view. In his *Meditations on First Philosophy* Descartes entertains two famous skeptical arguments, firstly the dream argument, and secondly the evil God argument. Both of these arguments claim that there is no way to distinguish between an illusion, or a deception, and a true perception. Another famous skeptic, David Hume, argues that there is ultimately no way to know for certain if our ideas of reality actually resemble what reality is. He goes on to argue that there is no satisfactory proof that can be given that justifies one’s belief in an external world. Hume argues that the only reason there is a continued belief in the external world is because of the “constancy and the coherence of impressions.” Most people however would object to the skeptic’s arguments. After all, it seems blatantly obvious that there is an external world, but is there any proof? In *A Defense of Common Sense* Moore claims to present a proof that defeats skepticism by proving that there can be certain knowledge about the external world. Moore’s strategy consists of refuting his opponent’s view, in this case that of skepticism, and from this refutation he finds support for his own position. Moore begins his argument by giving a list of what he calls “truisms,” or propositions that he knows with certainty to be true. The truisms, of which he gives many


examples, are considered to be common sense propositions, for example “there is a pen.” One problem that Moore faces at the beginning of his essay is, how are common sense propositions identified? Are there any particular conditions that a proposition must satisfy in order to be considered common sense? To help answer this question it will be helpful to examine the different kinds of truisms that Moore lists in his opening section.

The first set of truisms that Moore lists are ones that are primarily concerned with material objects. The list includes objects such as Moore’s body and the surface of the earth, but it also includes everyday objects such as a mantelpiece, a bookshelf, and a pen. The second set deals with particular experiences that Moore has and the experiences of other people. Thus, the second set primarily deals with nonmaterial things. It should also be noted that the second set includes the experiences of other people. Moore then asserts that most, if not all human beings, have known propositions that are similar to the ones he listed above. The only difference is that the propositions pertain to the particular individual rather than Moore.

Although Moore did not set out criteria for defining common sense in *A Defense of Common Sense*, he does provide the criteria in *Some Judgments of Perception*. In this work Moore argues that statements, such as the truisms listed above, are what he calls “judgments of perception,” and these judgments of perception are a part of what he considers to be common sense. Moore argues that a judgment is a judgment of perception if and only if it meets the following criteria. First, it must be stated, or it at least must be able to be stated, in such a way that it expresses the existence of a physical object. An example of this would be when Moore claims that “this is a bookshelf.” Any time someone formulates the statement, “that a is b” they are making a judgment of perception. Secondly, the statement that “a is b” must be grounded in the perceiving “that a is b.” In other words, “this is a bookshelf,” must be based on perceiving that the object is indeed a bookshelf. Thus, in perceiving that “a is b” it is certain that “a” is indeed “b” and therefore that all judgments of perceptions are true. So, when Moore says “this is a bookshelf” he knows this to be true, along with all the other statements that he calls truisms because they all could be classified as judgments of perceptions.

To make this point clearer Moore uses the example of a soap-bubble. When Moore says that “there is a soap-bubble,” is a true proposition, then it follows from this proposition that “there is an external object” or that “there is something to be met with in space”. Furthermore, when Moore says that “this is a soap-bubble” it is possible for that soap-bubble to have existed before the statement was made or after it. In other words, the continued existence of the soap-bubble does not depend on its being perceived. When

---


7Klemke, *Defence of Realism*, 136.

someone says, “that is a soap-bubble” it follows that “there is an external object,” and to be an external object the soap-bubbles’ existence must be logically independent of that person’s perception, or else it would not be an external object. From this argument it seems reasonable to conclude that there are objects external to the mind, not just one particular mind, but every other mind as well. Moore now claims that he can provide numerous proofs for the existence of the external world based on this formula.

Moore argues that if it is possible to have a pair of objects that are distinct from one another and to formulate them in such a way as to be a judgment of perception, then it would be possible to prove that the pair of objects were external objects existing in an external world. Moore uses his own hands to prove this point. Moore holds up both of his hands, and making a certain gesture with his right hand says, “here is one hand,” then doing the same thing with his left hand states, “and here is another”. Thus, two external things exist, namely Moore’s hands. While this proof may seem ridiculous, Moore insists that it has met all the conditions of a rigorous proof. The first condition for a rigorous proof is that the premises must be different from the conclusion. Moore says that this condition was met because both of the hands were different from one another, and even if these premises were false it still could have been possible that the conclusion was true. The second condition is that the premises are known to be true, and Moore claims to know with certainty that he has hands. Moore claims that it would be foolish to assert that he did not know for certain, but only believed that he had hands when he made the statement “here is one hand,” for this statement is obviously true. The third condition is that the conclusion follows necessarily from the premises. If there is indeed a pair of hands, then it follows that there are external objects, and thus the third condition is meet. If Moore’s proof is correct then skepticism is proven to be false. For if skepticism was true, then a person could not have knowledge of the outside world. However, a person can have knowledge because one can know for certain that he or she has hands. Therefore, skepticism is false.

Many objections have been raised against Moore’s argument. This paper will only address two. The first objection is that Moore has provided no reason to believe that his premises are true. In the previous paragraph one of the premises for Moore’s argument was “here is one hand.” The first objection claims that if Moore cannot prove the existence of the hand in the first place then he cannot claim to know that the premise is true. Thus, the skeptic who demands proof for the existence of hands remains unsatisfied, for Moore has not provided this sort of proof. In fact, Moore admits that he cannot provide an answer to this objection, but responds by claiming that he can still know things despite not being able to prove them. In his reply Moore uses the idea of subjective probability. Subjective probability is the

---


amount of confidence one has for a particular proposition. Throughout the day everyone has a varying degree of confidence in many different propositions. For example, I am confident that when I save this essay on my computer, the essay will still be there when I come back. Of course something may go wrong, for example, someone could delete my paper, but in general I am pretty confident. I have an even greater degree of confidence that there is a computer sitting in front of me at this very moment, in fact I am certain of it.

The proposition “this is a computer” would be considered by Moore to be a common sense proposition. For Moore, the degree of certainty in common sense propositions, such as the one listed above, is so great that it amounts to subjective certainty. Any argument, such as a skeptical one, that argues against the subjective certainty of a common sense proposition will have to use a premise that has less subjective probability than the common sense proposition. Take for example Descartes’ evil God argument. Could there be an evil God who at every moment is deceiving me to believe that there is a computer in front of me, when in fact there is not? This could be true, but at this moment it appears that there is a computer in front of me, and this has more subjective probability than the evil God argument. It is the subjective certainty of propositions that allows Moore to claim throughout the article that “I know with certainty” that his common sense propositions are true. Additionally, he usually prefaces his common sense statements with “in my own opinion,” and because he phrases his knowledge claims in this way he is expressing his own subjective certainty. Moore is convinced, despite any argument that a skeptic might bring up, that he is right. Moreover, Moore expresses that other people are subjectively certain of the same propositions that he is certain of. The fact that Moore has subjective certainty of common sense propositions and the fact that he is certain that other people are subjectively certain of common sense propositions is the main foundation of Moore’s argument. So, the reason Moore can say that he knows that he has hands is because he is subjectively certain that it is true and because it is more probable than any other option.

The second objection against Moore is that his argument involves a logical fallacy, circular reasoning. It seems that one of the premises that Moore uses is the same as his conclusion. The objection states that Moore begins his argument by claiming that his list of truisms is true. Moore then states that people who hold views that contradict his truisms, such as the skeptic, are wrong. All skeptical views are therefore wrong while the right view is that Moore’s list of truisms is true. Therefore, Moore’s conclusion is that his truisms are true and thus he commits circular reasoning. However, this is a misrepresentation of Moore’s main argument. It was stated at the beginning of this paper that Moore’s strategy was to refute the skeptic’s view and by doing so establish his own views. Moore begins A defense of Common Sense by stating his conclusion, which is that he can be certain that his truisms are true. It must be made clear that his truisms are his conclusion, not one of his premises. Moore then proceeds to argue that skepticism is wrong, by arguments such as the soap-bubble example. With the soap-bubble example Moore shows that it is indeed possible to

---

11Ferguson, Common Sense, 134.

12Ibid., 135.
have knowledge of the external world, and because knowledge of the external world is possible it proves that the skeptic's views are false. So, the views that disagree with Moore’s list of truisms, the skeptic’s views, are false. Therefore Moore’s list of truisms must be true, according to the argument that one can have knowledge of the external world. From what has been stated above it seems that the objection to Moore’s argument, that he committed circular reasoning, is a misrepresentation and that Moore did not intend for his argument to be stated in that way.
Let us weight the gain and the loss in wagering that God is. Let us estimate these two chances. If you gain, you gain all; if you lose, you lose nothing. Wager, then, without hesitation that He is.
– Blaise Pascal, Pensées

C.S. Peirce intended philosophical pragmatism to be a method for clarifying the meanings of difficult to understand terms, and he took this understanding of pragmatism to be uncontroversial among all those who would call themselves pragmatists: “[Pragmatism] is merely a method of ascertaining the meanings of hard words and of abstract concepts. All pragmatists of whatsoever stripe will cordially assent to that statement.”¹ More specifically, Peirce conceived of this method of linguistic meaning-finding as fundamentally experimental in nature: “All pragmatists will further agree that their method of ascertaining the meanings of words and concepts is no other than that experimental method by which all the successful sciences...have reached the degrees of certainty that are severally proper to them today.”² Because the pragmatic method is experimental, we ought to conclude – and Peirce does conclude – that pragmatism is necessarily oriented towards and concerned with that which is tangible and practical since these things are the possible objects of scientific experimentation.

But if we accept that pragmatism is primarily an experimental method for meaning clarification that is concerned with practical matters, it follows that most true pragmatists would tend to be antagonistic to outlooks that encourage non-evidentiary methods of belief formation as well as to outlooks that trade in the inarticulate and the unobservable. Bearing this in mind, it should seem odd to us if any genuine pragmatist were to offer up a defense of religion since, through the doctrine that faith is virtuous, many, perhaps most,

¹Pragmatic and Pragmatism, 57

²Ibid. Peirce may have overstated the case when he said that all pragmatists will assent to pragmatism thus conceived, but even the fellow pragmatist who he felt most gravely misunderstood pragmatism – William James – has a body of work largely centered around the theme of meaning-finding in one way or another. His theory of truth, for example (that we engender truth upon reality by framing reality in terms of our own interests) is little more than an application of Peirce’s thought to the term “truth” insofar as he says that whenever we ask ourselves what the truth of the matter is in any case, what we must do is first understand the meaning of what we are asking by determining what concrete interests we have in mind when we ask.
religions promote non-evidentiary belief formation about the inarticulate and unobservable as much as or more than any other major life outlook. Yet offer a defense of religion is just what a pragmatic light no less bright than William James did in a number of philosophical lectures and written works and what he did most famously in an address to the philosophical clubs of Yale and Brown universities. This address has come to be called “The Will to Believe.” In it, he argues, in a situation of epistemological indeterminacy (i.e. a situation in which nothing coerces our logical intellect to make a decision on a matter of genuine controversy), we have a right to choose our beliefs according to the dictates of our passions. In this essay, I intend to be critical of this position, drawing in part from William Clifford’s work “The Ethics of Belief” as well as from the fundamentals of pragmatic theory itself. My criticism of James’ voluntarism will be tripartite. The central prong of this trident criticism will be to question the ethical responsibility of James’ claim of the epistemic legitimacy of faith in light of the pragmatic view that belief is, in every case, tied up with action. This line of argument will be flanked by two supporting lines of argument that call into question the existence of actual Jamesian cases of indeterminacy, respectively asking whether James’ proposed state of epistemological indeterminacy ever naturally obtains in real situations and whether we can meaningfully choose our beliefs anyway, regardless of the presence of an indeterminate state. First, however, I will briefly explicate the important parts of the central thesis of “The Will to Believe” to set the stage.

The Skeleton of “The Will to Believe”

In his own words, James considers his essay to be "...a justification of faith, a defense of our right to adopt a believing attitude in religious matters, in spite of the fact that our merely logical intellect may not have been coerced." To understand what he means by this takes a little digging into James’ concepts of what an uncoerced logical intellect is as well as what the “religious matters”, about which we may adopt a believing attitude, are.

To one who gives “The Will to Believe” a cursory reading or to one who only knows of the text second-hand, James suggestion that we have a right to believe as we wish when our

3Strangely, James is not alone in this among the pragmatists, though he is arguably the pragmatist who most famously dabbled in religious apologetics. Even Peirce himself occasionally had good things to say about religious belief, though it should be noted that these remarks (such as his “argument” for God in “A Neglected Argument for the Reality of God”) are often regarded by scholars of pragmatism as confused at best, not to mention emerging from a method of thinking that is divergent from the orthodox pragmatic one.

4“Genuine” as it is used here is a complex technical term of James’. Its meaning will be explained in the next section.

5Actually, James’ statement is stronger than this, although it is often read as weaker. He says we must choose according to our passions in such a situation.

6The Will to Believe, 186
intellect is uncoerced sounds like an invitation to believe whatever we like, whenever we would like to. But this is a misunderstanding of James and it is one so widespread that “The Will to Believe” has become perhaps the most caricatured work in the entire body of pragmatic philosophy. The misunderstanding commonly results from ignorance about the kinds of situations of choice that James limits his recommendation to. In section I of “The Will to Believe”, James makes it clear that his position is that while we do have the right to believe as we wish, we have that right only when faced with what he calls a “genuine” option. The term “genuine” here should not be understood in the common way, so that we end up thinking that James’ “genuine” option means only an option that is “real” in some way, as opposed to merely “apparent.” While it is true that “genuine” options for James must be real or “live” and not just apparent options, there are two other conditions that must be met as well if an option is to reach the status “genuine.” Options must also be “momentous” and “forced.”

James contrasts “momentous” options with “trivial” options. Momentous options are high-stakes, unique opportunities that cannot be easily undone once chosen. The opportunity to marry someone is an example of an option that is not momentous because although the option is high-stakes and hard to undo, it is not a unique opportunity. The opportunity to marry a particular person that you love dearly, on the other hand, is momentous because such an opportunity is the only one of its kind (or is one of just a small few).

A “forced” option for James is simply one that is logically unavoidable. With a little help from President Bush, we can illustrate how a forced option differs from a merely apparently forced option. A week and a half after the attacks on the United States of September 11, 2001, George W. Bush famously said to the nations of the world about the coming “war on terror”: “Either you are with us, or you are with the terrorists.” Assuming that he did not mean that all nations not supporting U.S. policies shall be considered as being “with the terrorists,” bush’s option to the world was, logically speaking, an avoidable one. It is possible (indeed, it is nearly certain) that some nations neither backed the terrorists nor backed the United States in the subsequent “war on terror.” A more appropriately worded...

---

7The first of three conditions that make an option “genuine” is that the option must be what James calls “living” as opposed to “dead,” which is to say that each side of an option must make at least SOME appeal to your intellect as a viable choice. An example of a living religious option for a person living in the contemporary West is that of choosing between being a Christian and being a Buddhist, because a person living in the contemporary West may have been taught to a certain degree about the ways of both religions, knows people who have adopted both of these religions as a way of life, and feels tempted, even if in small ways, to further investigate both religions to see whether they are worth adopting. An example of a dead option to the same person might be the choice between being a Christian and being a Zoroastrian. Zoroastrianism, unlike Christianity, is likely to be an idea that is completely foreign, is an idea that has not been adopted by anyone the Westerner knows, and generally, as James puts it, is an idea that “refuses to scintillate with any credibility at all.” (Ibid, 187)

8Although, this may have actually been the intended message at the time...
forced option to the world (logically speaking) would have been “Either you are with us or you are not with us,” though one can see why such a wimpy ultimatum was never given.

In any case, because James’ restricts himself to only genuine cases (again, these are cases with options that are live, momentous, and forced) in which our intellect is uncoerced, he is able to make the following astute observation:

Our passional nature not only lawfully may, but must, decide an option between propositions, whenever it is a genuine option that cannot, by its nature, be decided on intellectual grounds; for to say, under such circumstances, “Do not decide, but leave the question open,” is itself a passional decision – just like deciding yes or no – and is attended with the same risk of losing the truth.9

In other words, because a genuine option is a forced option, even deciding to put off a decision remains a definite stance because the world does not wait for us to make up our minds. We might liken this to the state of affairs that obtains during a time of political election.10 In political elections, people choose who their leaders will be by employing various criteria. Some decide according to what a candidate will do for them, personally. Others decide according to what a candidate will do for the wider community. Still others decide according to a candidate’s personal liability or according to a stance on a pet issue. Some simply decide according to political affiliation. Some even decide to let others, such as a spouse or a religious authority, decide for them. Finally, there are also those who, for whatever reason, decide not to vote at all. But although this last alternative may at first appear to result in no vote, in practical terms, it results in a vote like any other. Not to vote is, intentional or not, a verdict submitted in favor of whichever candidate the other voters choose.

Similarly, to James’ mind, it is not practically open to one who meditates on religion to remain agnostic about the existence of gods until more data is in. One who attempts this, says James, is actually choosing to let the reigning conditions determine his fate in the same way that the non-voter has done. James supports this idea by noting that, if the entities supposed by religion really do exist, waiting to believe in them until more evidence is in may preclude one from obtaining that evidence before it is too late to matter. For example, many in the Christian tradition hold that evidence of the reality of their god can only be had in the afterlife. If this is the case, then only belief without evidence during life will put one in the epistemically desired position afterlife. Thus, says James, the better position with

9(The Will to Believe, 193 – My italics.) In a way, this amounts to a refutation of Buridan, who said that a person faced with alternative courses of action must always choose the apparently greater good and allowed that, in cases of uncertainty, the decision could be postponed indefinitely. James would say of that person that if there is no rational basis for choosing one course of action over the other, the passions will make the decision anyway in some cases (forced cases).

10I owe this analogy to John Smith. (72)
regard to religion, given a state of total ignorance, is probably to wager, as Pascal did, that God exists and to believe in him while still alive in order to have the chance of later making His acquaintance. However, this suggestion suffers from the same old problems that Pascal's wager suffers from. Most notably, one might object that there are any number of religions competing for one's endorsement, many of them mutually exclusive,\textsuperscript{12} and that therefore, this choice seems like it is not the simple choice between religious belief and religious nonbelief, but one between religious nonbelief and religious belief in some particular religion chosen from among a host of different religions. Furthermore, if the intellect is truly uncompelled in religious matters (as it must be if James' argument is to apply), any particular religion is as unlikely to be correct as any other, so the choice between religions must be made arbitrarily. Clifford puts the problem like this:

The Prophet tells us that there is one God, and that we shall live for ever in joy or misery, according as we believe in the Prophet or not. The Buddha says that there is no God, and that we shall be annihilated by and by if we are good enough. Both cannot be infallibly inspired; one or other must have been the victim of a delusion, and thought he knew that which he really did not know. Who shall dare to say which? [A]nd how can we justify ourselves in believing that the other was not also deluded?\textsuperscript{13}

Thus, James' preference for faith is not really the preference for truth acquisition over error avoidance he characterizes it as (since it is not realistic to think that arbitrary guessing from among a large number of alternatives will get us the truth), but is instead nothing more than a preference for activity over passivity with regard to decision-making about beliefs – a preference for a blind guess over no guess at all.

Although James' argument for preferring belief to nonbelief doesn't stand up, his assertion that we must believe according to our passions when faced with a genuine option that our intellect cannot help us with may be a different story. However, in spite of (or perhaps because of) the correctness of James' observation about the role of the passions in belief fixation, there are good reasons to think that such cases (hereafter referred to as “Jamesian cases”) do not actually exist – or if they do, they do not exist in any way that James should recognize as pragmatically meaningful. In the following two sections, these reasons will be explored.

\textsuperscript{11} This is similar to John Hick's idea of “eschatological verificationism”. Hick proposes that remaining agnostic about the Christian god's existence may preclude one from ever being able to verify that existence (because heaven will be closed to that person, according to many Christian teachings). (103-105)

\textsuperscript{12} “He is Allāh, besides whom there is none who has the right to be worshipped” (\textit{Qur'an}, Sura 59, Ayah 22) “You shall have no other gods before me” (\textit{King James Bible}, \textit{Exodus} 20:3)

\textsuperscript{13}\textit{The Ethics of Belief}, 8
The Left Prong – Beliefs are Never Chosen

Earlier, I mentioned the similarity between James and Pascal on the matter of religious belief and said that their positions suffered from some of the same problems. One of those problems, not mentioned so far, is the problem that both James’ and Pascal’s recommendations about religious belief seem to presuppose that beliefs are the kinds of things we can choose. In Pascal’s case, this presupposition has been widely criticized. If one begins as an atheist and then, after meditating on the logic of game theory, professes belief in God to maximize his advantage, surely we cannot take this as an instance of genuine belief, but only as one of belief in name only. It is a case of telling God what He is thought to want to hear, regardless of what is actually believed.\textsuperscript{14} Even James points out this flaw in Pascal’s argument in “The Will to Believe”:

...a faith in masses and holy water adopted willfully after such a mechanical calculation would lack the inner soul of faith’s reality; and if we were ourselves in the place of the Deity, we should probably take particular pleasure in cutting off believers of this pattern from their infinite reward.\textsuperscript{15}

However, while James improves upon Pascal in a sense by explicating the nature of choice, his improvement is what might be called a Pyrrhic improvement. When James says that our “willing natures” are the deciding factors in Jamesian cases, he means “…all factors of belief as fear and hope, prejudice and passion, imitation and partisanship, [and] the circumpressure of...caste and set”\textsuperscript{16} If this is his view of our “willing natures”, then James looks as if he must be making a descriptive statement about how beliefs are actually determined (i.e. they are determined by non-rational forces that are beyond our control) rather than making a prescriptive statement about how beliefs may legitimately be determined in “The Will to Believe.” Insofar as this could be James’ point, it is a good one. It seems true that in cases where we lack guiding evidence, the non-rational forces of our psyche must determine what beliefs we will adopt. However, note that these forces are also non-volitional. We do not get to choose what our fears, prejudices, passions, etc. will be when we enter into a situation of having to decide about an “uncoerced” genuine option any more than we get to choose the facts of the matter that coerce us when we enter into a situation of having to decide about a coerced genuine option.

\textsuperscript{14}Since Pascal thought of God as all-knowing, a strategy like this that involves deceiving God could not possibly have worked as Pascal wanted it to.

\textsuperscript{15}The Will to Believe, 189.

\textsuperscript{16}Ibid., 191.
However, one who pays attention to James’ language in “The Will to Believe” will probably come away from it with the impression that James intended for this work to be more a statement of what we may legitimately choose when given an option, rather than a mere statement of what we, in fact, choose when given an option. Thus, the first tension in James’ piece is laid bare.

The Right Prong – Our Intellect is Rarely (if Ever) Uncompelled

It has just been argued that there are no real Jamesian cases because we never get to freely choose our beliefs. Rather, beliefs are the kinds of things that are forced upon us, whether by reason or by the passions or by a combination of the two. James’ “genuine option” is thus not really an option at all if to have an “option” here entails being able to make a free choice. However, even if James’ “genuine option” were rightly called an “option”, it would also be worth considering what kinds of options leave a person’s faculty of reason uncompelled. Meditating on this, we may come to realize that such options may be more elusive than James would have us believe. Even where religious matters are in question, I would like to argue that there is usually (perhaps always) some evidence that compels our logical intellect, even if that evidence may not be conclusive.

To begin this argument, let us examine a centrally important question for many religions – the question of whether or not a god or gods exist. James seems to be of the mind that, with regard to a question such as this, our reason cannot be compelled because evidence about the gods is, for some reason, not available. But why suppose this to be the case at the outset? While Kant was able to reject most proofs of the existence of a god on these grounds, he was able to do so because he regarded that god as a transcendent being that could only be found beyond the limitations of human reason. However, one might expect that a god conceived of as a pragmatist would conceive of it would be much less out of human reach.

Furthermore, the people of James’ time for whom “The Will to Believe” was most relevant (i.e. those grappling with Christian beliefs in the face of the latest strong objections from both science and philosophy) normally conceived of God as able to be known, even if imperfectly.

The issue of what is to count as evidence for the existence of a god or of gods will of course be particular to the conception of god or gods that is being entertained, but with the exception of gods that are completely unknowable by definition, evidence both in favor of and against the existence of gods will always be available. Personal experience of a hard to

---

17Again: “I have brought with me tonight something like a sermon on justification by faith to read to you – I mean an essay in justification of faith, a defense of our right to adopt a believing attitude in religious matters.” (Ibid., 186 – Bold text is mine.)

18He was unable to reject his own moral proof.

19“[Pragmatist theists] believe that God is as real as sense-impressions, tables, quarks, and human rights.” (Religious Faith, Intellectual Responsibility, and Romance, 92)
explain nature, the testimony of millions, and the difficult to comprehend complexity of life and of the world may count as evidence in favor of a god. On the other hand, the fact that no gods have been observed under scientific conditions, the fact that the qualitative “feel” of religious experience can be created on command in the laboratory (by simulating neurological disorders in the brain of a subject), the queerness of the religious story, and the fact of religion’s human – and thus fallible – origins may count as evidence against a god. If there is a problem at all with reason compelling us in favor of or against a god, in the case of most people, that problem is not too little evidence for one side or the other, but too much evidence on both sides to know what to do with – a problem exacerbated by having a vague conception of the kinds of gods one is looking for at the beginning.20

———

20 James has a simple method for removing the insolubility from supposedly insoluble philosophical problems that he presents in “What Pragmatism Means” and it may be able to help us here. Whenever we are faced with a seemingly intractable philosophical problem, what we are to do is to simply clarify the motivations or interests that drive our examination of the problem. If we wish to know, for example, if there is a material and personal god who intervenes daily in the affairs of human beings, then the fact that no such god has been found in the world in scientifically credible way will count as evidence against the existence of such a god. If instead we wish to know whether there is an impersonal, immaterial god that is only revealed through private individual religious revelations, the fact that such a god has not been found in the world will not count as evidence against its existence and different facts, such as a preponderance of reports of religious revelations of this god, will count instead as evidence in its favor. That James sees the question of the truth of the proposition “the best things are the more eternal things” to be scientifically unverifiable may stem from his failure to apply his own recommendation of clarifying what is really being asked as he asks whether this religious hypothesis is true or not.
Nothing has been said, up to this point, about the specific religious hypothesis that James concerns himself with, so it is time to introduce that hypothesis. Aware of the different varieties of religious experience (as James’ book of the same name suggests), James surveys religious messages and distills down two assertions that he takes to be the foremost messages of all religions:

Religions differ so much in their accidents that in discussing the religious question we must make it very generic and broad. What then do we now mean by the religious hypothesis? Science says things are; morality says some things are better than others and religion says essentially two things.

First, she says that the best things are the more eternal things, the overlapping things, the things in the universe that throw the last stone, so to speak, and say the final word. “Perfection is eternal” – this phrase of Charles Secrétan seems a good way of putting this first affirmation of religion, an affirmation which obviously cannot yet be verified scientifically at all.

The second affirmation of religion is that we are better off even now if we believe her first affirmation to be true.21

James gives these two examples – the prime examples for him – of the kind of religious question that our reason cannot help us decide. Let us consider only the first, for the time being.22

We see that James writes, just after giving the first affirmation of religion, that it is not something that can be scientifically verified. But if James has trouble seeing how a statement like “the best things are the more eternal things” could be dealt with by reason, I would like to suggest that this is not because there are no facts for or against this hypothesis for the human intellect to take hold of, but rather because the hypothesis itself is vague to the point of being nearly meaningless. Of course it is clear that we are supposed to prefer the eternal things in some way,23 but that is all we know. In any proper statement concerning what is best, better, or good, there is always also a reference to what the thing in question is best, better or good for (unless what it is best, better, or good for is already understood). That is, we do not tend to say things like “gold is better” or “iodized salt is good” by themselves but instead say things like “gold is better at conducting electricity than other metals” or “iodized salt is good for preventing goitre”. Otherwise, we run the risk of having our messages importantly misunderstood, as they would be if they were interpreted as “gold is better because it is the cheapest conductor” and “iodized salt is good for your

21The Will to Believe, 204.

22Since the second affirmation of religion is parasitic on the first and since, as we shall soon see, the first affirmation is highly ambiguous, it is difficult to know what to say about the second affirmation by itself.

23Kwasi Wiredu once wondered in my presence if James had considered the prospect of eternal suffering when he wrote these words.
It is true that we sometimes simply say that something is good, better or best without mentioning what it is good, better or best for – as in the case “swimmer x is the best,” but it is usually clear what we mean in these cases. In the case of the swimmer, it is clearly meant that ’swimmer x‘ is the fastest swimmer and not that he is the safest swimmer or, for that matter, the most dedicated student. But in the case of “the best things are the more eternal things”, the context of the statement is not nearly so clear as that. Unless answers are given to questions like these, the religious hypothesis “the best things are the more eternal things” is not really a full hypothesis at all, but an incomplete one. If it means anything, it can rise to no higher level of meaning than mere complex interjection (an emotional sentiment that reads something like: “Hooray for the more eternal things!”) and so cannot be a proper object of belief. And since interjections are not beliefs and so cannot be considered for adoption or rejection, there is no real option between beliefs here (much less a “genuine” one), but only the appearance of an option.

However, it is possible that, rather than giving a specific religious hypothesis for us to accept or reject, James was merely laying out what he thought the general form of the most important religious message was when he wrote “the best things are the more eternal things” and that he left it to individual religions to fill in the details regarding the “what for” and “how” of the supremacy of the eternal things. But if this is so, then once these details are filled in by the different religions, it seems likely that our reason will once again have something (however small) to work with. For example, if a religion were to teach that the eternal things are best because one who regularly contemplates the eternal things will gain peace of mind, the religious hypothesis then becomes subject to scientific scrutiny (one test of such a hypothesis could consist of looking to see whether people who meditate on the eternal things have lower stress levels than those who do not). So, the dilemma that James’ particular religious hypothesis faces is simply this: Either it is not really a true hypothesis, or else it is a true hypothesis, but one that can be critically examined with the intellect.

There may be some exceptions to the rule I have in mind – that our intellect is always compelled in some way when a meaningful hypothesis is present – but most of these are probably trivial. For example, our logical intellect may have no way of judging (without further investigation) whether a stranger who walks into a bar is named John or Mark, but it seems unlikely that very much rides on this kind of judgment in most instances.

---

24Not only is it not clear what the eternal things are best for, but it is not even clear what the “eternal things” that James has in mind are. The soul (as opposed to the body), perhaps?

25Also see Anthony Flew’s “Theology and Falsification” for an excellent treatment of the problem of the vagueness of religious assertions.

26A person can perhaps still grasp at straws with his logical intellect in these kinds of cases – perhaps by already knowing that John is a more common name than Mark and then reasoning probabilistically or, in the absence of such knowledge, perhaps by
difficult to think of a pure case in which reason is in no way compelled to choose one hypothesis over the alternative and in which the stakes are also high. Perhaps the case of a police officer trying to tell if a stranger walking into a bar is a suicide bomber or not, is one such case. But even here, there are likely to be some clues for the logical intellect to work with (e.g. perhaps the suspect is wearing a heavy coat in warm weather, carries a backpack with protrusions or visible wires, is nervous, exhibits excessive sweating, displays an unwillingness to make eye contact, etc) and responsible officers of the law will do their best to look for these clues as thoroughly as they can before making the weighty decision to kill the stranger. In any case, as for deciding what our religious beliefs are going to be, there is surely much more time than in the previous case to look for clues before having to decide.

The Central Prong: The Will to Believe is Always Also the Will to Action

I have just argued that even if James’ “genuine option” is truly an option and not just an apparent option (as it seems to be), there are good reasons to think that there are still no Jamesian cases because in nearly every instance, our logical intellect is compelled in some way or other. Now I shall argue that even if it is also true that there are true Jamesian cases, and even if the option to have religious belief or not is one of those cases, the cost of deciding with the passions rather than waiting for the intellect to give its assent is often greater than James lets on.

Pascal was of the mind that while wagering in favor of God’s existence might win us “everything” if we are right (in the form of eternity in paradise after life), wagering so if we are wrong loses us nothing. Similarly, James believes that to choose religion over religious skepticism wins us the truth (and all the benefits that go with it) if we are right and loses us only our reputation (because we will look foolish) if we are wrong. However, the cost of losing such a religious wager may be significantly greater than this.

Let us once again examine James’ religious hypothesis. Immediately it strikes anyone who is at all familiar with world religions that James’ religious hypothesis is much simpler and aesthetically preferring one name to the other and then reasoning that other similarly constructed people might also prefer to give that name to their offspring.

27This exact behavioral profile of a suicide bomber was actually disseminated by an international organization representing police chiefs in July of 2005 with the recommendation that sudden and lethal force (in the form of a gunshot to the head) is warranted if an individual sufficiently fits this profile, even if the danger is not imminent. Two weeks later, Jean Charles de Menezes (an electrician who was unrelated to the bombings) was shot seven times in the head by plainclothes British officers looking for suspects in a failed bombing attack on the London transit system. Police claimed that Menezes was wearing a heavy jacket on a warm day (the temperature in London was about 62°F at the time) and ran from the police who chased him.
less demanding than the various hypotheses of actual religions. If James believes his religious hypothesis to be on par with other religious hypotheses, it should then come as no surprise to us that James believes the potential cost of religious belief is only the relatively light cost of being in error and looking like a fool. What should come as a surprise, however, is that James, the pragmatist, would offer a picture of religion that is as vague and that spurs us to so little action as the one he does offer.

At the heart of pragmatism, according to Peirce, is the doctrine that our beliefs are always plans for practical actions and not mere idle contemplations. This is seconded by James:

> Mr. Peirce, after pointing out that our beliefs are really rules for action, said that, to develope [sic] a thought's meaning, we need only determine what conduct it is fitted to produce: that conduct is for us its sole significance. And the tangible fact at the  

__28__Two difficult hypotheses of Christianity: It is good not to resist evil, but instead, if someone strikes you on your right cheek, it is good to offer that person your left as well. (Matthew 5:39) Thousands of years ago, a 500 year-old man built a ship out of wood (roughly of the proportions of a modern battleship), gathered 2-7 reproductive animals from every non-aquatic species on Earth into the ship, and sailed safely with all of them for several months while the world’s land surface was completely covered over with water, annihilating all other life. (Genesis 6 & 7)

Two difficult hypotheses of Islam: It is evil to play games of chance. (Sura 2, Ayah 219) At least some ants are capable of speech. (Sura 27, Ayah 18)

__29__“...[belief] involves the establishment in our nature of a rule of action, or, say for short, a habit.” (How to Make Our Ideas Clear, 85) “...belief is a rule for action...” (Ibid.) “...the whole function of thought is to produce habits of action...” (Ibid, 87) “The Protestant churches generally hold that the elements of the sacrament are flesh and blood only in a tropical sense; they nourish our souls as meat and the juice of it would our bodies. But the Catholics maintain that they are literally just that; although they possess all the sensible qualities of wafer-cakes and diluted wine. But [...] ...beliefs about wine are nothing but self-notifications that we should, upon occasion, act in regard to such things as we believe to be wine according to the qualities which we believe wine to possess. The occasion of such action would be some sensible perception, the motive of it to produce some result. Thus our action has exclusive reference to what affects the senses, our habit has the same bearing as our action, our belief the same as our habit, our conception the same as our belief; and we can consequently mean nothing by wine but what has certain effects, direct or indirect, upon our senses; and to talk of something as having all the sensible characters of wine, yet being in reality blood, is senseless jargon.” (How to Make Our Ideas Clear, 87 & 88)

__30__If we consider that we neither get to choose our beliefs nor have beliefs without being disposed to act on them, it is easy to see how wrongheaded the very idea of willing to believe is. Quite to the contrary, the truth is that we do not will to believe at all, but rather that we do the will of beliefs!
root of all our thought-distinctions, however subtle, is that there is no one of them so fine as to consist in anything but a possible difference of practice. [...] This is the principle of Peirce, the principle of pragmatism.31

So, to understand the meaning of an affirmation like that of Matthew 7:12 (the “golden rule”): “[It is good that] all things whatsoever ye would that men should do to you, [ye do] even so to them...” we need to know what sort of behavior will be elicited in one who believes it. And, in this instance, it is relatively clear what kind of behavior will be elicited. One who believes Matthew 7:12 will strive to treat others with kindness and consideration (since this is how the person himself would wish to be treated). One who consistently does not strive to do so, but claims nonetheless to believe in the golden rule can safely be said to be feigning belief rather than truly believing.32

What, then, is the behavior elicited by (which is the meaning of) James’ religious hypothesis – “the best things are the more eternal things”? It is difficult to think of any behaviors at all that belief in this hypothesis brings about. At least, one might say, one behavior it elicits is that a person will answer in the affirmative when asked whether more eternal things are superior to more temporary things. But to be asked such a question would be highly unusual and, whatever one answers, the words of that answer without any further accompanying behavior behind the words seem unlikely to make any difference in the daily waxing and waning of life. If there are no other behaviors elicited by James’ religious hypothesis – if there is nothing we can point to that makes a difference somewhere – there can be no pragmatic difference between believing in his religious hypothesis and not believing in it. After all, as James himself famously wrote in “What Pragmatism Means“:

There can be no difference anywhere that doesn’t make a difference elsewhere – no difference in abstract truth that doesn’t express itself in a difference in concrete fact and conduct consequent upon that fact, imposed on somebody, somehow, somewhere and somewhen.33

31What Pragmatism Means, 18. Clifford would have agreed with both Peirce and James on this point: “If a belief is not realized immediately in open deeds, it is stored up for the guidance of the future. It goes to make a part of that aggregate of beliefs which is the link between sensation and action at every moment of our lives, and which is so organized and compacted together that no part of it can be isolated from the rest, but every new addition modifies the structure of the whole. No real belief, however trifling and fragmentary it may seem, is ever truly insignificant; it prepares us to receive more of its like, confirms those which resembled it before, and weakens others; and so gradually it lays a stealthy train in our inmost thoughts, which may someday explode into overt action and leave its stamp upon our character for ever.” (The Ethics of Belief, 18)

32Again, Clifford agrees: “Nor is it that truly a belief at all which has not some influence upon the actions of him who holds it.” (Ibid.)

33Page 20.
Following James’ own recommendation above, if belief in his religious hypothesis makes no difference at all, mustn’t it be disqualified, in true pragmatic style, from having the status of even a meaningful hypothesis, much less the status of a profound religious truth? It seems clear that, failing further exposition on James’ part (unlikely at this late date), his religious hypothesis falls short of a specifically meaningful proposition. This may help to explain James’ willingness to affirm it without fear of losing anything more than the respect of others.

It should be noted that religious folk (Christian religious folk, most prominently) who follow creeds with much more concrete meanings have heard about James’ justification of faith and have used his words to affirm their right to the religious beliefs of their choice in the face of a dearth of evidence as well. Some of these beliefs are as benign as James’ belief that the eternal things are the best. Some of them are positively benevolent (such as the belief that wealth should be at least partially shared with the poor). But some of them (ranging from the literal demonization of other cultural groups to the justification of violence without limit) are also positively malevolent and this is why we must take seriously the implication of the pragmatic conception of belief for James’ view – that James’ will to believe is never just a will to believe, but that it is always also a will to action. All religions worthy of the name require not just one’s nominal assent to a belief or set of beliefs, but a real change in living behavior of some kind. It cannot be ignored that, in the absence of rational regulation, the actions demanded by faithful belief systems sometimes turn out to be very harmful to those sharing the world with faithful believers.

We must admit that it is at least logically possible that those who subscribe to even the most repulsive faithful belief systems have somehow been fortunate enough to have blindly grabbed hold of religious truth. It may, for example, logically be the case that there exists a supreme god who condemns homosexual acts and blood transfusions as great wrongs. Alternatively, there may also be a supreme god who requires regular, ritualistic human sacrifice. If so, blessed are those lucky enough to have faith in these truths (and cursed are the rest of us). But if they are not correct, then, unlike with James, the consequence of their choice of faithful belief rather than critical skepticism is not merely that they will look foolish, but that those around them will literally suffer the tortures of the damned for no good reason at all.

There is also a less obvious, but more insidious threat that is posed to the truth itself by faithful belief, and this results from the fact that faith is both habit-forming and contagious. It is first contagious in the sense that the habit of faithful belief can be mimetically spread to other people, but more importantly, it is contagious in the sense that the habit of faithful belief with regard to religion tends to be carried over, for the believer, into other areas of interest as well.34 The ongoing battle between evolution and creationism (which is pig-

34 “…if I let myself believe anything on insufficient evidence, there may be no great harm done by the mere belief; it may be true after all, or I may never have occasion to exhibit it in outward acts. But I cannot help doing this great wrong towards Man, that I make myself credulous. The danger to society is not merely that it should believe wrong things, though that is great enough; but that it should become credulous, and lose the habit
lipped under the pseudo-scientific name “intelligent design” these days) is just one manifestation of this crossing over of the "method" of faith from the realm of religion to the realm of science in the minds of faithful believers. The willingness of a people to uncritically trust the judgment and motivations of their political leaders is perhaps another. Furthermore, as Clifford points out,\(^\text{35}\) when we cease to have concern for the truth in our own minds, others will cease to have concern about the truth of the words they speak to us,\(^\text{36}\) thus compounding the problem of finding the truth even for people who have not, as a rule, adopted faithful belief as an epistemological way of life.

Given all this, we must judge that, far from being harmless, the \textit{real} potential cost of faith is measured in terms of tangible human inconveniences and/or atrocities...not to mention the cost of fostering a culture of Bullshit\(^\text{37}\) and half-truths. Could such risks \textit{possibly} be worth the prize of guessing the religious truth from an infinity of equally unlikely religious alternatives? Clifford posed the question this way in 1877: “What would be thought of one who, for the sake of a sweet fruit, should deliberately run the risk of delivering a plague upon his family and his neighbours?”

\[^\text{35}\] ...and as Frankfurt echoes over a century later in his treatise “On Bullshit" when the problem foretold of by Clifford is in full bloom.

\[^\text{36}\] “The harm which is done by credulity in a man is not confined to the fostering of a credulous character in others, and consequent support of false beliefs. Habitual want of care about what I believe leads to habitual want of care in others about the truth of what is told to me. Men speak the truth of one another when each reveres the truth in his own mind and in the other’s mind; but how shall my friend revere the truth in my mind when I myself am careless about it, when I believe thing[s] because I want to believe them, and because they are comforting and pleasant? Will he not learn to cry, “Peace,” to me when there is no peace? By such a course I shall surround myself with a thick atmosphere of falsehood and fraud, and in that I must live. It may matter little to me, in my cloud-castle of sweet illusions and darling lies; but it matters much to Man that I have made my neighbours ready to deceive.” (Ibid.)

\[^\text{37}\] This is not meant as a snarky remark, but a technical one. See Harry Frankfurt’s “On Bullshit” about the rise of the modern problem of meaningless communication.
The issue of free will has been discussed among philosophers for centuries. Many ask what free will is, provide their own definition, and decide whether we have it or not based on that definition. Some philosophers define free will as our freedom to choose our courses of actions, to be in control of our actions, our power to act freely. Where does free will tie into religion, and why is it important whether we have it or not? We think we have it when it comes to choosing one path instead of another, but as will be seen in this paper, we are merely fooling ourselves by believing that. The purpose of this paper is to answer these questions, among many others, in order to conclude that if God is omniscient, if God is in control of everything that goes on in the world, and if He has foreknowledge of everything that we will do in the future, then humans lack free will but still fool themselves into believing that they have it.

Why Free Will Should Matter

The argument for the apparent impossibility of both God’s having foreknowledge and humans’ having free will has troubled religious thinkers and philosophers for centuries. But why does it matter so much whether we have free will or not, one may wonder? If we have free will, if we have the power to choose how we act, then it makes sense, legally and morally, for us to be punished and rewarded for our actions. However, if we lack free will, if we lack the power to choose how we act, then we should not be punished or rewarded for what we do. After all, we are morally responsible for only those of our actions that are done freely, or “if we are not responsible for actions we cannot help doing, then it follows that nobody is morally responsible for anything” (Mates). This is critical when religious beliefs and the Judgment Day are involved; for example, if a person commits a sin, he does not do it so because he chose to but because it had been already been planned that he would do so; it is therefore morally as well as legally unjust for this person to be punished for the committed crime since God was using him to carry out His divine plan. As Davis, in Theory of Action, notes,

Many people think that an agent is responsible for his action only if ... his action, its antecedents, or he himself is ‘free.’ If he is not responsible, then neither praise nor blame, reward nor punishment can be fully appropriate from a moral point of view or (many would add) from a proper legal point of view (109).
If we have free will, it would mean that we are the masters of our own fates, "captains of [our] souls" (Davis 109). Having free will, thus, would confer upon us certain dignity, certain power that may make us feel good about what we choose to do or choose not to do.

Another reason why free will matters, or should matter, is that we should ask ourselves how much and what sorts – if any at all – of power we possess as humans and whether our own conceptions about our own selves are correct. For instance, "Am I, and are we all, 'cogs in a machine,' or are we the originators of courses of action?" Ekstrom asks (14). If we have free will, then that indicates that we are the originators of courses of action; if we lack it, then we are simply "cogs in a machine." If free will indeed exists, then we are free beings, but if it does not, then those of us who believe we have it should stop living a lie and critically evaluate what it means not to have free will, for knowing the truth about ourselves will indeed be valuable and gripping.

Theologians, too, are troubled by the idea of whether or not humans possess free will if God has foreknowledge. As Swartz points out, many religions insist that God has given us free will and we can thus choose right from wrong and that "wrongful acts are sinful and worthy of divine punishment, while good acts are righteous and worthy of divine reward."

Determinism points towards the nature of the physical world. It implies that all actions and their antecedents are explainable due to events and actions occurring before them, suggesting that we are neither autonomous in what we choose to do nor responsible for any of it. Every event is thus the effect of previous ones, which in turn were caused by event antecedent to them, and so on. Since all human actions, including those that are generally referred to as free, are the inevitable effect of events occurring long before the agent was born and over which he had no control, does this agent really possess any free will? This agent could neither prevent the existence of the causes nor avoid the occurrence of the effects. Consequently, "despite appearances to the contrary, human actions are no more free than the motion of the tides or the rusting of a piece of iron that is exposed to water and air" (Mates). In Theory of Action, Davis cites two arguments dealing with determinism:

According to one, determinism implies we are compelled to act (and to think and to feel) as we do. According to the other, determinism implies we are not in control of what we do (think, feel). Either way, we can hardly be considered autonomous or judged responsible for anything we do (110).

Determinism states that every event "stands in a chain of events stretching backward into history, the links of which are deterministic causal connections" (Ekstrom 16). This means that there is exactly one way the world can go, that there is exactly one physical possible future.\(^1\) If this is the case indeed, then is there any room for anyone to break this chain such that he may choose a course of action? Most likely not since he would be going against the laws of nature, as every event that happens, according to determinism, is the result of

\(^1\)This view is incompatible with the Quantum Mechanics theory, however, which argues that there are more than one possible ways an event can occur – hence negating the conception that there is only one way the world can go and only one physical possible future.
previous events that took place in the past. If whatever we do is the result of Laws of Nature and of our physical and genetic makeup and our personal history, “then – since all these 'factors' are 'set' (or 'in place') at the moment of one’s acting – you must undertake the action you perform, that action must occur, you are powerless to prevent yourself from undertaking that action” (Swartz). If a proposition about some future action you undertake (let's say tomorrow) is true, then it is true now. But if it is true now, then tomorrow you must undertake that action, that action must occur, and you are powerless to prevent yourself from undertaking that action. Consequently, we do not have free will since our choices and actions are caused.

It is worthwhile to note that there are two types of determinists: hard and soft. The former hold that responsibility for our actions is an illusion; the latter determinists, also referred to as compatibilists, hold that causation is neither a constraint nor a compulsion (in matters of free will) and to act freely is not to act unpredictably.

*Divine Foreknowledge*

Let us first consider the omniscience of God. God, in many religious views, is considered to be omniscient, and His being omniscient means that He knows everything that has happened in the past, everything that is happening right now, and everything that will take place in the future. Hence, God knows everything about us, our thoughts, our actions at every moment of our lives; and what we will do in the future, what we have done in the past, and what we are doing in the present. Because God can never be wrong or hold any false beliefs, whatever He believes will happen must happen, or He is not infallible – a quality that comes with the definition of God. The same religions that believe God to be omniscient – these religions include Islam, Christianity, and Judaism – also believe in divine foreknowledge and-or predestination, stating that God created a plan for the world and all of its inhabitants even before He created the world; this conception also comes with the definitions of foreknowledge and omnipotence, expressing that God knew from eternity what will happen at such and such point in time, that God knows what we will do before we do it. Hence, that God has foreknowledge of everything means that this plan He has in mind for His Creations is put right in front of him in a timeless zone; thus, all things are happening in front of Him in the present tense. As Boethius asks, “If God knows in advance what one is going to do, then how can one have any free choice in the matter?” (The Consolation of Philosophy 105). He further adds:

If God sees everything in advance and cannot be deceived in any way, whatever his Providence foresees will happen, must happen. Therefore, if God foreknows eternally not only all the acts of men, but also their plans and wishes, there cannot be freedom of will . . . If things could somehow be accomplished in some way other than that which God foresaw, His foreknowledge of the future would no longer be certain . . . And it would be wrong to think that of God (The Consolation of Philosophy 105).
Keeping this in mind, let us suppose, then, that a person has the option of going to either Store A or Store B. God knows in advance that this person will choose Store A, but the person does not know which one he will choose and (falsely) believes he can choose either one. However, because God has already planned for him to choose A, can one really say that this person has the free will to choose Store B if he wanted to? If he chooses Store B, he will be going against God's plan, against what God believes he will do, and since God can never be wrong nor hold any false beliefs, this person must choose Store A.

Let us now consider a real life scenario. Suppose that Lucy took a dancing class two years ago. She took this class believing that she does not have to but instead chooses to do so. However, the fact is that she had no choice but to do it according to God's foreknowledge about her actions. Because God is believed to be omniscient, He knew from eternity that Lucy would take this dancing class, much before Lucy took the action. Lucy's taking the class illustrates that God's foreknowledge was accurate and that she was acting according to God's omniscience; she was thus unable to refrain from not taking the dancing class because had she refrained from doing so, she would have proven God to holding a false belief about what she will do what she did two years ago.

Such a view creates a problem regarding Hell and Heaven. Do we go to hell because we are meant to go there, since we lack free will, or do we go there because of the actions we choose to do? But, then again, if we have no choice in what we do, is it fair that we will go to hell for our sins? Further, some religious followers claim that humans do have free will, yet they warn us that if we do not follow the religion they follow, God will punish us. In this sense, at least, do we really have free will? How are we justly punished for the paths we reject?

If, however, we are not to believe that God is not omniscient, then humans must have free will. If God did not know what will happen tomorrow, nor what happened yesterday, nor what is happening now, then it would eliminate the idea of God's foreknowledge, which will have the advantage of allowing us the belief that that in some, sense at least the future actions of human beings are free. A non-omniscient God would not know everything nor would He have any plan for the world with which it could act accordingly.

Divine Providence

Another challenge to the doctrine of free will lies under divine providence. Who controls the universe and everything inside and outside of it? Many of us will answer that God, being omnipotent, controls it all, including us humans. This means that God can do anything He wants, has authority over all of His creations. If this must be so, then do we as humans control anything at all, especially our actions? Must our actions not be such that they are in accordance with God’s plan? Must God not have it such that our actions do not go against His purpose for the universe? If so, does this leave any room for us at all to act as we wish, to choose freely what to do, including our religious beliefs? Is there any possibility “of our exercising any except the ones we do exercise, at the precise times we do so”? (Davis 110) The answer seems to be an obvious “no.”
Some people might argue that we have free will as long as we voluntarily choose to do whatever is in question, that if one can do something other than what one does, then one has free will. What they fail to realize, however, is that we are evaluating whether or not we can be masters of our own fates. Nonetheless, supposing that these people are right in claiming that if we can volunteer to act a certain way, then we have free will; this would then contradict our earlier claim that because God has foreknowledge of everything we do and that because we are predestined to act, live, and think a certain way, we essentially do not have the power to do contrary to what it is that we do, think, or even say. Indeed, we have no genuine freedom or free will since we can never do otherwise than what we in fact do. Divine providence, hence, should make us suspicious about whether or not we have freedom of the will.

Possible Solutions to the Problem of Free Will

Alvin Plantinga makes a plausible objection to the argument that if God is omniscient, then humans do not have free will. In his article God’s Foreknowledge and Human Free Will Are Compatible, he states that “the claim that divine omniscience is incompatible with human freedom seems to be based upon confusion.” His objection is specifically directed towards Nelson Pike, who believes that free will and God’s omniscience are not compatible, that if one had the power to do something such that if he does it, then a belief that God held at $T_1$ would be false. Plantinga responds that “if God is essentially omniscient (omniscient in every world in which He exists) what follows is that in [a world] $W$ [in which Jones refrains from doing $X$ at $T_2$], God did not believe at $T_1$ that Jones will do $X$ at $T_2$; He believed instead that Jones will refrain from $X$.” He also points out that if God is essentially omniscient, then He is omniscient in every possible world in which He exists, and the same notion that God never holds any false beliefs apply in every world in which God is question.

Another solution may be the idea of attributing “to God special properties, for example, being ‘outside’ of time” (Swartz). If God is solely watching every event that takes place in our world at our time, and He is not in time with us, then we can argue on behalf of free will while still believing that God is omniscient. Some other theologians have argued that God has a “special way” of knowing: unlike human beings, who must “interact with the world (e.g., read a report, see an event, examine evidence [such as ashes, skid marks, etc.]), God is said to ‘know directly’ – i.e., without the need of sensory data or of physical interaction with the world” (Swartz). Unfortunately, this notion is problematic as it raises too many questions and “it is hard to see how it solves the problem at hand” (Swartz).

We can also solve the problem of free will and God’s omniscience by arguing that just because God has foreknowledge about every event that takes place and every action humans make, it does not necessarily have to mean that He has already planned for those events and actions to occur; perhaps He is sitting outside of time, watching what happens, knowing that it will happen.
Conclusion

While we can always argue that we choose whom to befriend, which school to go to, whom to marry, etc., etc. the question of free will is not a question of what individuals are religiously, socially, politically, or legally allowed to do. We are more concerned with a deeper question, which is whether or not we can be masters of our own fates in such a way that we direct our own lives. Evaluating three main challenges with regards to free will, it can be safely concluded that humans lack free will, no matter how desperately we hope, think, or believe we have it, or wish we had it. First, scientific determinism, stating that every event that takes place does so due to events that have already taken place in the past – thus disallowing us the freedom to choose according to our own will; from this, we can make the conclusion that if determinism is true, then it would counteract human morals and ethics.

Second, God’s being omniscience means that He has foreknowledge of everything that happens. God is all-knowing and all-powerful. He would know the position of every particle in the universe at every time. He must surely, therefore, know everything that would happen in the universe, infinitely into the future. Further, He made us, so He knows our thoughts and knows what choices we make in the present, have made in the past, and will make in the future – just as He knows about every event that takes place. Even if God does not control our will, an omnipotent and omniscient God would know what it would be. As a result, this contradicts the idea of our freely choosing as we wish since He already has planned for us to make certain actions that prevent us from doing what He has not planned for us. And if we were to choose something other than what we choose to do, we would be exemplifying that God held a false belief before we encountered that action.

Third, if God is the King of the Worlds and all that is inside and outside of it, where does our personal freedom come into place? He controls everything that happens, including our actions, and that leaves no freedom for us to control anything, even our actions. Accepting this, how can we have free will? It must just be an illusion of choice. We are destined to live and die in a certain way.

Still, some philosophers do propose ideas in which God’s being omniscient and humans’ having free will do not clash. For instance, if we consider Alvin Plantinga’s perception, we see that an omniscient God would hold no false beliefs in any world, which eliminates the idea that if a person chooses Path A as opposed to Path B, then he could not have chosen Path B; this would be because if the person had chosen Path B, the omniscient God would have known that information instead. We can also consider the suggestion that perhaps God is not with us in time but sees every choice that humans freely make.

However, because the argument against free will is stronger than that of for it, we can conclude that if God exists and is omniscient, then humans do not have free will.
Joshua Horn – Kant’s Political Philosophy: An Evolution of Social Contract Theory

Few scholars would disagree with the claim that Kant is one of the, if not the most, important and influential philosophers of all time. There are several reasons for this, but what I see as the best reason is the fact that he wrote on almost every conceivable philosophical topic—metaphysics, epistemology, ethics, political philosophy, etc. While other prominent philosophers can also lay claim to this broad scholarship, what separates Kant is that he is perceived as the pinnacle of modern philosophical thought. Most history of philosophy texts that one will see, explain the fact that he combined the best parts of rationalism and empiricism to pave the way for faith and reason together.

In much the same way, his political philosophy is often understood as an amalgamation of earlier thought, principally Social Contract Theory. However, what makes this endeavor so important is the fact that there are such crucial differences between the political philosophy of Hobbes, Locke, and Rousseau, that it could be very dangerous to pull possibly incompatible ideas together into one new narrative. Most philosophers would say that Kant succeeded in his reconciliation of continental rationalism and empiricism, my goal here is to see if he does the same with his political philosophy. While it would take a full dissertation to make every comparison between only one of these authors and Kant, I will restrict my essay to what I see as the most contentious claims that Kant may have a problem with, namely, Rousseau’s conception of the “general will”, Locke’s conception of private property, and Hobbes’ conception of revolution.

Before we delve into the subtle underpinnings of Kantian deontology and how this carries over into his political works, it is essential to lay the groundwork of my argument, by giving some sort of exegetical discussion first. After all, the aim of this essay is not merely to say what Kant is pulling out of each Social Contractarian, but instead, whether the combination of each is consistent. Instead of moving historically through each philosopher, I want to begin with Rousseau and his conception of the “general will” because it is the foundation for understanding both his political philosophy, as well as Kant’s. I will then move into Locke’s notion of private property, before finally ending with the fitting conception of revolution, the end of the state, as seen in Hobbes.

Rousseau’s depiction of the Social Contract varies over the course of his career. What he illustrates in his earlier work is a trick by the rich against the poor. This is apparent in his well known sentiment that, “Such was, or should have been, the origin of society and laws, which gave new fetters to the weak and new forces to the rich, irrevocably destroyed

---

1While I understand that Social Contract Theory encompasses far more philosophers than just Hobbes, Locke, and Rousseau, in the interest of brevity, I will limit this discussion to what I, along with most philosophers see as the most important figures in the tradition.
natural liberty, changed adroit usurpation into irrevocable right, and for the profit of a few ambitious men henceforth subjected the entire human race to labor, servitude and misery.”\(^2\) He then proceeds to tell us that somebody should have said, “Do not listen to this impostor. You are lost if you forget that the fruits of the earth belong to all and the earth to no one!”\(^3\) However, in his later work, he acknowledges that there is a better way for the contract to come about, and it results from coming together for means of preservation. “Since men cannot create new forces, but merely combine and control those which already exist, the only way in which they can preserve themselves is by uniting their separate powers in a combination strong enough to overcome resistance, uniting them so that their powers are directed by a single motive and act in concert.”\(^4\) The general will is infallible, and it is composed of every member of the state. The general will is composed of every member of the state because, either people would consent based on reason, or the minority should be forced to be free. Rousseau writes, “it is tacitly implied in that commitment—which alone can give force to all others—that whoever refuses to obey the general will shall be constrained to do so by the whole body, which means nothing other than that he shall be forced to be free...”\(^5\)

Kant, like Rousseau, believes the original state of nature and the rise of the Social Contract to be purely hypothetical exercises. For Kant, this is essential because “right” must be grounded in *a priori* reasoning—not in any historical context. Kant also follows Rousseau regarding the general will must involve all members of the society. While this may be a *prima facie* restraint on individual freedom, Kant argues that it is necessary to force everyone to submit to the sovereign; else it cannot protect everyone universally. He writes, “Now a unilateral will cannot serve as a coercive law for everyone with regard to possession that is external and therefore contingent, since that would infringe upon freedom in accordance with universal laws. So it is only a will putting everyone under obligation, hence only a collective general (common) and powerful will, that can provide everyone this assurance.”\(^6\)

While this may sound eerily similar to Hobbes’ justification for consenting to the sovereign in his second law of nature, namely, “that a man be willing, when others are so too, as far forth, as for peace, and defense of himself he shall think it necessary, to lay down this right to all things; and be contented with so much liberty against other men, as he would allow other men against himself”\(^7\), Kant’s justification is more emphatic on preserving the moral


\(^3\) Ibid.


\(^5\) Ibid., 64.

\(^6\) Kant, Immanuel, *The Metaphysics of Morals* [6:256]

\(^7\) Hobbes, Thomas, *Leviathan*, 190.
law as such, than the egotistical justification that Hobbes provides. Since this is only possible in a civil society, it is necessary for Kant, as well as Rousseau, to have the general will supporting the public. The difference is that for Hobbes, the motivation is based more upon self-interest, than anything else.

One of the obvious objections to Rousseau’s concept of the “general will” is that he is terribly inconsistent. One side to the coin is that it is only through the general will that we can have freedom (I will return to this point soon), whereas on the other side of the coin, the emphasis of the general will is on the collective body of society. One of the other objections that should be noted is that Rousseau is ambiguous with how the general will is identified and recognized. The fear then is that if he is inconsistent, and Kant is using this notion from Rousseau, then, Kant is inconsistent too by association. However, since these are critiques of Rousseau’s philosophy, and my aim here is to address Kant’s consistency, I will table this discussion as a red herring away from my argument.

Another close analogy that Kant shares with Rousseau is that the Social Contract is necessary to preserve freedom. Freedom for Rousseau is only following the laws which we legislate through the general will. Kant defines freedom as a universally administered right according to law. Since this is only possible in civil society, then a sovereign is necessary and legitimate.

These two primary examples are sufficient to show that Kant is very sympathetic to Rousseau’s political philosophy. Most notably, he is partial to the ideas of the general will and freedom. With the foundations of society in place, let us move into one of the most important topics in political philosophy, namely, the right to property, in order to see where Kant stands in relation to a key proponent of property rights, John Locke.

Property is immensely important for the political philosophy of Locke. In fact, it would not be an understatement to say that his entire justification of the state up through his validations for revolution concern property. Locke’s state of nature is drastically different from the Hobbesian war of all against all. Instead, Locke envisages a scenario where everyone lives according to reason in a state of perfect freedom. This reason “teaches all mankind, who will but consult it, that being all equal and independent, no one ought to harm another in his life, health, liberty, or possessions.” He groups life, health, and liberty all under the term “property”, so in effect, the justification of the social contract is to protect our property. “This makes him willing to quit the Condition, which however free, is full of fears and continual dangers: and ‘tis not without reason, that he seeks out, and is willing to joyn in society with others who are already united, or have a mind to unite for the mutual preservation of their lives, liberties and estates, which I call by the general name, property. The great and chief end therefore, of men’s uniting into commonwealths, and

---

8 I by no means intend this to be an exhaustive list of the comparisons between Rousseau and Kant, these are what I see as the most important in arguing for my ultimate position.

putting themselves under government, is the preservation of their property.”

Similarly, we may overthrow the sovereign when he fails to fulfill his contractual obligations.

Since property plays such a vital role in Locke’s overarching arguments, it would be pertinent to proceed further in asking how one gains property. He argues that when we mix something that is intrinsically our own, namely, our labor, with an external possession, we will then have a rightful claim to it.

How does Kant address the role of property in a political system? Does it play such a vital role as it does in Locke, or is it the ultimate reason that we leave our fundamentally good state of nature, as Rousseau will hold? The answer is somewhere in the middle. One of Kant’s prominent political works, *The Metaphysics of Morals*, begins with a discussion of property. So, clearly it has some higher importance than Rousseau, and yet his justification for the state is not grounded on property, so it plays a less important role than with Locke.

Kant divides property into two categories, the “empirical” and the “intelligible”. Empirical possessions are merely what we observe *a posteriori*. We have empirical possessions in close physical proximity to them. For an intelligible possession, we need not have the close physical proximity to it. So, if I am holding an apple, I have an empirical right to it, only insofar as I am holding it; whereas, if I claim a right to an apple tree a mile away, then it is a claim based upon reason. The intelligible possession then, is the higher justification for “right” here as Kant conceives it. Due to its higher status, the only way in which we can have an intelligible possession is in a civil society with a sovereign to protect these rights for us. It is interesting to note here that the sovereign for Kant is not a Hobbesian authoritarian state, but instead a general will by everyone in the society, rationally agreeing to it. So, “John Locke had famously avoided this problem in his theory of property by making property a product of a single individual’s activity. By “mixing” one’s labor with an object in the commons, one comes to have property in the object. Kant objects to Locke’s theory of property on the grounds that it makes property a relation between a person and a thing rather than between the wills of several persons”

At this point, we may ask how Kant is actually pulling the thread of property through Locke. After all, Kant is critiquing to some degree Locke’s notion of the relationship between labor and the individual as justification for property right. And if Rousseau is critiquing Locke’s theory of property too, then should not Rousseau and Kant at least be moving in the same direction? No. We should remember that Rousseau’s state of nature is actually man’s best

10 Ibid., 350.

11 Locke also gives other qualifications for claiming private property. For instance, he argues that we only have a right to property, even if we mix our labor with it, if the property is not wasted. So, we could not pick twenty apples, and let eighteen rot. We would only have a right to the two that we had actually used.

state to live in, and the only way that we form a civil society is based on property. Rousseau believes that the rise of civil society results in the inevitable decline of the human existence as it was meant to be. Although the state exists to protect property, property is viewed as a bad thing. For Locke, the state exists to protect property, and this is the only way in which we can be free and live our lives to the maximum capacity. This is a more Kantian line of thought than Rousseau’s.

It should be apparent now how Kant is pulling parts of political thought from different philosophers, and how they are beginning to come together. There is one more variable in the equation that I wish to distinguish though, namely, revolution. Prima facie, it would seem that if we follow Kant’s moral philosophy and his doctrine of the categorical imperative, then the state, which exists from the general will of the citizens to protect intelligible property rights, must not treat its citizens merely as means to an end. Surprisingly, Kant actually holds, similar to Hobbes, that any sense of revolution would be completely unjust and contradictory to reason.

Recall that Hobbes’ state of nature is radically different than Rousseau’s state, where the “noble savage” can flourish, then we can come close to understanding one of his several arguments for why a subject can never revolt against the sovereign. Hobbes’ original state is the war of all against all, where every individual is purely self-interested, living a life that is “solitary, poor, nasty, brutish, and short.”13 Hobbes argues that we would do anything to remove ourselves from this state of constant fear, even at the expense of implementing a sovereign, whose sole obligation is to protect our life—not our liberty or property as Locke will hold. If we would do anything to get out of this state of nature, then we would certainly not risk revolution where we have the chance of returning to it.

At least one other argument that Hobbes gives for denying any right to revolt against the sovereign involves harming ourselves. In a sense, we would be harming ourselves by returning to the state of nature, however the harm involved here is a contradictory type of harm. More specifically, he argues that the sovereign has unlimited, justified power based upon our explicit consent to enter into the contract. Therefore, any grievance that we have against the sovereign would, in principal, be a grievance against ourselves, since he exists only through our will. Hobbes writes, “nothing the sovereign representative can do to a subject, on what pretense whatsoever, can properly be called injustice, or injury; because every subject is author of every act the sovereign doth...”14

Kant follows suit and argues that we never have a right to revolt against the sovereign. The critical difference though is that Hobbes’ prohibition is based upon self-interested claims, whereas Kant’s prohibition is based on maintaining the rule of law as such. Revolution is also a contradiction in terms. A sovereign, by its very definition, is the supreme law of the land. If there were a right to revolution maintained in the contract, then there would have to be a third party to adjudicate the dispute, which would be de facto higher in power than the sovereign. Since this is absurd, revolution cannot be a just action. Kant writes, “For a

---

13Hobbes, Thomas, Leviathan 186.

14Ibid., 265.
people to be authorized to resist, there would have to be a public law permitting it to resist, that is, the highest legislation would have to contain a provision that it is not the highest and that makes the people, as subject, by one and the same judgment sovereign over him to whom it is subject.”

Another explanation that Kant gives to attempt to argue that revolution is an unjust action will bring us full circle, and take us back to Rousseau and the general will. If we remember that a “right” for Kant is only possible with the consent of the general will, and that the general will is sovereign, then it would be another contradiction to hold that one could legitimately revolt against the very system which gives it legitimacy.

While there are obvious similarities between Hobbes and Kant that should be apparent now, the question now becomes how Kant’s account of revolution is not as similar to Rousseau. This is a tricky question to answer, because in a sense, both Hobbes and Rousseau argue that the sovereign is absolute and infallible. There is an important distinction that Rousseau makes though, that Hobbes does not account for, and it is this distinction that makes all the difference. For Hobbes, the sovereign is the state. They are one and the same. Rousseau diverges here and argues that the government is separate from the sovereign, and while the sovereign (the general will) is infallible, the government, which acts as a medium between the sovereign and the subjects, can be usurped and overthrown. Rousseau writes in Book III of The Social Contract, “What then is the government? An intermediary body established between the subjects and the sovereign for their mutual communication, a body charged with the execution of the laws and maintenance of freedom, both civil and political.” Then in Book IV, he tells us that the general will is indestructible, but not before he ends Book III with the dissolution of the state in two ways. “First it takes place when the prince ceases to administer the state according to the law and usurps the sovereign power...The same situation occurs when the members of the government separately usurp the power which they ought only to exercise as a body; for this is no less an infraction of the law, and it produces an even greater disorder.” Rousseau’s distinction between the sovereign and the state is exactly what makes Kant’s prohibition of government closer to Hobbes.

---


17Ibid., 133.
In essence, here is my overarching argument thus far:

P1. Kant uses Rousseau’s conception of the “general will” and freedom to justify a state.

P2. Kant adapts Locke’s conception of property to make it more specific to a justification for the state by distinguishing between the “empirical” and “intelligible” possessions.

P3. Kant agrees with Hobbes’ assertion that we never have a right to revolt against the sovereign.

It would take a great deal more time to critique the entirety of Kant’s political philosophy. So, I will limit the end of my enquiry in answering one question: Does Kant adequately succeed in forming a cogent political philosophy while drawing on so many seemingly incompatible thoughts? So, let us begin by simply analyzing the individual strands of thought, and suppose that Kant has not combined them in such a way. If we were to just evaluate the philosophy of Hobbes, Locke, and Rousseau, would their thought be compatible? The short answer here is, absolutely not. Locke is writing his Second Treatise as a critique of the absurdity of Hobbes’ philosophy. In the same way, Rousseau wrote his Social Contract in response to the tradition beginning with Hobbes, and continuing into Locke.

Locke sharply disagrees with Hobbes that we would remove ourselves from the state of nature merely for security, with no other assurances by the sovereign. In a sense, their very beliefs about human nature are different. Hobbes’ fundamental man is purely self-interested, whereas Locke’s fundamental man can still live in a peaceful state without a sovereign. If we were to throw Rousseau into the mix, the contrasts become even more telling. In contrast to Hobbes’ fearful man in the state of nature, Rousseau believes that the fundamental state is actually when man was at the height of his very being. Man in this state is stronger, faster, more agile, etc. So it seems as if Kant is in a bad position to attempt at a reconciliation of Social Contract Theory. However, Kant succeeds not by focusing on the differences in Hobbes, Locke, and Rousseau, but instead on the subtle similarities that are often overlooked. For example, it may seem as if Hobbes must maintain that the basic state of man is without reason. Otherwise, why should he be so fearful and self-interested? Would not reason dictate that we work together in a primitive state to survive? This is exactly where Locke gets the idea that the state of nature is a state with reason. However, Hobbes and Locke both maintain that the predecessors to humanity in the state of nature had the capacity to reason. Indeed, how could any man without reason, come to understand their basic natural rights, enough to surrender them to a sovereign for protection?

This may seem like a minor point, namely that Hobbes and Locke agree that man in the state of nature can reason, but it is exactly this point which is crucial for Kant’s possible compromise in their philosophy. Instead of focusing on human nature and what the classical social contractarian’s believe results from this state, Kant follows suit with the rest of his enlightenment philosophy with an emphasis on reason. In other words, it is
irrelevant whether man is fundamentally egotistical as Hobbes holds, or whether the state rises merely for convenience sake as Locke will hold.\(^{18}\)

Now that it has been shown that Kant’s emphasis on pulling reason as the primary strand through the social contract tradition, and the general will, property, and prohibition of revolution are merely dependent on that, the last critical question is whether there is an inherent, necessary incompatibility in these terms which would make Kant’s political philosophy fall short. Consider a triangle with these three terms on each point. What I am essentially asking here is whether there is a necessary conflict between the connections of the points of the triangle.

First, there is no necessary conflict between the general will and property. After all, if the general will as Rousseau envisages it is infallible, and the general will consents to have property, then there is no problem. The fact that this may ultimately be a bad thing is irrelevant to whether it is possible, and whether that possibility produces problems.

Second, as we mentioned previously, there is no necessary conflict between the general will and revolution either. In fact, Kant holds that this would be a contradiction to maintain that the general will is infallible, and also that it can be justly revolted against. An important thing to note here that could be problematic for Kant is how he can argue that subjects can passively resist the sovereign from within the system, but have no right to openly remove the sovereign from power. In fact, this is one of the major points of criticism of Kant’s political philosophy. This is one of the very criticisms that Kant addresses in the appendix to the *Doctrine of Right*, where he elucidates the fact that subjects must obey the sovereign in every circumstance except those which conflict with the inner moral law. This sounds very close to a justification for civil disobedience. Kant writes, “Obey the authority who has power over you (in whatever does not conflict with inner morality.)”\(^{19}\) So, it is not clear then how we can be justified in rebellion, but not revolution, if the general will is infallible and “right” is only intelligible in society. There seems to be an apparent

\(^{18}\)Rousseau also holds that man in the state of nature has reason. Although Kant may very well pick up on his idea here too, the state does not result from reason for Rousseau, as it does for Hobbes and Locke. The state rises instead from pride due to people living together, and what Rousseau believes is a necessary result from property.

\(^{19}\)Kant, Immanuel, *The Metaphysics of Morals* [6:371].
distinction between whether the categorical imperative may exist prior to society. In any event, revolution in the way that the comparison was made earlier with Hobbes, has no necessary conflict with the general will.

Lastly, is there a necessary incompatibility between property and revolution? Once again, Kant’s attention to detail in his terminology saves his overarching argument. If he had not distinguished between the sensible and intelligible possessions, a conflict may have resulted. However, property is not the justification for the state for Kant, in the way that it is for Locke. So, not only would Kant have had to conflate property rights, he would have also needed to base the sovereign on protecting our property, for a conflict to arise. Instead, the state exists to ensure freedom in an attempt to reach a state of perpetual peace, where the categorical imperative can flourish in a kingdom of ends. Since intelligible property is only possible in society, it would be absurd to justify revolution to preserve these rights. Therefore, there is no necessary conflict that Kant must address here either.

While it may be the case that Kant’s political philosophy is nothing more than slight variation on Rousseau’s thought, this claim is not necessarily correct. It may be the case that Kant derives even the majority of his political thought from Rousseau, but it is certainly too much to say that Kant had no other influences. Obviously who I have in mind here is Hobbes and Locke. So, while Kant may have meritoriously declared Rousseau the “Newton of the moral world”, this should not be taken as a universal acceptance of Rousseau’s social and political philosophy. My claims here were more modest. I did not argue that Kant manipulated several different aspects of the social contract tradition, but instead, that he was influenced enough to use some small threads from each throughout his political philosophy. The hat trick then for Kant is not combining political philosophy from his predecessors to make something of his own, but instead, combining the correct premises from each philosopher in the correct way to fend off incompatibility. This was a task that he succeeded at admirably.
Christa Johnson – A Fresh Take on the Non-Identity Problem

The Problem

Consider two cases about having a child:

(1) A woman, Mary Kate, conceives a child. Her doctor warns her that taking recreational drugs during pregnancy will cause her child to be born with a moderately debilitating heart condition. Despite her doctor’s advice, she knowingly takes the recreational drugs, and her child is born with the heart condition.

(2) Another woman, Ashley, is suffering from German Measles. Her doctor informs her that if she conceives a child now, the child will be born with a moderately debilitating heart condition. However, if she waits two months, her child will not develop the heart condition. Despite her doctor’s advice she conceives a child immediately. Her child, Michelle, is born with the heart condition.

In each of the cases, the heart condition will have certain consequences for the child: fewer opportunities in life, the need for a heart transplant at a young age, and potentially a shortened life. Despite these setbacks, the child’s life is worth living.

Many would argue that Mary Kate and Ashley’s actions are equally wrong. Both women could have avoided having a child with a heart disorder: Mary Kate by not taking recreational drugs during her pregnancy and Ashley by waiting a couple months to conceive. However, there is an important difference between the cases that complicates arguing that the wrongs are equal. The case of Mary Kate is cut and dry. She conceived a child, and then went on to harm that child through her drug use. Ashley, however, conceived a child while recovering from German Measles, which caused her child to have a heart condition. Her alternative to this was to wait two months to conceive, which would have caused a different egg and sperm to join and created a completely different child.1 While that child would have been healthy, he or she would not have been Michelle. Thus, Ashley’s options were to bear Michelle with a heart condition or a different child who would be healthy. She could not have given birth to a healthy Michelle. When one argues that Ashley has harmed Michelle by not waiting, he seems to be saying that it would have been better for Michelle if she had not been born and that another child was born. Can Michelle really say that it would be better for her if Ashley had waited two months to conceive? If Michelle has a life worth living, which we claimed she did, she does not seem to have an argument against her mother for her heart condition. Despite this difference in the case, most people still want to maintain that Ashley did something wrong. Derek Parfit calls

1Some have argued that the same child would exist despite the two month wait. Given the scope of this paper, I will not argue that this is not so.
this the non-identity problem. Many philosophers have attempted to address this problem and its many applications. The goal of this paper is to argue for a fresh theory that will illustrate that wrong and work to avoid it.

**Casper Hare and the De Re / De Dicto Distinction**

Before we turn to a solution to the non-identity problem, I want to discuss the goals of a moral theory. It is clear that every moral theory should address a variety of cases and exception cases, be simple enough to follow, and have elements that are both necessary and sufficient. In addition, a moral theory will ideally work both as a decision making tool and a criterion of rightness and wrongness. Some ethicists argue that it is acceptable for a moral theory to only work as one of these two. However, I will be seeking an ideal moral theory, which does both. As a decision making tool, one should be able to look to a moral theory in the face of a decision and use it to discern the right thing to do. Moreover, as a criterion of rightness, it should show who, if anyone, is harmed or benefitted, and if anyone has a right to retribution for harms done. Finally, a moral theory should encapsulate our intuitions—what ethicists call commonsense morality. A theory should explain the deeper reasons for why our intuitions find something to be wrong or else show why our intuitions happen to fail us in a given case.

Now that we have established the type of moral theory we are seeking, we can look at potential solutions to the non-identity problem. Casper Hare has recently suggested using the *de re / de dicto* distinction as a solution. In order to assess his theory, let’s first get a grasp of the distinction through considering a simple case:

**Quarterback case**: The star quarterback claims “I am always faithful to my girl.” When asked how he stays faithful, he explains “Well, Jan is my girl before the game, while Marsha is my girl during halftime, which makes Cindy my girl at post-game celebration.”

If one is discussing the quarterback’s girl *de re*, there is no way in which he is faithful to his girl, Jan. After all, he is with two other girls throughout the course of the night. However, when *de dicto* factors are considered, he is able to say he is faithful to *his girl* by changing who he is referring to when he uses the phrase “my girl”. To discuss someone *de re* is to discuss a fixed person. For instance, if you are *de re* considering the job of your boyfriend, Greg, you will be worrying about Greg’s job, even if Greg ceases to be your boyfriend at a later date. However, to discuss a boyfriend *de dicto* is to discuss whoever happens to be

---


3See Hare, “Voices from Another World” and further discussion of his position below.

4Thanks to Kelly Sorensen for this example.
your boyfriend at any given time. Thus, in de dicto talk, one may say “my boyfriend has a different job every month” if one month you date Greg the contractor, the next month you date Peter the teacher, and this month you date Bobby the police officer. Essentially, de re talk of your boyfriend refers to the specific person who is your boyfriend at the time, while de dicto talk of your boyfriend refers to all of your boyfriends past and present, whoever they happen to be.

The majority of moral theories conceive of harm in the de re sense. In this sense, the theories argue that when Ashley conceives knowing that Michelle will have a heart condition, she harms Michelle, despite the fact that Michelle would not have existed otherwise. Again, it is suspect how Michelle could argue that Ashley harmed her given that she would not exist if Ashley had not conceived right away. However, Casper Hare has recently introduced the idea that a moral theory could use de dicto considerations to handle the non-identity case. Rather than trying to explain how Michelle is de re harmed in some weird way—given that she would not exist in any other scenario—Hare claims that Ashley commits a de dicto harm. When Ashley is considering whether to conceive now or wait two months, she should be considering the well-being of her future child—whoever that child turns out to be. If she conceives now, her future child will have a heart condition. If she waits two months, her future child will be healthy. By choosing to conceive now, she knowingly inflicts a heart condition on her future child.

Hare establishes two criteria for when de dicto considerations should be used. First, the acting agent should be partial to the affected person. For instance, in the case of Ashley, it can be expected that she is partial towards her future child. Hare’s other consideration notes that the circumstances should have no de re expression. For Ashley, either act she performs, Hare claims, will be de re better for her future child so long as his or her life is unambiguously worth living. That is, if Ashley conceives right away as she did, Michelle is benefitted by her existence since her life is worth living. Further, if Ashley had waited and conceived a different, healthy child, that child would likewise be benefitted through his or her existence. Thus, no matter what Ashley does she performs a de re benefit. To Hare, this shows that there is no de re expression and thus de dicto considerations must be followed instead.

Assessing the Distinction

There are a number of problems with Hare’s use of de dicto considerations. Take his two criteria for when de dicto concerns are morally relevant. Leaving partiality aside, there is a question of whether there is a de re expression in the non-identity problem. Hare claims that all decisions Ashley makes before conception will affect the timing of her future child’s conception and thus the identity of her future child. Further, so long as Ashley’s future

---

5Partiality seems to be an obvious condition for de dicto considerations. If one lacks partiality to the object of de dicto concerns, the situation will be as nonsensical as considering the health of whoever is sitting to your left.
child has a life worth living, the child—whoever he or she is—will be de re benefitted by any and all decisions Ashley makes before conception that affect the timing of the conception.\(^7\) Since Ashley chose to conceive when she did, Michelle exists. Further, since her life is worth living, this existence—albeit with a heart condition—is a de re benefit.\(^8\) If Ashley had made other decisions that affected the timing of conception and bore another child, that child would similarly be de re benefitted by the decisions Ashley made. One must accept a number of premises in order to claim that every decision Ashley makes de re benefits her child. The first is that existence is a benefit. If it turns out this is not the case, then bringing someone into existence with a heart condition would not be a de re benefit. Many argue that since people are not waiting in another dimension to be brought into existence, it cannot be said that existence is a benefit. If one is not conceived, then one is not harmed—one is not anything. However, since the argument is largely unsettled, Hare may simply state that for his purposes, he will assume as many do that existence is a benefit.

Even if we grant that existence is a benefit, Hare still faces arguments against a sliding scale of harms and benefits. Seana Shiffrin argues that harms are often more morally weighty than benefits.\(^9\) Take pleasure and pain. If one can conceive of similar amounts of each, it seems that inflicting the pain is a larger wrong than bestowing the equivalent amount of pleasure is a right. Further, even if existence turns out to be a benefit, the harms attached could potentially outweigh the benefit. Thus, if a heart condition turns out to be an outweighing harm, the decisions Ashley makes could potentially de re harm her future child.

Setting the timing of de dicto concerns aside, the theory faces another issue. In his argument, Hare utilizes a case that trades de re harms for de dicto benefits. However, when the theory is applied to the non-identity case, de re benefits are washed out, and in the case of Michelle, a de dicto harm must be assessed.

---

\(^6\) It is noncontroversial to claim that most if not all decisions one makes in everyday life will affect the exact timing of a conception. However, some claim there would still be no identity change. Hare is of the faction that finds a slight change in timing of conception to affect one’s identity. I will thus address the issues that follow if we grant this claim.

\(^7\) See Hare, “Voices from Another World.”

\(^8\) See further discussion below regarding the possibility that existence is not necessarily a benefit.

As was mentioned above, harms often weigh heavier in morality than benefits. Consider his case:

Hare’s Safety Officer: “Tess is a state safety officer, whose job it is to regulate those features of the automobile that protect its occupants in the event of a collision—airbags, crumple zones, and so forth. Noticing that people in her state are not wearing safety belts, she implements some tough new regulations and, a year later, is pleased to discover evidence that they have been effective, that the severity of injuries sustained in automobile accidents has been reduced as a result of people belting up.”

In this case, Hare argues that many accidents are based on split second timing, and the extra seconds spent buckling a seatbelt could have been the difference for these victims between having and not having an accident. Thus, many of the people who sustained accidents were actually de re harmed by Tess’s actions as the regulations led to their accident. However, Hare notes that some accident victims would be de re harmed no matter who they were. That is, since Tess implemented the law, the accident victims are de re harmed. Likewise, if she had not implemented the law, the accident victims that came to be would also be de re harmed. Thus, de re considerations wash out and de dicto concerns must be taken into account.

There are three main problems causing an asymmetry between Hare’s case and the non-identity problem. First, in the Safety Officer case, the de re concerns wash out because either choice Tess makes, the victims are de re harmed. However, in the non-identity case, the de re concerns wash out because either choice Ashley makes, her child is de re benefitted. Secondly, along the same lines, Tess makes a decision in which de re harms wash out in order for her to bestow a de dicto benefit. This allowed us to show that it is morally permissible, if de re harms wash out, to bestow a de dicto benefit. However, in the non-identity case Ashley chose a de dicto harm once the de re benefits washed out. Thus, we are trying to show that it is wrong that she inflicted a de dicto harm even though it turns out to be a de re benefit. Lastly, the Safety Officer case is not necessarily a same numbers case since there is no way of knowing if there will be more or less accident victims in each scenario.

---

10See Hare, “Voices from Another World.”

11It cannot be said that all of the accidents sustained were due to Tess’s regulations. Rather, a number of accidents would have occurred regardless due to inevitable car failure, drunk driving, disregard for regulations, and previous seatbelt wearing individuals. I only wish to consider those whose lives actually were changed by Tess’s actions.
However, in the Ashley case, the numbers are clear. In order to assess Hare’s de dicto reasoning, we will need a symmetrical case.

The Coach: Coach Baymore is a track coach. The Admissions Office at his college tells him that he can select only one more student to get accepted with a scholarship. Tom has a history of knee problems and will need his knee replaced if he continues running and will be very limited in physical activity the rest of his life. However, Coach knows that Tom would compete anyway and be a large asset to the team. Jerry has no history of injury, and will remain healthy throughout his college career if accepted. However, Jerry’s grades are not up to par and he will have to spend the fall semester in the local county college to get his grades up. Coach Baymore chooses Tom. Tom runs the fall season without realizing his risk of knee injury, and goes on to need his knee replaced which limits him the rest of his life.

In either scenario, the runner Coach Baymore chooses will receive a great education that he would not have otherwise received. However, if not chosen, both runners will still receive a moderate education most likely at a trade school. Furthermore, Tom’s injury does not make him completely immobile. He still benefits from his education and goes on to have a worthwhile life. Coach Baymore should be partial to the health of either athlete he chooses, and it seems that either way, the athlete chosen will be de re benefitted by his choice due to the great education received. Thus, according to Hare, Coach Baymore should base use de dicto concerns for his future athlete. However, by choosing Tom, Coach Baymore de dicto harmed his future athlete.

This case better parallels the non-identity case in a number of ways. First, either athlete will be de re benefitted if Coach Baymore chooses him. Likewise, either child Ashley conceives will be de re benefitted if chosen. Secondly, the goal of the Coach Baymore case is to show that the de dicto harm he inflicted is wrong despite the inevitable de re benefit Tom had bestowed upon him. Likewise, with Ashley’s case we wish to show that the de dicto harm she inflicted is wrong despite the de re benefit Michelle received. Although the Coach Baymore case still considers people already living, it is able to serve as an instructional case to better assess the application of de dicto considerations to the non-identity problem. The Coach Baymore case shows that given the same de re result in either scenario, the de dicto considerations can work as a decision making tool to allow the moral agent to do the right thing. Thus, the result Hare sought remains and the de dicto theory seems to survive.

The final issue Hare’s de dicto theory must address is whether or not it is a sufficient moral theory. We saw above that it seems to succeed as a decision making tool. Once one is able to distinguish when de dicto harms and benefits have moral weight, one can use de dicto considerations to weigh options and choose the best moral decision. However, the theory does not seem to provide a criterion of rightness. Hare claims that the wrong is in the de dicto harm that Ashley inflicts upon Michelle. However, it is hard to assess what this harm really means. Hare claims in a quick concluding paragraph that “the de dicto concern account puts a finger on [Michelle’s] special grievance. [Michelle] alone can say, ‘You failed
to show appropriate de dicto concern for your child, and I am your child.”\textsuperscript{12} It seems to me that Michelle can say this, but the truth is, if Ashley had shown appropriate de dicto concern for her child, Michelle would not exist. Although one may argue that this is a complaint, it seems to be an attenuated complaint at best—that is, a complaint with considerably less moral force than our original intuition about the wrong Ashley has done. In Hare’s story, Michelle is not actually arguing that Ashley harmed her. Rather, Michelle’s complaint is that Ashley harmed her future child who Michelle ended up being. This does not seem to account for the wrong many feel is done in this case. Hare does have room for possible responses to the argument. However, these responses will either hinge on the asymmetrical Safety Officer case or argue that there are no complaints or victims needed to assess the harm. The Safety Officer case has already been rejected. Further, just because one need not find a victim to assess a wrong, does not mean there is not a victim present. Similarly, just because we may or may not need to find a complaint in order to assess responsibility and retribution does not mean there is not one. Thus, let us move on to an alternative approach.

The Union of the Distinction

Given the shortcomings of Hare’s de dicto approach, I suggest an approach in which both de re and de dicto considerations are taken into account. I argue that de dicto concerns should be used as a decision making tool, while de re interests are a criterion of rightness to assess harms and complaints. First, let us take the de dicto approach. We learned above that it functions well as a decision making tool. However, Hare’s criteria for when the use of de dicto concerns should be enacted seem to fail. So I must argue for new considerations. First, we should utilize Hare’s initial observation. That is, the moral agent in question should be partial towards the person whom the decision will affect. Further, due to this partiality, we should expect the moral agent’s concern for the person affected to guide some of the agent’s decisions. For instance, we would expect a mother to be partial to her child and thus for these concerns to affect the decisions she makes. A second observation is that de dicto concerns should only be considered for cases in which the state of affairs of the affected party could be potentially changed. For example, de dicto effects should not be consulted when a man is debating where to sit and the decision technically affects the de dicto health of the person to his right. Whether he sits next to a terminally ill person or someone who is perfectly healthy does not change the state of affairs of either person affected. These observations work to avoid unnecessary and unimportant de dicto consequences.

A final observation relates de dicto concerns to the non-identity problem specifically. That is, de dicto concerns should be taken into account when the object of the decision has yet to be determined. This may seem implicit in de dicto reasoning, but it is important to note this in considering the non-identity problem. Before conception, it is not determined who exactly will be the object of the decision. Thus, a mother must consider all possibilities and decide what will be de dicto best for her future child, whoever that turns out to be. In this

\textsuperscript{12}See Hare, "Voices from Another World."
way, using de dicto concerns as a decision making tool allows one to avoid de re harms once the object of the action is determined. Once a child is conceived, these de dicto concerns drop out, and we must turn to de re consequences.

As was just noted, de re considerations are necessary once the object of the moral agent’s action is determined. Here, de re concerns work as a criterion of rightness. There are many ways in which one may be de re harmed. First, there is the standard form of harm: an existing person’s level of wellbeing is somehow lessened by another. There is also the type of harm in which one’s rights are violated by another. Further, one may have obligations to another person and harm that person by not meeting those obligations. It is important to note that with these last two types of harm, there are cases in which one is said to be harmed despite an overall rise in one’s level of wellbeing.

Given the scope of the paper, I will just focus on the rights violation type of harm. Imagine a Muslim trying to purchase a plane ticket. The airport is highly prejudiced and refuses to sell him a ticket. We would all agree that the Muslim’s rights have been violated, which harms him in a way. However, say that the plane he wished to ride crashes. The Muslim, in terms of pure outcomes, was benefitted by the actions of the airport. If the airport had sold him a ticket he would have died. However, despite the overall result, we still wish to say that his rights were initially violated and that harmed him. In other words, the plane crashing does not justify the prejudicial actions of the airport.

Let us now apply the notion of rights violation harm to the non-identity problem. This seems to be tricky since the child does not exist and therefore does not seem to have rights. However, the authors of *From Chance to Choice: Genetics and Justice*\(^\text{13}\) argue that while a fetus does not have rights that can be violated, after birth one becomes a fully morally weighted person and rights as well as violated rights come into existence. To elaborate, a child—that is, a fully morally weighted person—has the right to have harm prevented when it is possible. Thus, when methods such as prenatal interventions are available to prevent serious disabilities for the future child without compromising the health of the mother, parents should utilize these means. Failure to do so “will violate the moral right that the child now has to have had the harm prevented when it was possible, earlier, to do so.”\(^\text{14}\) In this way, when the child’s rights come into existence, his or her right to harm prevention will come into existence as an already violated right. Furthermore, this reasoning is warranted because there is little doubt that the fetus will become a full fledged moral being and eventually have these rights.\(^\text{15}\)

---

\(^{13}\)See Allen Buchanan, Dan W. Brock et al., *From Chance to Choice: Genetics and Justice* (New York: Cambridge, 2000), pp. 204-257.

\(^{14}\)See Allen Buchanan, Dan W. Brock et al., *From Chance to Choice.*

\(^{15}\)The authors do not use this theory to assess preconception cases, but leave the possibility open. I find that so long as we consider future people that will eventually have a full fledged moral status, the argument can work for the non identity problem.
The rights violation theory does face a major objection. That is, many object to my claim that Michelle is harmed at all. They argue that it can be assumed that Michelle would waive her rights in order to receive the benefit of existence. If this is the case, then it would not be wrong to violate her rights. This objection may not be warranted, however. Shiffrin makes a distinction between benefits and pure benefits. Pure benefits are those that one could do without, such as breast augmentation surgery, while regular benefits are those that are necessary for survival, such as the benefit of being saved from drowning. Shiffrin finds existence to be a pure benefit. She argues that prior to existence there are not beings waiting to be conceived who are then benefitted through existence. Rather, in a sense, one can do without existence. After all, if one is not conceived, there is no one that is harmed in any way. Shiffrin argues further that in the case of pure benefits, one may not assume that a person would agree to some harm in order to receive the benefit. We cannot be sure if, in our case, Michelle would accept the harm of a heart condition for the pure benefit of existence. If she does not have the heart condition, she could never regret the decision since she would not exist. It is not as if she will be losing something if the benefit of existence is not bestowed upon her. Rather, the only way a harm is introduced is if this pure benefit is introduced as well. Thus, since we cannot be sure she would agree to the heart condition, Ashley is still held responsible for the harm she does to Michelle.

In sum, I have argued that a harm was done in our case that would have been avoided if Ashley had used de dicto considerations to make her decision. However, Michelle’s complaint is not that Ashley harmed her future child de dicto as Hare argues. Rather, once Michelle is conceived, de dicto considerations wash out, leaving de re concerns to be assessed. Once conceived, Ashley de re harms Michelle through German Measles. Ashley assumes Michelle’s consent to harm her for a pure benefit thus violating her rights. In the end, Michelle’s worthwhile life does not justify the heart condition inflicted upon her.

**Benefits of the Theory**

The combination of the de re / de dicto distinction creates a complete moral theory in a way that neither approach achieved separately. As I argued from the start, a complete moral theory ideally works as both a decision making guide and a criterion of rightness. Within my proposed theory, de dicto concerns work well as a way to assess possible options and outcomes before one makes a decision and guides a moral agent in taking the best action. However, if one fails to make a morally right decision based upon de dicto assessments, de re concerns work to assess the wrong. The de re side of the theory allows the person or persons harmed to have a complaint against the moral agent, and works to uncover the harm done from choosing the de dicto wrong. Some may argue that I have separated the roles of a moral theory too much. They claim that I actually have two separate theories, rather than one theory achieving both roles. These thinkers argue that part of a moral decision must necessarily be considerations of what is right and wrong.

---

16 There are times where consent may be implied. For example, consent is implied when one is unconscious and needs surgery to save his or her life.
What I have done appears to be an after-the-fact criterion of rightness. However, what I find to be the criterion of rightness, de re considerations, must be considered when assessing what is de dicto better—my decision making strategy. What is de dicto better, and thus the choice that should be made, is inevitably whatever of the choices is de re less or not harmful at all. Given the use of de re results in the decision making role of de dicto, the theory is unified as one that accomplishes both goals.

I admit that the combination theory does not work alone to solve issues of different number considerations, nor does it make arguments for or against bringing children into the world with certain disabilities if there is not a better option. However, different theories surrounding these issues may be applied to allow the theory to handle any and all types of the non-identity problem. These outside theories go beyond the scope of this paper, but I find it beneficial for the combined theory to be able to utilize different answers to these issues in order to solve the non-identity problem.

After introducing the non-identity problem, this paper worked to assess Hare’s use of the de re / de dicto distinction as a solution to the problem. I found that Hare failed to show the real harms done when a moral agent failed to choose the de dicto best option. Therefore, I introduced a combined theory that worked to solve the problems that the de dicto approach faced. My theory also has the advantage of being open to different solutions regarding disability cases as well as different number cases. Finally, the theory works to explain commonsense morality’s answers to the non-identity problem. It is compatible with our intuitions and works to show the deeper morality behind them. The de re / de dicto combined theory thus succeeds as a complete moral theory in a way that previous person affecting and non person affecting theories have failed in the past.

---

\(^{17}\)My thanks to Kelly Sorensen for pointing this argument out.
For a philosophical sketch of education to succeed, it must first avoid the lure of reducing education to its epistemological exclamation point – to the proverbial and allusive ‘ah-ha’ moment of understanding. The ‘experience’ of education reveals itself to be, if nothing else, collaborative – an inter- and inner-personal event, uniquely capable of shedding light on the long-since muddied dynamics of body and mind, self and other, exteriority and interiority. The kinds of questions raised by an analysis of education reflect this same breadth: How fluid is the boundary between teacher and student? ‘Where’ does learning occur? How do the phenomenal experiences, mental capacities, and emotive intentions of teacher, student, and even the wider community, contribute to and alter the probabilities of pedagogical success?

To fully address each of these necessary questions, both in their particularities and moments of overlap, would demand more than our current space allows. Instead, here I hope only to offer one example of a thinker who wrestled with the spacious and inclusive breadth of education – a thinker no doubt struck by its decisive irreducibility to the ‘ah-ha’ moment. In the early text De magistro, Augustine invites readers into a uniquely dynamic investigation of education. Not only is the text formally didactic, but the analysis presented therein parallels the simultaneous education of one of its interlocutors, Augustine’s son, Adeodatus. As such, the reader is dually fortunate, confronted with at least two sketches for inquiry: a theory of education, and a performative one.

Yet, in spite of such richness, interpretations of De magistro are often limited to unpacking one small (though no doubt substantive) component of the text – the so-called ‘theory of Divine Illumination,’ the disclosure of the enigmatic Source of the ‘ah-ha’ moment of understanding. I would like to suggest that while Augustine may offer Illumination as a kind of epistemological climax, it is only one (albeit key) contribution to his wider, more

---

1 All citations, unless otherwise noted, are from: Augustine. De magistro. Translated by Peter King (Cambridge: Hackett Publishing Company, 1995).

2 While ‘uniquely dynamic’ it is far from ‘unique.’ The most obvious example might be Glaucon, educated alongside an analysis of education in Plato’s Republic.

3 The dialogue form was common (especially) for the early Augustine. One might also look at De ordine and De musica, to name but a few textual examples.

4 To be fair, the text suggests that Augustine is also ‘educated,’ is also a student – most notably, after the ‘entrance’ of the Inner Teacher.
nuanced account of education. Such a reading demands that we hold the author accountable to his own articulation of the project at hand, one not explicitly focused on knowledge (or even its acquisition) per se. On the contrary, a much broader subject-matter motivates De magistro, and with it the defense of a particularly pedagogical answer: “when we speak, what does it seem to you we want to accomplish [...] it is clear that by speaking we want to teach” (DM 1.1.5-8).

In light of this starting point, and given a hermeneutics of trust, the success of our task today will largely hinge upon the answers to three questions: (1) What is the accomplishment of speech; (2) What picture of education is suggested by the intention, the hope, of this accomplishment; (3) What is the role of the Inner Teacher within this broader sense of education, (i.e., in contrast to education understood as Divine Illumination alone)? Assuming what we might call the normative ‘patterns’ of Augustine’s philosophy writ large, this third question requires a special sensitivity. Although I am not suggesting that Illumination is tertiary or gratuitous for Augustine, I am even less suggesting that the role of the Inner Teacher can, somehow, be abstracted from education itself. On the contrary, my aim is to show that it is only, and precisely, by widening our interpretation of De magistro, and thereby our understanding of the picture of education presented by the text, that the import of the Inner Teacher is allowed to flourish with full Augustinian possibility.

5Of course, the use of the term ‘epistemology’ is anachronistic. I use it only for simplicity’s sake.

6The ‘problem’ of knowledge comes up most explicitly as the tangential concern of the hierarchy of knowledge (9.27-9.28).

7As such, what follows will be focused on ‘teaching’ and ‘learning,’ ‘teachers’ and ‘students.’ One could argue that this is not quite ‘education,’ but education reducible to its parts. Yet, my sense is not only that Augustine’s treatment of these ‘parts’ comes to comprise the ‘whole’ of education, but, furthermore, that hints of this ‘whole’ are suggested by the conclusion of the text.
Augustine seeks to uncover the accomplishment of speech by first determining the accomplishments of its building blocks—signs and sounds or, though somewhat of an idiomatic misuse, words and things. Beginning with the former, Augustine’s initial exchange with Adeodatus reveals their basic agreement about the ‘nature’ of words. Stating what he sees to be the obvious, Augustine proposes: “we are in agreement, words are signs” (2.3.1). As signs, words function relationally as flagships for things. In this way, the nature of words largely consists of their ability to attract attention for the sake of directing attention away, directing it towards the appropriate ‘thing.’

Fortunately for consequences of praxis, Augustine’s tone suggests that this referential aspect of words is virtually inevitable. His confidence can be seen in the playful example of *homo* (human). Turning to his son, Augustine asks:

*Augustine:* “... these two syllables [‘hu-’ and ‘-man’] conjoined are [hu]man. Will you deny it?

*Adeodatus:* Who could deny it?

*Augustine:* [...] are you those two conjoined syllables?

*Adeodatus:* Not at all, but I see where you are headed [...] You think it follows that I’m not a [hu]man” (8.22).

---

8 In contrast to other texts (such as *De dialectica*), in *De magistro* Augustine does not clearly distinguish between ‘words’ (a ‘sounding sign’) and ‘speech’ (the activity, the delivery, of ‘sounding signs’). However, it is clear that his initial analysis of ‘words’ is almost entirely devoted to the ‘sign’ of the word, ignoring both its ‘sounding’ and its ‘delivery.’ In contrast, and only after an analysis of such ‘sounding’ (i.e., an analysis of things), his concluding comments seem to retrieve a fuller, more ‘sensible,’ meaning of ‘speech’: as the activity of a human ‘teacher’ who gives a ‘sounding sign.’ As such, my treatment here, of first ‘words’ (loosely conceived of as ‘signs’) and only later ‘speech’ (as the activity of ‘sounding signs’) is simply intended to reflect Augustine’s ‘emphases.’

9 Augustine’s analysis of words is divided into two ‘kinds’: words which are ‘signs of signs,’ and words which are ‘signs of things.’ Given the eventual, albeit general, reduction of ‘signs of signs’ to ‘signs of things,’ I am beginning with the latter.

10 Cf. 8.22.72-74: “We can’t carry on a conversation at all unless the words we hear direct the mind to the things of which they are the signs.”

11 Cf. 8.23.95-110.

12 Given that the Latin is ambiguous, I have opted to translate ‘*homo*’ as ‘human,’ rather than as ‘man.’ This has the added textual benefit of maintaining the two-syllable effect towards which Augustine seems to be driving.
Of interest here is hardly Adeodatus’ humanity, although he himself (at least initially) might beg to differ. Rather, what intrigues is the emotive reaction, the offense predictably elicited by such a word game;\(^{13}\) for, even the accomplished philosopher cannot help but be startled when told that she is neither ‘in-tell-i-gent’ nor ‘con-vinc-ing.’ Augustine suggests that the impetus behind these reactions stems from the unremitting success of words – from the ability of signs to reflect things so compellingly that the distinction between sign and thing seemingly blurs. He explains that, “as soon as the words are uttered, I can’t help thinking that what is signified by the syllable[s] ‘[hu]-man’ is relevant […] so that once the signs are heard the attention is directed to the things” (8.24.148-151). As such, we (like Adeodatus) can hardly be blamed for our unrest – maintaining the mental distinction between word and thing is a task easier said than done.

Augustine’s understanding of the accomplishment of words wrestles with this difficulty, with discovering how one might avoid the unpleasantries that result from erroneously collapsing signs into their things.\(^{14}\) Most generally, the author’s strategy works to accentuate the epistemological and ontological ‘burden of proof’ for the word/thing pair. Augustine contends that words cannot reveal, cannot teach, the meaning of the things they signify.\(^{15}\) On the contrary, words are only meaning-full, only effective as signs, if the things which they signify have been previously encountered, if the things are already known. Words are not the source (the origin) of the matter (or content) of their signification. This responsibility rests with the things themselves. For, “we learn the meaning of a word – that is the signification hidden in the sound – once the thing signified is itself known” (10.34.157-159).

As a terminological aid, we might say that words are teleologically oriented towards things – their beginnings are indebted to their ends, their ends to their beginnings. This relationality comes to be reflected in Augustine’s final articulation of the accomplishment of words: words ‘remind.’\(^{16}\) Reminding suggests two things. First, the use of words does not mark a self-sufficient or original event; rather, to use words is to enact a repetition. To ‘know’ a word (and to know a word as a word) is to re-know, to re-member, a thing already-somehow present to one’s self. Words may remind us of things in more or less meaningful or successful ways, but the intention of these signs remains clear – words are

---

\(^{13}\) Cf. 8.24.145-147: “Then why is saying ‘Hence you are not a [hu]man’ offensive to the mind, since according to what we granted nothing more true can be said?”

\(^{14}\) The temptation of this collapse seems to motivate Augustine’s request that we “not attribute more to words than is suitable” (13.46.18-20).

\(^{15}\) The most noted example is that of ‘sarabarae,’ or ‘saraballae’ in the Latin Vulgate (10.33-11.37).

\(^{16}\) Cf. 11.36.1-2: “Words have force only to the extent that they remind us to look for things.”
the means to a greater end, a dynamic means, but one whose objective falls outside of its own boundaries.17

Secondly, reminding locates the agency of the accomplishment of words entirely within the individual.18 This is an important subtlety – for, it is not so much the thing itself that the users of words depend upon but, more precisely, the images of things as they exist within the contents of one’s memory.19 Although the success of words depends upon previous encounters with things, the use of words depends upon a subject’s encounters with things, upon the mental instantiations (the memories) of these experiences. Given the privatized nature of Augustine’s conception of memory – the inaccessibility of its depths and caverns to ‘others’ – the confinement of words to acts of reminding likewise restricts their meaningfulness to each individual alone (12.39.23-26). Even with memories seemingly held in common (such as the imprint of our most fundamental Desire),20 this commonality does not translate into public admittance.

It is largely due to these consequences that words fail to succeed as an exhaustive condition for the possibility of education. The author is plain: “most of all I’m trying to persuade you [...] that we don’t learn anything by these signs called words” (10.34.155-156). Augustine’s explicit argument for the pedagogical failure of words reflects a theoretical, rather than practical, concern. In short, he proposes what we might call a ‘theory of dependence’: given the dependent status of words (on things, memories, etc.), they cannot be expected to function as the sufficient (in-dependent) end of education.21 If anything is to account for education, it will necessarily be one of these ostensibly ‘independent’ entities, or, perhaps, some greater ‘whole’ not as of yet discovered.

However, more pointed concerns are hinted in the margins of the text. It seems clear that Augustine is hesitant to equivocate the general activity of reminding with that of education.

17Cf. 10.33.120-125: “The word doesn’t show me the thing it signifies [...] my conception of them wasn’t fashioned because they were named.”

18 Augustine’s respective use of the terms ‘anima’ and ‘mens’ as representative of the ‘human individual’ is nuanced throughout his corpus. In this paper, my language gestures towards the more general, anima, only to sidestep a deeper examination which time disallows.

19In acts of reminding, words do not literally recall “the things themselves, but the images impressed by them and committed to memory” (12.39.18-20).

20To be clear, Augustine does not suggest that we hold individual ‘wisdoms’ but that we seek one Wisdom individually. See Augustine, De libero arbitrio. Translated by Thomas Williams (Cambridge: Hackett Publishing Company, 1993). Hereafter, DLA.

21Cf. 9.25.1-6: “Whatever exists on account of another must be worth less than that on account of which it exists.”
A more ‘particularized’ account might satisfy many of us, and examples from the Western canon are no doubt on the tips of our tongues. Yet, given the neglect of any real specificity, the unanswered questions raised by reminding appear obvious enough: Are all acts of reminding to be deemed educational? Are we willing to sacrifice the assumption that education involves some sort of transformation, or even ‘addition,’ to the contents of one’s mind? What might it mean for the process of education to be entirely limited to the realm of the subject – for teacher and student to be rendered one and the same?

**Sounds/ Things: Exhibiting**

In light of these kinds of questions, in light of the limitations of words, Augustine shifts the focus at hand. The failure of words to serve the ‘whole’ of education has been chiefly presented as the more reductive failure of ‘signs to teach.’ Yet, the counterparts of signs, things, have not suffered this same fatality. While the sign of a word fails, the thing of a word – namely, the sound that occurs when a word is spoken – hints of greater educational possibility.

For Augustine, sounds, like all objects of sense experience, are things. As already hinted by his investigation into signs, the uniqueness (and the ‘power’) of things rests in their ability to reveal themselves, to reveal their ‘meaning,’ self-sufficiently. Things require no mediators, no detours of self-expression, and seemingly disclose their own ‘truth’ with relative ease – conceivably, because things (unlike signs) have only to indicate the content, to carry the epistemological weight, of themselves.

It is in the self-revealability of things that Augustine finds, at least a reflection, of education. When a thing presents itself, it does so as uniquely knowable. In fact, it presents itself as uniquely known. Augustine suggests that the self-display of a thing is virtually inseparable from the perception or cognition of that thing – in other words, for a (‘fully-functional’) rational being to gaze upon a sensible object is precisely to have the data of that object presented to the mind. As such, and barring unusual circumstance, the presentation of a thing equates to our learning of that thing from that thing.

This pedagogical emphasis is reflected in Augustine’s final determination of the accomplishment of things: things, including sounds, ‘exhibit.’ Augustine does not spend extensive time detailing the character of this ‘exhibiting,’ nor its resultant ‘sensible

---

22 Of course, a more particularized account of reminding might look something like the, equally ‘so-called,’ ‘theory of recollection’ presented in Plato’s *Meno*.

23 This is at least, or especially, true for the more ‘public’ senses of sight and sound. See *DLA*, Book II.

24 While later nuanced by Augustine’s seeming interpretation of the entire created world as a sign of God, this turn is not explicitly made in *De magistro*.

25 See *DLA*, Book II.
understanding’ (the knowledge that results from the self-display, the teaching, of things). His only real specification occurs by way of analogy – the exhibiting of a thing is comparable to the revelation of its boundaries. In other words, to ‘know’ a thing, to ‘perceive’ it, is to recognize that thing as distinct from the greater collective of things keeping its company – to recognize the pencil apart from the table it rests upon.26

Of greater significance to Augustine is highlighting the origin of the success of this perception: an origin which, he claims, resides with things alone. One might object that the ability of things to exhibit, to self-teach, is all the more dependent upon an anima: upon one who is able to collect the disparate data of a thing and transcribe it into some meaningful whole. Initially, Augustine does toy with the idea that individual intelligence is the true apex of the meaning, and the success behind the revealability, of things.28 However, his later tone is clear – no human capacity, neither sensation nor perception, can account for a thing’s exhibition. While the fluidity of the process of ‘sensible understanding’ surely helps to confirm Augustine’s confidence in its success, exhibiting remains an event of the sheer present-ness (the immediacy) of the thing itself.29 In terms of our ability to grasp the ‘truth’ of things, we learn and are, thereby, able to “answer [only] so long as the things we sense are present” (12.39.11-12, emphasis mine). Augustine’s consistency is confirmed in the opposite case: when a thing is no longer present, and an individual is forced to look upon nothing more than the residues in her mind, the influence of exhibiting dissolves. For, to resort to the images of things is to resort to signs, to return to the limitations of reminding.30

26 As King’s translation notes, the Latin (perceptio) maintains a tie to the ‘intellect’ in a way the English (perception) does not. As such, I am purposefully playful with my word choice (using ‘knowing’, ‘perceiving’ and ‘understanding’) as an attempt to be true to Augustine’s meaning.

27 Cf. 10.29.9-17: “For example, if anyone should ask me what it is to walk while I was resting or doing something else, as was said, I should attempt to teach him what he asked about without a sign, by immediately walking, how shall I guard against his thinking that it’s just the amount of walking I have done? He’ll be mistaken if he thinks this. He’ll think that anyone who walks farther than I have, or not as far, hasn’t walked at all. Yet what I have said about this one word applies to all the things I had agreed can be exhibited without a sign.”

28 For instance, in the example of bird catching, the paradigm of the sheer transparency of objects, Augustine explains: “I add that he’s so intelligent that he recognizes the kind of craft as a whole on the basis of what he has seen” (10.32.94-95).

29 Augustine explains, “I wouldn’t learn the thing I was ignorant of by the words spoken, but by looking at it [...] When I learned the thing itself, I trusted my eyes” (10.35.160-170).
However, in spite of this success, in spite of the fact that exhibiting presumably represents an 'educational moment,' Augustine is far from satisfied with the ability of things to grapple with the fullest range of pedagogical possibility. The capacity of exhibiting to self-display is not denied, and Augustine never refutes that "someone who presents what I want to know to my eyes, or to any of my bodily senses [...], does teach me" (11.36.3-5). Yet, notwithstanding this alleged triumph, Augustine remains unconvinced, suggesting that the 'mere' knowledge attainable in exhibiting (i.e., the knowledge of sensible things) is precisely that: 'mere.' Education need not be limited to the realm of sensible experience and, thus, cannot be limited to exhibition alone. Perhaps a different kind of exhibiting would suit, an exhibiting capable of revealing the presence of a different caliber of objects. Still, when restricted to its sensible formulation, the model limps.

While we may find ourselves more or less on board with Augustine's dismissal of 'sensible education,' other implicit concerns seem less contentious. For example, the nearly uncompromising accessibility of things, once again, calls into question the customary dynamic between teacher and student. The transparency of the meaning (the data) of exhibiting is so pronounced as to render such relationships incoherent; for, what can a teacher disclose that the thing itself has not already perfectly revealed? If anything, the correspondence between teacher and student is reduced to second-order gestures – gestures such as pointing, gestures about which Augustine is far from certain. In this way, it is the very ability of things (like sounds) to reveal themselves so perfectly that disengages exhibiting from the central concerns of education. Exhibiting, like reminding, is only able to enact a shortsighted response to the questions of Augustinian pedagogy.

**Speech: Prompting**

Having grappled with the limitations of both words and things, of both signs and sounds, Augustine teeters dangerously close to abandoning his project altogether; for, “what is more absurd than thinking that [someone is] taught by my speaking” (12.40.35-40). However, although Augustine toes the line, he does not cross it. On the contrary, the

30Augustine is frank: “When a question is raised not about things we sense at present but about things we sensed in the past, then we speak of not the things themselves but of the images impressed by them and committed to memory. I don’t know how we state truths even though we look upon these false [images], unless it’s because we report not that we are seeing or sensing but that we have seen or sensed” (12.39.17-23).

31Cf. 12.39.5-8: “What then can be said to show that we learn something by words aside from the mere sound that strikes the ears.”

32In other texts, such as the *Confessions*, things are presented not only as a matter of disinterest, but as a menacing threat: as a key source of the distentio of the anima. This ‘threat’ is less prevalent in *De magistro*.

33For, by pointing, or “by aiming the finger, I can’t know either the thing [...] or the sign” (10.34.145-150).
conclusion of *De magistro* turns to the combined front of signs and sound (of reminding and exhibiting), to the accomplishment of speech: speech ‘prompts.’

Augustine spends little time addressing speech in itself. Even less explicit is the time he devotes to prompting – the reader is seduced by only a few scattered passages, the majority of which are parceled between the distinctly human and divine roles that come to comprise prompting as such. And yet, these brief sections not only excite with the lightness of pen that has come to suggest great Augustinian seriousness, but they likewise allow for similarly rich interpretation. In both its human and divine expressions, Augustine’s sketch of prompting strikes with the rigor of its resolve and the certainty of its indebtedness to all the dialogue has, thus far, brought to bear.

By the end of *De magistro*, human teachers have been stripped of their title. However, they have not been estranged from the event of education. If anything, the text provides their role with a right clarity of purpose: through speech, human ‘teachers’ prompt. The utility of human prompting addresses a specifically educational concern: although we may or may not agree that education is irreducible to its epistemological finish, it is more difficult to deny that the normative ‘experience’ of education reflects a more intricate process (or, at the very least, a greater length of time). More specifically, to claim that education *ends* with a flash of understanding is to say very little of how that same education came to be. In prompting, in human speech acts, Augustine locates a kind of ‘beginning,’ a kind of educational catalyst: through speech, human ‘teachers’ prompt students (slowly and methodically) through their error (or partial understanding) and towards the arena of the possibility of right understanding.

How does this educational nudging come about? In one of his more Socratic moments, Augustine re-defines human ‘teachers’ as the facilitators of necessary, and appropriate, questions. The speech acts of prompting are, thus, specific in form. With the help of

---

34 I find King’s translation of *admonitio* as ‘prompting’ incredibly helpful, as it reflects the directionality of a ‘warning’ while remaining open to the activity of ‘reminding.’ It would be interesting to contrast Augustine’s use (or tone) of the *admonitio* in *De magistro* with its more prevalent formulation, that alongside *correptio*.

35 In other words, in as much as the nearest contender for the award of brevity is Divine Illumination itself.

36 Cf. 13.46.19-23: “As a result, we should by now not only believe but also begin to understand how truly it has been written on divine authority that we should not call anyone on earth our teacher, since there is one in heaven Who is Teacher of all.”

37 Augustine is clearly a thinker intrigued with ‘beginnings.’ For example: the ‘coming to be’ or ‘beginning’ of creation (cosmology), of the Christian soul (conversion), and of the self (autobiography).
prompters, a student is re-appropriated to questions through questions, a student is “prompted to [...] it part-by-part, when he’s questioned about the very parts that make up the whole” (12.40.45-50).

The result is two-fold. Through their questions, human prompters secure education, secure its ‘beginning,’ and do so outside of the realm of internal gymnastics. In other words, as prompting, education demands the human other and, thereby, serves to justify Augustine’s (no doubt budding) pedagogical community. Furthermore, and perhaps more importantly for those disagreeable to any traces of medieval orthodoxy, this community is one secured not by the content of its ‘answers’ but, rather, by its methodological form.39 The clearest example of this is proffered by the internal workings of the dialogue itself, as Adeodatus is prompted to learn by Augustine’s questions, by the many pages and twists of argument that come to comprise De magistro.40

However, prompting (as all speech acts, as all signs) is essentially relational. If the dialogue has revealed anything, it has shown that insofar as it is word-ed, the human side of prompting can comprise but one part of the whole of this model of education. Human prompting is necessarily linked to its ‘peak,’ to the ‘solution’ hidden behind its many questions, to that ‘thing’ towards which it ultimately provokes: namely, the Divine Illumination of the Inner Teacher. In fact, for Augustine, the work of human prompting is precisely intended to remind us of the Agent of Illumination.41 Following the way already paved by an otherwise digressive examination into signs, Augustine suggests that human prompters both remind us of ourselves (and the nature of our pedagogical limitations), and remind us of our Orientation (and the ground of right understanding).42 In spite of their blatant exteriority, the ultimate accomplishment of human prompters is, paradoxically, their ability to direct us to the hidden recesses within.

38Prompting, in its human articulation, seems very near to Socratic midwifery. See Plato’s Theaetetus, 149a-150d.

39This need not stand against Augustine’s treatment of the skeptics in Contra Academicos, a dialogue in which he insists upon the ‘content’ made possible by the Christian Christ figure. Rather, it seems to highlight Augustine’s remarkable ability to address the ‘concerns’ of his respective ‘audiences’ with great specificity.

40Adeodatus is explicit, “I have learned from the prompting of your words that words do nothing but prompt men to learn” (13.46.34).

41In other words, to claim that “words prompt us to consult Him” is, at least provisionally, to claim that ‘human prompters prompt us with words to consult Him’ (11.38.45).

42Cf. 11.36.15-19: “When words are spoken we either know what they signify or we don’t; if we know, then it’s reminding rather than learning; but if we don’t know, it isn’t even reminding, though perhaps we recollect that we should inquire.”
But what might we say of this Illumination, this fullest completion of prompting, this climactic ‘crown’ of education? While Augustine takes full advantage of the space allowed for having chosen a metaphor for his account (avoiding any specific remarks), a few general comments are possible in light of details also-already presented by the dialogue. Of particular interest is the striking familiarity of language with which Augustine clouds Illumination, a terminology nearly indistinguishable from that previously used to discuss the exhibiting of things. Although the force of Divine Illumination surely resides in its ability to extend human education to the category of eternal things, it nonetheless mimics the exhibitive method of the sensible – most notably, in the import given to the present-ness, the immediacy, of its revelation. With Illumination, “we perceive directly by the mind” (12.40.29-30, emphasis mine). The ‘epistemological problem’ of education, the problem of the allusive ‘ah-ha’ moment of understanding, is, therefore, largely broached as a problem of immediacy – how might a ‘truth’ be so very ‘at hand’ as to become one’s own, how might eternal things smack with the same transparency as the sensible realm? Illumination, and the most intimate of Teachers, allows for precisely this immediate disclosure of meaning. However, it is a realization indebted to fruits already harvested by the text.

The import of the ‘whole’ these pieces come to comprise (of human questioning/reminding and Divine Illuminating/Exhibiting) is clear: with prompting, Augustine is able to draw a picture of education begun on human ground while concurrently propelled deeper into the Divine. Prompting is able to wed the otherwise paradoxical elements of education, without collapsing the tensions between them: question and answer, human others and Divine Other, method and insight. As in the relationship between words and things, the temptation of such a collapse – of reducing education to illumination, or De magistro to the questions of modern epistemology – remains all too real. And yet, or, perhaps, precisely through this difficulty, stands the text itself, a dialogue whose intricacy serves as both an example, and exercise, of such restraint.

It is true that Augustine never explicitly says that prompting is the surest picture of education. However, he also never makes such a claim regarding Divine Illumination. In this way, by calling attention to prompting, I am only asking the reader to consider the interpretation that best makes sense (1) of the author’s self-stated project, (2) of the entirety of the dialogue (and all of its multiple parts), and (3) of our own experiences of education (in all of their gray, extended, nuance).

The Inner Teacher

However, Augustine has set himself, and his readers, for a different kind of interpretive conundrum. While Augustine does not claim that Illumination is the whole of education, he

43For a clear examination of the similarities between exhibiting (the revelation of things) and Illumination (the revelation of God), see Gareth Matthews, Thought’s Ego in Augustine and Descartes (Cornell University Press: London, 1992), 154-156.
does say that the Christian God is the only Teacher. Given that God is also the only Illuminator, does it not make sense to infer, along with the majority of medieval scholarship, that Illumination, that the ‘ah-ha’ moment, is the undeniable picture of education as it is presented by the text? In order for prompting to be a satisfying account of education, in order for it to succeed with all the theistic gusto of Divine Illumination, it must resonate with the activity of Augustine’s God.

Fortunately, especially for you generous readers of my otherwise very un-Augustinian analysis, Augustine offers precisely this divine consent. At the near end of the dialogue, Augustine adds a key detail to his sparse description of the Inner Teacher: “furthermore, He Himself will teach us [...] He who prompts us externally through men by means of signs, so that we are instructed to be inwardly turned toward Him” (13.46.20-25, emphasis mine). This expansion adjoins a crucial dimension to the divine involvement in education: the Inner Teacher is a Prompter, one not only engaged in the use of signs, but (as with human prompters) one engaged for the purpose of reminding humankind within. In this way, through prompting the Inner Teacher comes to stand with humanity in education (rather than only ‘after’ or, worse yet for Augustine, nowhere near). Furthermore, having already clearly marked the culmination, the ‘crown’ of education, it is the extension of the role of the Inner Teacher to prompting which allows Augustine’s God to re-comprise its ‘beginning’: a ‘beginning-before’ and a ‘beginning-through’ a human ‘beginning.’ Of course, this need not negate the role of human prompters, nor the reality of the experience of an all-too human ‘entrance’ into education. Divine prompting details, nuances, human involvement, rather than replaces it.

As such, the involvement of the Inner Teacher, the necessary ‘final Word’ on any appropriately Augustinian account of education, incorporates the totality of the model, and the whole of the text bearing Its name: the Inner Teacher serves as the prompting that is Reminding, the prompting that is Illuminating, and (we can only assume) the whole of Prompting that is ‘larger’ (more divine) than the sum of its parts. Is not this why Augustine sees his God as the only Teacher – not because Illumination is the pinnacle moment of education, not because God ‘is-only’ an Illuminator – but because only God ‘is-also’ the Illuminator, only God participates in the whole spectrum of human education?

---

44Cf. 11.38.45-48: “He who is consulted, He who is said to dwell in the inner man, does teach.”

45In other words, my argument is not intended to reduce Augustine to the later work of Aquinas, which tends to primarily focus on the ‘human’ (and subjective) element of education.

46If otherwise, we are left to ask why Augustine does not simply do away with all human involvement in education, a possibility that the work of his own life seems to blatantly disallow.
Conclusion

The conclusions that this interpretation might derive are most appropriate when formulated as questions. Might such a model of education rescue Augustine from the seemingly paradoxical claim that there are no human teachers, when clearly (though it hints of my own biases) Augustine is one of many such ‘teachers’ – human ‘teachers’ who continue to prod, inspire, and transform us all? How might prompting render education a co-creative event, an elaboration of imago dei whereby creature and Creator come to share a common vocation? And, finally, what more might be said about the relationship between speech and education, and the ‘usefulness’ speech\textsuperscript{47} gains because of it?\textsuperscript{48}

But for Augustine, no doubt the truest asset, and the crucial allowance, of prompting is its ability to unfold a ‘space’ for a question he had already asked – a space for the presence of a uniquely Christian Teacher, for One who not only reveals but works to reveal in each of the expansive reaches and intertwining parts of a multilayered account of education. This space, this God, was always-already ‘wider’ than a flash of light, more nuanced than a burst of sun through the clouds. On the contrary, Augustine’s hope for education demanded the ‘World.’


\textsuperscript{48}Cf. 13.46.15-20: “At another time we shall, God willing, look into the whole problem of the usefulness of words – which if considered properly is not negligible.”
I.  Introduction

Language is “a productive, recursive and infinite system,”¹ which means that from a finite number of rules and semantic primitives we are able to produce and understand an infinite number of novel sentences. The concept of an infinite number of sentences may at first seem to be an exaggeration, but we can explain that there is an infinite number of sentences in various ways. One such way is to appeal to language’s recursive nature, which allows us to embed clause within clause ad infinitum. For example:

John told me that Mary thought that Kim said that Arthur believed that I bought the kitten with the orange stripes that was born in April, but I did not.

The infinite nature can also be explained by language’s capacity to take as many modifiers as we would like to add, or how longer sentences can be made by continuously using conjunctions, but some may still find it hard to believe that from this we can produce an infinite number of sentences. An easier way to provide proof then is to start with a simple sentence such as sentence one.

(1.) I own one cat.

The determiner ‘one’ in sentence A can then be replaced with the name for any natural number 2, 3, 4, 5, n, n+1… and we have then constructed a sentence schema² which produces sentences that correspond to the natural numbers. Because there is always going to be some next natural number (n+1), we can always produce a novel sentence.³

What may be even more remarkable than the fact that the number of possible sentences is infinite is that humans, even at a young age, have the capacity to understand sentences that they have never encountered. As pointed out by B.H. Partee, the principle of compositionality in its most general form is uncontroversial and necessary for accounts


²The sentence schema does require modification for ‘cat’ to become the plural ‘cats’ when it is filled in with any number greater than one.

³Of course, this infinity is always a potential infinity, and we are constrained by physical limitations “that impose practical limits on the length of actual spoken sentences, whereas working memory imposes limits on the complexity of sentences if they are to be understandable... as well as other limitations—for example, on concept formation or motor output speed...” (Hauser, Chomsky and Fitch, 1571).
explaining our ability to understand novel sentences. The principle of compositionality can be formulated in many slightly different ways. A standard formulation is that “[…] the interpretation of a complex expression is a function of the interpretations of its parts. This is the principle of compositionality of meaning, also referred to as ‘Frege’s principle.’”

II. The Role of Compositionality in Davidson’s Meaning Theory

As most attempts to explain our ability to understand the novel sentences of a language that we encounter do, Davidson explains this remarkable feature of language by employing the principle of compositionality in his meaning theory. In this paper I shall demonstrate how Davidson’s definition of a semantical primitive fails, though it appears to be cleverly constructed.

The principle of compositionality looks at the components, or “semantical primitives” as Davidson refers to them, and syntactic rules for combining them. The combination of primitive involves more than mere concatenation, as hierarchical structure is also a necessary component of compositionality. More simply put, semantics can be determined from an understanding of the hierarchical combination of constituents in a grammatical sentence. Davidson takes this principle to mean that every construct of syntax is associated with a schema T sentence, and the meaning is then determined truth functionally.

Davidson’s meaning theory is modeled after Tarski’s criteria for a truth theory, which requires a satisfactory definition of truth to be materially adequate and formally correct. Material adequacy requires that the definition accords with the notion that truth corresponds with. Any materially adequate definition of ‘is true’ must entail (or “satisfy,”


6That compositionality requires something more than concatenation of primitives that we already understand can be seen easily even at the morphological level. ‘Unlockable’ is made up of morphemes: ‘un,’ ‘lock,’ and ‘able,’ yet it must mean more than just the stringing together of those meanings, as it can mean both “able to be unlocked” and “unable to be locked.” Moving to the sentence level, it is not (usually) the case that one refers to a transformational grammarian that he speaks of a grammarian who can transform his person. Meaning thus must be explained as consisting in hierarchical combinations.

as we see Tarski use) all instances of what Tarski calls a schema T sentence. Formal correctness requires that the definition work within the logical language and not be circular. A language that meets these requirements would be a formal language. Tarski is not confident that the same can be done in any natural language. Natural languages fall short of these requirements because they contain indexicals and ambiguous terms. Natural languages are not finitely axiomatizable, and in a formal language it is important that theorems are the only sentences that can be asserted.8

Davidson, however, is much more optimistic about natural language, and models his meaning theory after Tarski by using Tarski’s schema T to truth functionally determine meaning.9 In order to avoid the problems that Tarski predicts, Davidson identifies a criterion of adequacy for natural languages: that they be compositional. Davidson goes on to say:

Another, and more interesting condition is that we must be able to specify, in a way that depends effectively and solely on formal considerations, what every sentence means. With the right psychological trappings, our theory should equip us to say, for an arbitrary sentence, what a speaker of a language means by that sentence (or takes it to mean).10

Here we see that Davidson’s project includes an examination of the productive nature of language, and a language user’s ability to understand sentences that they have never heard before.

---


9Meaning is truth functional as it is based on the value of the biconditional of the schema T sentence. Davidson follows Tarski in this manner, though in Davidson’s meaning theory, truth is taken to be the primitive notion.

He presents a learnability argument, which is easy to follow in the form given by Lepore and Ludwig.

1. Natural languages have an infinite number of nonsynonymous expressions, and, in particular, an infinite number of nonsynonymous sentences.

2. We can (fixing our cognitive capacities and life spans) learn natural languages (since we have done so.)

3. “...we do not at some point suddenly acquire an ability to intuit the meanings of sentences on no rule at all.”

4. “...each new item of vocabulary, or new grammatical rule, takes some [minimum] finite time to be learned.”

5. “...man is mortal.”

6. Thus, natural languages have “a finite number of semantical primitives” [1-5].

7. Thus, it must be possible to account for how an infinite number of nonprimitive expressions in natural languages can be understood in terms of a finite vocabulary of semantical primitives [1, 6].

We see from Davidson’s learnability argument that he refers to the composite parts of language as “semantical primitives.” As we shall see, his definition is quite clever, yet leaves many questions. He defines them as follows:

Let us call an expression a semantical primitive provided the rules which give the meaning for the sentences in which it does not appear do not suffice to determine the meaning of sentences in which it does appear. Then we may state the condition under discussion by saying: a learnable language has a finite number of semantical primitives.

What is very clear is that there are a finite number of semantical primitives. We can understand that this must be the case, or it would not seem that natural languages are learnable. The other part of his definition is clever as it provides a recursive procedure that is meant to determine if any possible case is a semantical primitive or not. The only further information he gives about semantical primitives is negative. Davidson tells us that they are not quotation marks, indirect discourse as described by Scheffler, belief sentences as described by Quine, or Church’s sense of denotation. This does not help us to


understand his positive view of what semantical primitives are, but is a reaction against what others had claimed could be semantical primitives in the past. Lepore and Ludwig seem to take Davidson’s definition to speak for itself, as they again just re-emphasize their finite number.\textsuperscript{14}

\section*{III. Four Reasons to Reject Davidson's Definition}
Upon examination, Davidson’s recursive definition of a semantical primitive appears to be problematic. I believe there are (at least) four reasons why this definition should be rejected:

1. There is the empirical problem of the difficulty in its application
2. It is unclear what the “rules which give the meaning” are.
3. Any attempt to understand what these rules may be leads to circularity.
4. Do-insertion is an example where his definition clearly fails.

As I have said, Davidson’s definition is quite clever as it is meant to handle every possible case. Examples aid in clarification:

(2.) That black cat is scary.

First we can take an easy case. Suppose we want to know if ‘black’ is a semantical primitive. ‘Black’ is a semantical primitive “provided the rules which give the meaning for the sentences in which it does not appear do not suffice to determine the meaning of sentences in which it does appear.” Then we can take a sentence in which ‘black’ does not appear:

(3.) That cat is scary.

It seems simple enough to say that the rules which give the meaning to sentence 3 do not suffice to determine the meaning of sentence 2. Nothing within sentence 3 provides information to determine the meaning of ‘black’ in sentence (2). Thus, it seems that ‘black’ is a semantical primitive.

Next we can pull out the constituent ‘that black cat’ and apply Davidson’s test to tell if it is a semantical primitive. ‘That black cat’ would be a semantical primitive “provided the rules which give the meaning for the sentences in which it does not appear do not suffice to determine the meaning of sentences in which it does appear.” This does not seem to be the case for the whole constituent, so we can say that it is not a semantical primitive, which

\textsuperscript{13}Ibid., 9-15.

\textsuperscript{14}Lepore and Ludwig, Donald Davidson, 29.
makes intuitive sense, as a constituent composed of multiple words should not be a primitive.

Upon closer examination, Davidson’s definition begins to seem problematic in application. Let us return to the seemingly easy sentence (2). Davidson’s definition appears to work well when ‘black’ is the word we choose to test. But suppose we want to test ‘is’ instead. If we set up our test in the same way as we did with sentences (2) and (3), we would then look to sentence (4) for an answer.

(4.) *That black cat scary.

Here we run into a problem. While it was easy to test ‘black’ by using sentence 2 minus that word, this strategy does not work to test ‘is,’ as we are not left with a grammatical sentence. I think it is safe to assume that when Davidson writes of “the sentences in which it does not appear” he has grammatical sentences in mind. Since simply extracting the word we want to test leaves us lacking a grammatical sentence in this case, we are left to conclude that this method for application of the definition will not suffice.

It may seem that this problem is easily fixed, as Davidson’s definition does not speak of a sentence, but “the sentences in which it does not appear.” We have seen that without using the word ‘black’ in a sentence we cannot determine its meaning in sentences in which it does appear, and the same for ‘cat,’ though this works only if we are charitable to Davidson in our understanding of his definition. Allowing for a plurality of sentences actually makes it more difficult to include the components that intuitively belong in the set of semantical primitives. I think that it may be possible to understand the meaning of ‘black’ on some level by giving a description not using the word itself, but by giving its mathematical properties. The word ‘that’ may appear to be slightly trickier to propose as a counterexample to the utility of Davidson’s definition, but it seems easy to understand the meaning of ‘that’ only from sentences in which ‘that’ does not appear. It can be explained (for this type of usage; excluding clause introduction usage) as a demonstrative, extensionally picking something out, namely the black cat to which we are referring to. This is again problematic, as I think we’d like to say that ‘that’ is a semantical primitive.

Davidson’s formulation of the definition, which talks of the sentences in which whatever we aim to test does not appear leaves room for many possible sentences which can explain that thing without using it. At best, we are left with a difficult decision procedure to carry out for practically any component we wish to test. At worst, we are left with a dearth of semantical primitives.

---

15Here we run into a problem for this particular sentence in English. It should be noted that this particular problem may not arise in some other languages that do not require the use of the copula. However, as Davidson wrote in English, and most likely conceived of his definition of a semantical primitive using English language examples, the counterexample is still problematic. More so even than if the counterexample existed in a language other than Davidson’s native language.

16Italics are mine.
That being said, let’s return to the definition again with the earlier more charitable interpretation in mind. The second problem arises here. I think more charity is required than originally thought. Not only do we need to look with charity to the sentences in which our alleged primitives do not appear, but charity is also needed to understand what is meant by “the rules which give the meaning.” Davidson never elucidates what these rules are, so we are left to come to some understanding of this on our own. I don’t believe that there is any reasonable way to flesh this part of the definition out.

I say no reasonable way, as it seems that a common sense understanding of what Davidson might have had in mind leads us to the third problem. In fact, the only understanding I can come up with for “rules that which give the meaning” leads to circularity, once we recall that we are looking for rules that give the meaning to something longer than just a word. If the meaning needed here were lexical meaning, we might not be confronted with circularity, but Davidson’s holism lets this way out quickly be eliminated.

The circularity problem arises when Davidson’s definition of a semantical primitive is examined together with the definition of the principle of compositionality. The definition of a semantical primitive relies on rules which give the meaning of sentences. Yet, the principle of compositionality is supposed to yield meaning from a combination of the parts (primitives) that make up the expression (or sentence) together with its syntax. This means that some notion of meaning is required to determine what is a semantical primitive, yet meaning isn’t yielded from the principle of compositionality until after we examine the primitives and the syntax of an expression. Thus, we need meaning to yield meaning, which makes Davidson’s definition circular.

Finally, we come to the fourth reason why Davidson’s definition fails. ‘Do’ is a special case for a few reasons. First of all, English contains two different verbs ‘to do’. One roughly has the meaning “to perform some action” (which could be any action), and the other seems to have no meaning at all. The second ‘to do’ verb in English is described as “a dummy (meaningless) auxiliary, which is inserted under “do-support”.

The first case is difficult to handle because of the fact that this verb ‘to do’ generically stands in for any action, as it means to perform some action. We can look at the simple sentence “I will do so.” A compositional meaning theory can at best give the meaning of some sort of affirmation. However, what action is being affirmed is unknown from the semantical primitives and their syntax alone. This is because of the ability of ‘do’ to mean so many things. The sentence could mean various things ranging from “I do take so and so’s hand in marriage,” to “I do own a cat,” to “I do watch House every Monday night,” and these by no means exhaust the list of possible meanings.

---

17 Carnie, Syntax, 264.

18 This sort of problem in fact applies to many verbs. For example, consider ‘run,’ which can range in meanings from “run a mile,” “run a machine,” “run a test,” “run a risk,” and so on.
The second case is also problematic, as the verb 'to do' of do-insertion/do-support is a dummy auxiliary with no meaning. Use of this second 'do' verb fulfills purely syntactic constraints imposed by the parameters of English when used in question formation and negative sentences. Some languages handle question formation with the inversion of main verbs, and others require the addition of a question complementizer, such as 'an' in Irish. For example, in French it is perfectly grammatical to say:

(5) Aimez-vous la citronnade?

However, the parameters settings of English differ from those in French. In English, yes/no questions use subject-aux inversion, but when no auxiliary is present, the dummy verb 'do' is required. This is seen in sentences (6a) and (b):

(6) a. *Like you lemonade?
   b. Do you like lemonade?

The question is only well-formed when the dummy verb 'do' is present. Similarly, we can see the same phenomenon in English negative sentences (7a) and (b):

(7) a. *I not drank the lemonade.
   b. I didn’t drink the lemonade.

In the cases shown, 'do' is meaningless,¹⁹ and thus hard to fit under Davidson's definition for a semantical primitive, as meaning is required for the definition's application. Yet if it is not a semantical primitive, but it is needed for proper syntax, when would this necessary word be added back into a constituent or sentence in a compositional meaning theory?

IV. Conclusion

These reasons are sufficient to show that Davidson's definition of a semantical primitive cannot be correct. It is likely the case that Davidson, while probably accepting the principle of compositionality, only explicitly brings it up in his writing to satisfy Tarski's criterion of adequacy. If this is true, it would explain why the definition he gave seems so clever in theory, but fails to work in application. For these reasons I find fault with Davidson's theory, and perhaps compositional meaning theories more generally. It is my intention that such criticism opens the door for examination of alternative theories of meaning that can be used in ways that Davidson's cannot.

¹⁹Of course, 'do' is not meaningless, but Davidson's definition leaves no room for functional meanings.
Chelsea Ruxor – Externalist Foundationalism

Understanding the nature of knowledge is of primary importance because it is of such great personal significance to establish justification for thinking we are able to interact in the world as intelligent agents; without genuine justification for our knowledge, it is impossible to apply our beliefs and experiences to an external world or to consider any field reliant on our ability to know things, from the fundamental laws of the sciences to complex ethical theory. In Epistemology, Classic Problems and Contemporary Responses, Laurence Bonjour examines the necessity of justification and explains how we might utilize certain a priori truths as a basis for further knowledge. If we are to establish adequate justification for some of our beliefs, however, it seems that the notion of a priori knowledge must be expanded to include some beliefs and internal states which are directly caused by external sources. Just as foundational knowledge is necessary to verify the truth value of our beliefs, externalism must anchor our internal experiences in a real external world, and considering the role of both internal processing and external causes should give us strong justification for saying we know things about ourselves, our experiences, and the world external to us.

Bonjour explains how one can use certain foundational beliefs, which do not require justification by appeal to anything else and consist of “a person’s immediate awareness of his or her own conscious states of mind, together with his or her a priori grasp of self-evidently true propositions,”¹ in order to gain further knowledge. Among the most fundamental forms of a priori knowledge are those solely dependent on internal processing, our direct access to our internal states and sensory experiences and capacity to understand semantic associations independently of experience. Surviving Descartes’ method of doubt is proof that we have direct knowledge of our immediate internal experiences, our sensations and our perceptions, about which we cannot be mistaken. We also seem to know some things by virtue of their meaning. Our knowledge of propositional content or of relations of ideas does not rely on our sensory experiences, but rather on the sort of reasoning that holds true in all possible worlds and at all times. We know, for instance, that something that is alive is not dead or that all mothers are parents merely by virtue of the meanings of the words and without any experience with dead or alive things or with parents.

Epistemological foundations, which are known to be true without reference to anything else, are vital to the justification of further beliefs. If we are to accept that the things we

know are true, representing some sort of objective reality external to our minds, they must
be justified on the basis of such a priori truths rather than through a merely coherentist
approach to our belief system. These foundational beliefs, whose truth value does not rely
on appeal to any others, establishes a basis for further inquiry in that more extensive
beliefs may be justified by appeal to them.

Beyond the fundamental truths that establish knowledge of our internal experiences, it
seems that we can use knowledge directly accessible to us to verify the existence of some
sort of external world. Bonjour explains John Locke’s “ideas of sense”\(^2\) and considers how
direct sensory experiences can be considered as evidence of an external world. Even if we
tentatively adopt the idea that an evil genius like the one Descartes considered could
theoretically exist and be controlling our experiences, it seems that there must be
something external to our minds which causes many of our involuntary internal
experiences; while our beliefs, experiences, and memories may greatly alter our
perceptions, there are nevertheless many aspects of our immediate experiences which do
not appear to be the result of conscious mental states. We do not have to be thinking of a
tree, for instance, in order to have a visual experience of it, nor need we believe in a tree at
all in order to hear the wind blowing through its leaves or feel pain when hit by one if its
branches. It seems that, regardless of whether it is an evil genius or a world which more or
less reflects our perceptions of it, our internal states are at least partially affected by
something external to our minds.

Once we have established the existence of some kind of external world, it is evident that
there are actually many things we know about it independently of our experiences. This
kind of knowledge applies not to abstract relations of ideas, but rather it establishes
necessary features of the universe. For instance, we know that it is impossible for
something to be both red all over and green all over, though nothing about the words “red”
or “green” leads necessarily to this conclusion.

Because it is possible to know such features of the world a priori, these beliefs, along with
our sensory experiences and memories, may allow us to draw certain conclusions
regarding actual instances within it. Although our memories are sometimes inaccurate and
there is likely an even greater margin of error in our sensory experiences consistently and
accurately reflecting things that are true of the external world, it seems nevertheless that
we can obtain very strong justification for beliefs which are contingent upon one another.
For instance, a student may use his grasp of what a sidewalk is, a stationary object which
one can walk on and in general does not talk to squirrels, fly, or do anything else non-
sidewalkish; together with his perceptions of the strip of concrete beneath his feet, which

\(^2\)Bonjour, 130.
looks and acts like a sidewalk; and inductive reasoning, his experiences with both this sidewalk and others; to rationally justify his beliefs about the thing on which he is walking.

The use of inductive reasoning to support beliefs regarding things in the external world, however, has encountered a great deal of skepticism by such figures as Hume, who suggests that the circular nature of the argument for induction yields no justification for trusting it as a means of attaining knowledge. Bonjour explains that induction “has to do with what reasons or justification there are for accepting general conclusions on the basis of observations of particular instances falling under them.”3 For instance, the student might infer that, because all the sidewalks in Evansville remain stationary, he should not expect the sidewalks at Purdue to levitate or those in London to slide around to different locations to confuse tourists. Induction relies on the instances we observe in the world on a regular basis holding true throughout the universe and the future being generally like the past.

Hume’s suggestion is a serious form of skepticism in that it deals not only with having absolute knowledge of principles or sidewalks in the external world, but in that it threatens our having any sort of rational justification for believing anything on the basis of inductive reasoning at all. Bonjour explains, “Hume’s claim is not merely that such inferences are not conclusive, that we cannot be completely certain that the conclusions are true...It is rather that inductive inferences yield no justification at all for their conclusions, that is, that they fail to increase or enhance to even the smallest degree the likelihood that their conclusions are true.”4 This is a very serious claim not only against knowledge, but against the very manner in which we are able to obtain justification for our beliefs regarding the external world.

Despite the implications of Hume’s argument, there seems to be a means of justifying the use of certain instances of induction that we know a priori. In considering our justification for knowing certain necessary features of the world, we observe general patterns in the manner in which it operates. We know, for instance, that something that cannot be red all over and green all over today cannot be simultaneously red and green tomorrow, and that a shirt that is solid cannot also be polka-dotted, even in the summer or in Europe. These patterns seem to imply that there are certain laws by which the universe operates which provide strong justification for trusting some kinds of induction. We should conclude that shirts cannot be solid and polka-dotted in France and that things were not simultaneously red and green in the 1970’s, just as sidewalks do not fly at Purdue, even though we may not have encountered all the red and green, polka-dotted, or sidewalk-like things in the universe.

3 Bonjour, 53.

4 Bonjour, 61.
While it does not guarantee complete knowledge of the actual thing under his feet in that his sensory experiences could be deceiving him, the inductive reasoning the student employs while walking on the sidewalk nevertheless seems to be an important means of furthering his justification for his beliefs. Because our strongest connection with the external world is through our sensory experiences of it, it seems that induction is necessary to understand and to interact with certain aspects of the world. Using his understanding of what a “sidewalk” is and any internal or a priori knowledge he may have regarding sidewalks, walking, etc, would be of little practical assistance to the student if he could not have a visual experience of the concrete thing or the sensation of feeling something hard and stable beneath his feet. If his justification for his belief that he is walking on a sidewalk does not rely, at least to some degree, on his sensory experiences, the student could have a sensation of drowning in a swimming pool, flying, or trying to get out of quicksand and still be equally justified in his belief that he is safely walking to lunch on an ordinary, student-friendly sidewalk.

Induction, as well as the semantic associations, contingent beliefs, and other forms of justification discussed above, which we know on the basis of our reasons, however, cannot account for all of our true justified beliefs of the external world; there are some things we are caused to believe or experience which are beyond the scope of rational justification. Our knowledge of a specific tree, for instance, relies not only on the tree-like image we conjure when we think about it and on internal processing and semantic associations, but on a real, external tree which causes our experience and serves as a basis for our beliefs about it. In this manner, some things are known not only by relying on reason, inferences, arguments, or careful considerations of our beliefs and their implications, but rather on their being reliably caused by the external world.

These instances of knowledge seem to be caused in the direct sense proposed by externalism rather than by a series of intermediate interactions. In order to have knowledge of a tree, it must be the tree that causes our immediate visual experience and sets off the series of mental states relating our physical sensations of it to our semantic identification of it as a “tree.” In a similar manner, the student’s valid justification for thinking he knows things about the sidewalk is dependent on the particular sidewalk he is experiencing directly causing him to have certain immediate sensory experiences. Externalists rely on some things in the world being reliable causes for beliefs which we have, and indeed know, without rational justification—what makes them true are their causes, not our reasons.

If these causes are indeed reliable, it seems they may be utilized as foundational beliefs for further inquiry regarding the things they represent. Much in the same manner as the foundational beliefs established by internal processing, those things like semantic associations which are known independently of experience, external causes are necessary
to ground our experiences in an objective reality and serve as a sound basis for further knowledge regarding it. If he is to attain knowledge of something external to him, the student must be able to justify his beliefs about the actual thing on which he is walking and his ideas of this particular sidewalk not by appeal the sidewalk in his head and his semantic understanding of what a sidewalk is, but rather by the real sidewalkish thing in the external world which causes his experience.

Bonjour finds fault with the externalist view of justification primarily due to the difficulty of showing one can know something without being able to rationally justify his reasons for knowing it. It is commonly accepted that true knowledge cannot be the result of an “accident,” and it may be that the world is not such a reliable cause, therefore rendering our beliefs insubstantial. For instance, there being horses being in a field is only one possible cause for our believing that there are horses in a field; we may be watching a movie about horses in a field, dreaming about them, or imagining them, or we may mistake horses for mules, cows, or muddy zebras. Bonjour suggests that proving that external causes are reliable depends on cognitive processes which are at best “merely hypothetical and insecure as long as they cannot be arrived at from the resources available within a first-person epistemic perspective.”

It seems, however, that to give up externalism altogether is to give up our very conception of knowledge—mainly, that it is true, contingent on a real external world. Because most of our immediate experiences are internal, it is necessary for them to causally relate to things in the external world if we are to attain any sort of justification for beliefs regarding things beyond our conscious mental states. For instance, while our full awareness of our internal states may allow us to acquire complete knowledge of the visual experience we are having of a sidewalk, it is necessary for there to actually exist some sidewalkish thing in the real world which is directly causing our experience if we are to have knowledge of it. It seems, furthermore, it is this physical cause, not the sidewalk in our head, that must serve as the foundation for further beliefs about the sidewalk. Without a reliable connection between external causes and sensory phenomena and the establishment of foundational beliefs grounded in a real world, there is little hope of justifying our belief that our experiences represent any kind of objective reality external to our minds.

Just as we require certain internal experiences, our sensation and perception of the sidewalk in order to acquire knowledge of it, externalism is necessary to establish that these sensations constitute a genuine causal connection with the sidewalk outside of our minds. Otherwise, all of our successful interactions with a supposed external world are completely accidental and, worse yet, the things we think we “know” about it have no basis in a real external world at all. Without first establishing this causal connection between

---

5Bonjour, 236.
our mental experiences and physical phenomena, the sidewalk in our head can only refer to the sidewalk in our head, and there is no secure means of establishing a basis of any kind of reality in our experiences. If we are unable to obtain “true,” non self-referential knowledge, furthermore, we cannot consider any kind of truth value of our beliefs, and it is clear we need not bother validating the rational justification we have for believing in the sidewalks in our heads.

It seems that reliance solely on externalism for foundational beliefs also proves insubstantial. If we consider only external causes and do not take into account the role of internal processing in obtaining knowledge, we must attribute the same kinds of knowledge to the sidewalk as we do to ourselves if it is subject to the same external causes as we are. If only causes matter, we should be able to obtain complete knowledge of the external world simply by experiencing it, and both people and sidewalks everywhere should have a comprehensive understanding of the universe. It is equally evident, however, that reliance on internalism alone ignores the reasons we have for seeking knowledge at all. Strict internalism forces us to accept that our studies of science, personal ethics, and even our understanding of our lives are irrelevant, just as hypothetical and insecure—if not more so—than Bonjour’s suggestion of those beliefs arrived at by means of external causes.

It seems, therefore, that it is impossible, and indeed incoherent, to attempt to justify our knowledge of the external world without considering elements of both internalism and externalism; even perfectly accurate perceptions of the world cannot aid us in gaining knowledge if we are unable to process and understand them, and the kind of “knowledge” we obtain by means of internal processes alone are at best self-referential and unconnected with any sort of objective reality. Despite Bonjour’s claims against external causes, it seems that we can be rationally justified in accepting some instances of externalism, and in fact that we must do so if we are going to talk about sidewalks and trees that exist outside of our heads or consider our rational justification for believing anything about the external world at all. These causes, in turn, may serve as foundations for further knowledge. It is necessary to utilize both external and internal foundations if we are to have genuine knowledge of the world, and establishing the validity of these foundational beliefs is vital to grounding our experiences in an objective reality.
Abstract: The question of the coherence of William James's “Will to Believe” argument depends on how we understand James's religious hypothesis. This paper will consider the contrasting interpretations of Robert O'Connell and Ludwig Schlecht and argue the neither satisfactorily correspond with the context of “The Will to Believe.” While O'Connell errs by reading James in overly Christian terms, Schlecht’s interpretation does not synthesize with all presented aspects of William James’s religious hypothesis. Finally, I will offer an alternative interpretation which better synthesizes with the immediate context of the “The Will to Believe.”

Is William James’s “Will to Believe” (WB) argument consistent with the concept of religion which is appears in the rest of his work? Essential to his argument is the assertion that we should adopt a believing attitude when “faith in a fact can help create the fact.” 1 Faith, on James’s terms, was necessarily manifested in action. It produced the necessary behavior which could bring about the realization of the object believed in. This aspect of faith was more than just a trump of the positivistic prohibition on unverified belief; it was, for James, an essential aspect of the religious hypothesis. “This feeling...that by obstinately believing that there are gods...we are doing the universe the deepest service we can, seems part of the living essence of the religious hypothesis.” 2  

Robert O’Connell’s, William James on the Courage to Believe, is an attempt to reconstruct James’s argument so as to make it plausible for a traditional Christian theology.3 O’Connell claims that James failed to adequately distinguish between “outcome cases” and “over-beliefs.” Outcome cases are situations where a belief in a certain outcome can be a necessary element of that outcome obtaining.4 However, in what are called weltanschaulich issues, or “over-beliefs”, such as the existence of God or the meaningfulness of moral statements, James may have been taken by “momentary seizure of optimism” and have

---


2Ibid., 31.


been “carried away too far” by claiming that one’s faith in a certain fact could help create that fact.  

James’s definition of religion is rather ambiguous, and it may be possible to understand his religious hypothesis in such a manner that what O’Connell thinks is an overstatement is no overstatement at all. Ludwig Schlecht criticizes O’Connell for interpreting James in Christian terms and claims that James’s primary religious concern was meliorism. Schlecht draws from the latter half of James’s work to develop an interpretation which avoids O’Connell’s criticism. He reduces James’s religious hypothesis to the assertion that the universe has substantive moral meaning and that by adopting a strenuous moral mood, we can help create this fact. On those terms, Schlecht says, James’s argument stands as is. 

It is this author’s contention that Schlecht’s interpretation does not do justice to the immediate context of the WB argument. I will argue that Schlecht illicitly discards what James saw as essential aspects of the religious hypothesis. If Schlecht is correct that James cannot, and should not, be construed as an orthodox Christian theist, neither can meliorism be held as the sole element of James’s religious hypothesis. Using sources published contemporaneously with the WB, I will show that James’s concept of religion did entail propositions, or “over-beliefs,” which cannot hold the weight of certain aspects of his WB argument. We, like O’Connell, are forced to modify and clarify certain elements of James’s argument if we hope to arrive at a suitable understanding that is faithful to the context of the WB.

The Will to Believe Argument

James, as a self described “radical empiricist,” regarded experience as the only source of content for philosophy. However, he rejected any account of experience which excluded the human component and contribution to the universe. Radical Empiricism is a significant part of James’s WB doctrine due to its large implications for the role of human activity and belief in shaping the world. It is also a crucial part of James’s pragmatism. For the pragmatist, the truth value of propositions is worked out in experience. When confronted with a proposition the pragmatist asks “What concrete difference will its being true make in anyone’s actual life?” The active change in behavior that results from accepting a

5 O’Connell, 163.


7 Stanley Hauerwas, With the Grain of the Universe: the church’s witness and natural theology: being Gifford lectures delivered at the University of St. Andrews in 2001 (Grand Rapids: Brazos Press, 2001), 59.
proposition will either bring about its verification or falsification. “True ideas are those that we can assimilate, validate, corroborate, and verify.”\(^8\) When James speaks of a religious hypothesis being shown true, what he means is that accepting such a hypothesis will result in action conducive to its verification. Belief in a certain truth necessarily results in a change in behavior. “Cognition, in short, is incomplete until discharged in act.”\(^9\) Thus the thinker, or participant, plays a vital part not only in the observation of facts, but in their verification and actualization.

The role of the passion is indispensable in determining truth. This was one of James’s fundamental disputes with those who claim “objective certainty” through reason, and that reason is the final authority for the will. For James, the thinker was not a passive mirror which reflected the outside facts, but the “actor and coefficient of the truth.” “Mental interests, hypotheses, postulates, so far as they are bases for human action—action which to a large extent transforms the world—help to make the truth which they declare.”\(^10\) Since the passion takes part in the verification of the facts, it cannot be derided to the position of a deceiver or hindrance. Any positivistic rule which would exclude the passion from the belief making process is illegitimate. James’s goal in the WB lecture is to show just this.

“*Our passional nature not only lawfully may, but must, decide an option between propositions, whenever it is a genuine option that cannot by its nature be decided on intellectual grounds.*”\(^11\) This is James’s thesis in his WB lecture. The target of this argument is positivists like W.K. Clifford who assert; “it is wrong always, everywhere, and for anyone, to believe anything based upon insufficient evidence.”\(^12\) But is Clifford consistent? James asks whether the

---


\(^9\) Ibid., 169.

\(^10\) James, “Sentiment,” 72.


\(^12\) James, “Will,” 14, 15, 20. A “genuine” option is one which is live, forced, and momentous. A live option is one which appeals, even in the smallest manner, to the passion or intellect of the thinker. A momentous option is one which has a limited opportunity to be chosen. Finally, a forced option is one which cannot be avoided. There is no third option between the two alternatives. If the option is forced, one cannot simply neglect to decide, for to do so would result in behavior identical to one of the original alternatives.

The proposition that “truth exists” is something which can survive Clifford’s principle. “Our belief in truth itself, for instance, there is a truth, and that our minds and it are made for each other—what is it but a passionate affirmation of desire?” 14 If the skeptic asks for the evidence for such a conviction, what can one say? “As a rule we disbelieve all facts and theories for which we have no use.” Clifford has no use for Christian beliefs; just as scientists disregard the evidence for telepathy because they have no desire that it be corroborated. Clifford’s principle is based upon no more than this, his own willful predisposition to forbid all propositions for which he has no use. 15 The intellectualist’s rule to abstain from belief until all the evidence is “in” is no more than a manifestation of his own passionate desire not to be fooled. Forbidding any influence of our will, in regards to the desire to have knowledge, is equivalent to a “general informing his soldiers that it is better to keep out of a battle forever than to risk a single wound.” 16

James is not giving a license for wishful thinking. There are two final conditions which a hypothesis must meet to qualify for the WB doctrine. First, it must be a desirable thing that this hypothesis be true. Second, and more significantly, it must be the case that adopting a belief ahead of verification will help bring it about that the hypothesis obtains. 17 A genuine option which, by its very nature, cannot be decided on empirical or intellectual grounds is valid for belief only if that belief is one which contributes to actions which help verify the fact. Thus we have a specific class of beliefs for which this condition applies. “The future movements of the stars or the facts of past history are determined now once for all, whether I like them or not. …in all that concerns truths like these subjective preference should have no part.” However, “In every fact into which there enters an element of personal contribution on my part,” it is better to believe in the desired result, for doing so may help create that result. “Thus the future fact is conditioned by my present faith in it.” 18

But do our personal contributions enter into the questions of religion? This depends, of course, on what type of facts the religious hypothesis entails.

**James’s Religious Hypothesis**

James’s religious hypothesis in the WB consisted of two rather vague and ambiguous propositions. The first was that “the best things are the more eternal things, the overlapping things, the things in the universe that throw the last stone, so to speak, and say

---

14James, “Will,” 19.

15Ibid., 19.


18James, “Sentiment,” 80-81.
the final word.” Second, is that we, as humans are better off now if we believe the first proposition to be the case.\textsuperscript{19} This does not at all seem to resemble what one thinks of when religion is considered, but James’s immediate task is not to defend a particular religion, but religious belief in general. Thus a broad, generic, and ambiguous definition of religion is unavoidable. We, as James’s interpreters, are forced to search elsewhere in pursuit of a more precise understanding of what James’s religious hypothesis entails.

The significance which James attributes to the “faith in a fact can help create the fact” aspect of his argument leads to misunderstanding.\textsuperscript{20} While some of this is the fault of James’s interpreters, who fail to adequately assess the scope of his claim, a good deal of the misunderstanding arises from James’s ambiguous representation of the argument. James fails to distinguish between whether a fact is and whether it will be. For example, note the “subtle shift,” in his question, “Do you like me or not?” \textsuperscript{21} This is a question of fact; whether or not you currently possess a positive disposition towards him. James has not given us a means to an answer to that question. Instead, he recommends trust and faith that you will come to like him, and behavior that is according with this trust. This answer is appropriate for the question, “Could you like me or not?” “To believe that I can vault some Alpine crevasse is to believe not only in what “will be, but to some extent in what “is now that case.”\textsuperscript{22} If a bride asks her groom, “Do you love me?” she is surely not inquiring about a future possibility, but a present actuality. If she presses for evidence one may hope that this would have little effect on whether or not her spouse may answer in the affirmative. In this case, the faith in the future possibility has little or no effect on the present actuality. However, if one is only asking the question “Could you like me?” then faith in this possibility is almost certainly a necessary condition for the liking to come.\textsuperscript{23}

How broad is this category of facts which require belief for their verification? Perhaps in the alpinist’s case this argument is tenable, but does James truly hold that in weltanschaulich issues-- like the existence of God or the morality of the universe-- that our willingness to believe inspires actions which bring about God’s existence?\textsuperscript{24} It seems clear that James does, and does so rather explicitly. He says,

\begin{itemize}
\item \textsuperscript{19}James, “Will,” 30.
\item \textsuperscript{20}James, “Will,” 29.
\item \textsuperscript{21}O’Connell, 14.
\item \textsuperscript{22}O’Connell, 76.
\item \textsuperscript{23}O’Connell, 14.
\item \textsuperscript{24}It should be noted here that James used both the plural and the singular when speaking about God. He was quite clear that monotheism was by no means forced upon the religious believer, and that a plurality of finite gods was certainly on the table for discussion.
\end{itemize}
I confess that I do not see why the very existence of an invisible world may not in part depend on the personal response which any one of us may make to the religious appeal. God himself, in short, may draw vital strength and increase of very being from our fidelity.25

This claim is rightly seen to be incompatible with the traditional Christian concept of God. James’s religious understanding, vague as it was, cannot be understood in Christian terms. James had an aversion for most every type of institutionalized religious expression that he ran across. He dismissed ecclesiastic organization on the premise that such an organization is based solely on the authority of the personal religious experiences of its founders.26 For him religion must be a firsthand, primordial thing. The most fundamental views of religion must be those derived from experience of the subconscious self. If one sheds all over-beliefs, all that remains is “the fact that the conscious person is continuous with a wider self through which saving experiences come.”27 The theological tenets which Christianity affirms, such as the unity of the Godhead and the incarnation of Christ, do not fit within the parameters of the WB doctrine.

O’Connell’s Distinction

In an attempt to make more sense of James’s claim, O’Connell argues that James would eventually come to distinguish between “outcome cases” and “over-beliefs.” James makes this distinction clear when he wrote to H.M. Kallen in 1907, eleven years after the WB lecture.

It is usually poor policy to believe what isn’t verified; but sometimes the belief produces verification—as when it produces activity creative of the fact believed; and again, it may without altering given facts, be a belief in an altered meaning or value for them.28

---


26Hauerwas, 44.


O’Connell gives the former instances the title “outcome cases” and claims that these are the circumstances of the alpinist. However, in the latter case of “over-beliefs,” such as the weltanschaulich issues upon which James so often dwelt, faith in the fact does not create the fact. Instead, faith puts the believer in a perspective from which the evidence better corroborates the fact believed in. O’Connell uses this distinction to show that James has a tendency to “treat weltanschaulich issues in outcome terms.” He argues that the most conspicuous instance of this is James’s M + x argument in “The Sentiment of Rationality.” However, even in this circumstance, O’Connell argues, James thinks better of the optimism of his claim by the end of the lecture. For O’Connell, this is the profound contribution of James: “that only the thinker of developed moral character can be expected to “see” our universe in appropriate moral terms.” This moral character exists, of course, only because one has already accepted a belief in a moral reality.

O’Connell’s distinction provides insight into how James’s argument may be understood in a more cogent manner. However, his attempt to show that James doubted the extension of the argument seems rather thin. Schlecht’s reading is more accurate when he argues that James clearly held that the “faith can create the fact” argument did apply to weltanschaulich issues, and that this is only one of several instances where James puts forth the same argument. If Schlecht is correct that James repeatedly asserts that “our “over-belief” can itself contribute to the “outcome” desired, he must show that James’s concept of religion is compatible with such a claim.

James according to Schlecht

Schlecht focuses upon the moral and pantheistic world order which was typical of James’s later religious sentiment. He cites texts in which James repudiated theism as support for the claim that “what James identifies as central to religion is the affirmation that the universe as a whole is melioristic.”

---

O’Connell, 77. Briefly put, the M + X argument, found in “The Sentiment of Rationality” pg. 81, is this: M = the universe “minus the reaction of the thinker,” and x = “the thinker’s reactions and its results.” M +x = all that can be the subject of philosophy, or all that is. James is considering the issue of whether or not the universe has moral worth, a weltanschaulich issue. X can either be optimistic or pessimistic, that is, the individual may react in a brave and serious moral manner, or react negatively and give way to the evil in the universe. The ultimate outcome of the M+x situation is determined by the stance which we take.

O’Connell, 3.


Ibid., 345.
To be sure, meliorism is a central aspect of James’s religion, but it is not the sole tenet of James’s religious hypothesis. There are certain ideas which repeatedly appear in James’s discussions of religion which indicate that, at least during his penning of the WB lecture, there was more to the religious hypothesis than just meliorism.\footnote{Meliorism is the optimistic belief that the universe is indeterminate in regards to moral worth. The active part played by free human moral decisions in the universe is what incrementally makes it good or bad. “M has its character indeterminate, susceptible of forming part of a thorough-going pessimism on the one hand, or of a meliorism, a moral (as distinguished from a sensual) optimism on the other.” James, “Sentiment,” 84.}

Before attempting to tackle the question of what James religious belief concerned, the vague and noncommittal nature of his presentation of religion must be noted. James was reluctant to give an emphatic, clear, or final word on what he takes the whole of the meaning of religion to be. This is partly due to his popular lecture style of delivery, which favored eloquence and metaphor rather than precision.\footnote{O’Connell, 107.} Also, in view of James’s radical empiricism, no systematic presentation of God or the universe was ever truly final or complete. Not only did James’s arguments compel him to favor a broad and generic presentation of religion, his understanding of what religious belief entailed developed over the course of his career.

The James who wrote “The Sentiment of Rationality” in 1879 differs, from the James who wrote “Faith and the Right to Believe” more than forty years later.\footnote{Chronology taken from the introduction to William James: Essays in Pragmatism, Ed. Alburey Castell. (New York: Hafner Press 1948), p. xv.} A rough characterization of James’s religious belief shows him to move from a form of finite theism to a pluralistic pantheism later in his career.\footnote{Nicholas Lash, Easter in Ordinary: Reflections on Human Experience and the Knowledge of God (Charlottesville: Univ. Press of Virginia, 1988), 78-79.} The texts which Schlecht cites to substantiate his portrayal of James’s religion draw from the four major articles in which this particular argument was championed, “and in many other writings during the last decade of so of his life.”\footnote{Schlecht, 345.}
However, James’s argument that our faith can help bring about its object appears throughout the breadth of his work, so in assessing its coherence with his concept of religion, it will not do to rely upon his later views alone as the sole indication of James’s religious hypothesis.

James’s religion, according to Schlecht, is one which broadly contrasts to “materialism’ and its conviction that science provides us with the last word on the nature of reality.” He interprets James’s statement: “A man’s religious faith...means for me essentially his faith in the existence of an unseen order of some kind.” as merely the assertion that religion affirms the existence of a spiritual or moral order to the universe in which we are a part. The universe is a place where our desire for “harmony” and “meaning” can be fulfilled. Schlecht emphasizes James’s pluralistic pantheism, and the inter-relational aspect of reality. The universe is open, still able to be impacted, for good or ill, by our actions in it. This brings about the demands for the strenuous moral mood which recognizes “that our choices and efforts can make a difference.” “Proceeding with the kind of openness and good-will that would be necessary to develop and sustain a meaningful relationship with another person can contribute to making the universe one in which 'life is worth living.’”

Schlecht employs James’s analogy of relationships to illustrate the ability of faith to contribute to the reality of meliorism and the centrality of that belief to James’s religious hypothesis. James sought an I-Thou intimacy with the deepest aspects of reality, and in his mind that was the craving of all mankind. “Just as receptivity, trust, and good will are essential for an intimate relationship with another person to be established, the same kind of faith is needed for the world to be experienced as a place in which intimacy can be found and melioristic activity can be realized.” Meliorism, not theism, according to Schlecht, is at the heart of James’s WB doctrine.

James’s theism as the basis for Meliorism

Schlecht’s representation does not salvage the WB argument because it fails to address the finite theism which was prevalent in James’s earlier writing. Neither does his focus on the interpersonal aspect of the human relationship with the universe allow James to elude the criticism of O’Connell. The chronological development of James’s religious thought and the ambiguous presentations of the “faith creates the facts” argument show that O’Connell’s distinction between “outcome cases” and “over-beliefs” remains essential to understanding James’s WB doctrine.


40Ibid., 346.

Apparent in some of James’s earlier works, as well as in the *Varieties of Religious Experience*, is a finite theism, wherein a God or gods form the basis for the moral meaning of the universe. Schlecht dismissed the essays, “Reflex Action and Theism” and “The Moral Philosopher and the Moral Life,” both of which affirm some variety of theistic ideas, on the supposition that James would later explicitly repudiate theism. Due to the appearance of the “faith creates the facts” arguments both before and after the publication of these articles, this is not an acceptable dismissal. If James saw fit to publish these two essays alongside the WB, and further declare that they “shed explanatory light upon each other, and taken together express a tolerably definite philosophic attitude,” it cannot be accepted that he had already abandoned these views in order to adopt ones compatible with his argument.\(^42\) He clearly thought that these ideas were, to some degree, consistent. Schlecht does not salvage James’s argument by limiting it to the faith that there is an unseen moral order of reality. For James, this unseen moral order was only possible on the premise that certain things can be judged to be good or ill by some “divine thinker.” James substantiates this understanding of morality in “The Moral Philosopher and the Moral Life.” “Goodness, badness, and obligation must be realized somewhere in order really to exist.”\(^43\) Objects, in as far as they exist independently of the sentient mind, are neutral, neither good nor bad. They only become good or bad in relation to being perceived, or more appropriately, felt to be so. Therefore, the moral universe for which the philosopher strives is only possible on the supposition of a

  
  divine thinker with all-enveloping demands. If such a thinker existed, his way of
  subordinating the demands to one another would be the finally valid casuistic scale;
  his claims would be the most appealing, his ideal universe would be the most
  inclusive realizable whole.\(^44\)

If we, as humans are interested in a “systematically unified moral truth”\(^45\) we must believe in this divine thinker as real, and act as if it were so. Our limited efforts to realize morality are, in fact, our perpetual attempts to find and actualize this prior moral order.

In “Reflex Action and Theism” James asserts that the most essential features of theism are that God be “conceived as the deepest power in the universe; and, second, he must be conceived under the form of a mental personality.”\(^46\) God’s personality should be considered just as another personality, “something lying outside of my own and other than

\(^{42}\)James, *The Will to Believe and other Essays in Popular Philosophy*, preface, pg. i.


\(^{44}\)Ibid., 161.

\(^{45}\)Ibid.

me, and whose existence I simply come upon and find.” 47 At this point at least, James is not thinking of some internal essence of which we are all a part. We are distinct. God, whatever he may be, is other and prior.

Even in James’s later works, ‘other-ness’ and ‘priority’ remain distinct features of the object of religious experience. In giving his own “over-belief” he suggests that “The farther limits of our being plunge, it seems to me, into an altogether other dimension of existence from the sensible and merely ‘understandable’ world.” “So far as our ideal impulses originate in this region..., we belong to it in a more intimate sense than that in which we belong to the visible world, for we belong in the most intimate sense wherever our ideals belong.”48 Even as late as 1902, when James was writing his conclusions to the Varieties, he was still emphasizing the otherness and priority of the object of religion.

“Otherness” and “priority” seem to offer a problem for Schlecht’s interpretation. For not only is the object of religion “other,” it is a “reality.” “When we commune with it, work is actually done upon our finite personality.” “That which produces effects within another reality must be termed a reality itself.”49 But how can our actions in reality A bring about the existence of reality B when the ideals which drive our actions are said to originate in reality B? An alternate reality, consciousness, dimension, or whatever we may call it; if it exists as the basis for ideas or concepts within the world A, has to exist prior to and regardless of any human knowledge thereof. Reality B would have to be eternal, and exist before and independently from human consciousness. It could never be a fact which was “created” by any amount of belief inspired behavior in world A.

If we consider a world which is devoid of any consciousness as being amoral, how exactly does it go about being made moral? James seems to leave us with a chicken or the egg dilemma. Which existed first?

A1- the God or gods whose demands are the basis for human morality

B2 - the human moral behavior which is creates A1

James holds that both of these things derive being from one another. This might be possible if both have always existed. But James never supposed that humanity has always existed.

The theme of interpersonal relationships is vital to Schlecht’s argument. He uses James’s statement: “The universe is no longer a mere It to us, but a Thou, if we are religious; and any relation that might be possible from person to person might be possible here” to argue that in the WB we find an assertion which is “very much akin to the religious mysticism developed in later works.”50 How is one to understand James on this issue? The decision to

47Ibid., 98.

48James, Varieties, 406.

49Ibid.

50Schlecht, “Mysticism,” 260.
believe in the religious hypothesis is the decision to believe that the universe is good, or at least, that it can be made to be good, according to Schlecht. The decision to believe that someone else can like me is supposedly analogous. However, my decision to believe that someone else likes me presupposes that they exist. Likewise, the decision to believe that the universe is moral presupposes that a mind, outside our own, perceives it to be moral. What is more, this mind transcends our experience completely. So in order to suppose that the world is moral, and in order to suppose that you could like me, I must presuppose the prior existence of another, who exists regardless of my assent. Thus the religious hypothesis must be more than simply the belief that the universe is melioristic.

The importance of O'Connell's distinction

How does this understanding of a divine thinker reconcile with James’s words, published five years earlier, that “This world is good, we must say, since it is what we make it—and we shall make it good.”\(^{51}\) It cannot be that by adopting a strenuous moral mood, an openness and good-will, that humanity will bring about the existence a moral order. Perhaps James can be understood as arguing that if we adopt an over-belief that a divine thinker exists, and that his demands are the most binding, we might, by the resulting virtuous actions and dispositions, bring our ethical and moral convictions in line with his. In this way, the faith--that there is a standard which makes our moral actions meaningful--has not altered the fact of the moral order’s existence, it has altered our understanding of the empirical facts, and the outcome which results from the changes in our behavior towards them. Thus we are placed into a perspective from which the facts may, someday, corroborate the existence of that moral order.

Accepting O’Connell’s distinction may not be altogether consistent with James’s writing, but neither is it altogether alien. O’Connell admits that he is attempting to bring out what he sees to be the best in James’s argument.\(^{52}\) Consider that when James posits the reality of the I-Thou relation, he says that “We feel, too, as if the appeal of religion to us were made to our own active good-will, as if the evidence might be forever withheld from us unless we met the hypothesis half-way.”\(^{53}\) His claim here is not that the existence of the appealer might be dependent on our “meeting it half-way.” Instead our belief illuminates evidence which indicates that such an appealer does exist, thus it may be part of the verification of that reality. This does not entail that our behavior has brought about the existence of the appealer, but simply that we have, by our active good-will, contributed to that “final integration of things” in which it is verified that such a moral order exists.\(^{54}\)

\(^{51}\)James, “Sentiment,” 84.

\(^{52}\)O’Connell, 122.


\(^{54}\)James, “Sentiment,” 87.
“If a ‘melioristic’ universe were really here, it would require the active good-will of all of us, in the way of belief as well as of our other activities, to bring it to a prosperous issue.”\(^{55}\) James’s claim is not that our active good-will can bring about the melioristic universe, but that it might bring the already existing melioristic universe to prosperous issue.

Attempts to reconstruct James’s WB doctrine have merit, then, in as far as they bring to the fore the ambiguities in James’s argumentation and attempt to apply some clarity. O’Connell’s attempt has remained within a broad parameter of James’s thought and if it is corrected in regards to James’s theism, it still stands as a valuable contribution to the literature on William James. Schlecht’s attempt to show that James’s argument is tenable ‘as-is’ does not do justice to the early stages of James’s religious thought and ignores the basis for a melioristic universe. O’Connell’s analysis, slightly filtered through Schlecht’s correction, offers an insightful development and defense of James’s WB argument.

Matt Schneider – Self-Ownership, Freedom and Responsibility: An Argument for an Enforceable Obligation to Help

Abstract:

The principle of self-ownership claims that all persons have full and exclusive property rights to their own person. Libertarians such as Robert Nozick traditionally utilize the self-ownership thesis to argue against enforceable, political obligations to help. Their argument centers around two claims. First, they deny a connection between moral duties to help and political rights to be helped. Second, they claim that forcing persons to help others uses their bodies and/or labor against their will, constituting a violation of their rights of self-ownership. This paper rejects the libertarian argument against coercive assistance and offers a positive argument for enforceable, political obligations to help, based on the self-ownership thesis. I claim first that the distinction between duties to help and rights to be helped is not as clear or obvious as libertarians suggest. Second, I argue that in some cases, not helping constitutes a violation of persons’ political rights. When not helping leads to foreseeable harm, a person who withholds assistance can be blameworthy for the harm in question. This blame is what ultimately generates enforceable political obligations to help in such cases.

Self-Ownership and Obligations to Help

Within the liberal political tradition, the self-ownership thesis claims that all persons have robust, exclusive property rights in their own person. Libertarians commonly argue that forcing persons to assist others in need of help constitutes a violation of their self-ownership rights. This paper aims to defend against the criticism that when applied as a political thesis, self-ownership fails to account for common intuitions regarding mutual aid. I argue against the foregoing libertarian view, and claim that self-ownership can and does generate enforceable, political obligations to help others.

It is because of the exclusivity of persons’ property rights in their own person that libertarians argue explicitly and extensively against any and all enforceable, political obligations on persons to assist others. Their judgment holds even if assistance could be lent at little or no cost, and even if a person may rightly be viewed as morally despicable for refusing assistance. It is this view which I intend to undermine, while retaining what I take to be fundamental about the self-ownership thesis.

I have broken my project into two parts. First, I provide an argument for enforceable, political obligations on the basis of special relationships, such as between parent and child. However, this analysis fails to derail libertarian arguments against coercive assistance. My second argument rejects the logic used by libertarians in their position against coercive
assistance and offers a competing analysis of self-ownership, need and assistance. I believe this argument presents a compelling case for positive, political obligations to offer help.

Just Desserts

Libertarians often argue that if you work and gather food for winter while I sit around (able-bodied) and do nothing, you are under no obligation to help me survive the winter, even if you have a surplus. Likewise, if another person gathers more than she needs for survival, and can afford time to rest and do philosophy, she is owed that luxury, so long as her surplus did not come at the cost of another’s involuntary poverty. However, while I may bear some responsibility for failing to provide for myself when I am clearly able to do so, it is not clear that you or I are in all cases free from non-contractual or non-punitive obligations to others, as many libertarians suggest.

Though persons own their own body, our bodies and minds are variously equipped to provide for our own sustenance, safety, and general well-being. Consequently, we cannot hold children or the disabled blameworthy for failing to do what they cannot do. We do not expect children to be able to provide sufficiently for themselves. Nor would we expect an individual without legs to be able to walk around and gather fruit, carry heavy loads, etc. The case is similar for mentally handicapped persons. We cannot expect them to be able to always provide for themselves in a sufficient manner. Though such persons may garner our sympathies and possibly our charity, the political reality for them on a libertarian account amounts to ‘tough luck.’

The absence of non-contractual obligations to children (for example) on a libertarian account is troubling to many in the liberal tradition. It is not clear, however, that no such obligations exist. By knowingly and willingly creating children (as opposed to being forced to bear a child by way of rape) we may accept an implicit obligation to care for them, much as adopting a pet comes with the assumption that we must feed and care for it. The needs of a child are foreseeable, and as such, any parent who knowingly and willingly creates one may be held accountable for its care until the child can care for itself. In the same way, if I injure you, I may be held accountable for your care until you are capable of handling it yourself. However, if our account only obligates the child’s parents to care for it in virtue of knowingly and willingly creating the child, then we have not completely bridged the gap between a Nozickean interpretation of self-ownership and one that accounts for a broader range of liberal intuitions about children and society. For example, if a child’s parents are unable to care for the child as a result of economic or medical reasons, then on the libertarian account, the child is left to the mercy of benevolent family, friends or strangers. Proponents of a minimalist (Nozickean) interpretation of self-ownership might argue in favor of this result, and find it unobjectionable. On their account, someone in the society would care for such children. Childless individuals and couples have had and will likely continue to have, a desire for their own child. Social organizations, religious faiths, and other voluntary-membership institutions in society have and will continue to assist in the adoption of unwanted children. In short, the needs of children would be met, not through a mandate on society but through the voluntary actions of citizens. Unfortunately, this
'solution’ is not a solution at all. It simultaneously (though perhaps implicitly) acknowledges the rights of children to be cared for without demanding that society fill that need when necessary. An advocate of this position must accept that children whose needs are not met by their parents or private institutions in their society are tragic cases but not unjust ones.

The above analysis can only get us so far. To adequately address liberal concerns about care and political obligations to assist, we must face the libertarian argument against coercive assistance head-on. Nozick and other traditional libertarians argue that self-ownership prevents any citizen from being forced to help another. He claims that even if we may have a moral obligation to help, no one can force us to do so. This is a subtle but important distinction. It means that though I may have a moral duty to help you, you do not have a legitimate, political right to my help. This claim is popular in libertarian and neo-Lockean literature. The validity of the claim itself is somewhat unclear, however. Let us see how it might work.

Nozick bases his argument against any sort of coercive ‘helping’ of others on two vague principles. The first denies the connection between moral duties to help and rights to be helped. We can have the former (duties) without the latter (rights). The second principle claims that coercive assistance is unjust because it violates a person’s self-ownership by using their body and labor without a person’s consent. Even if the first principle is ultimately defeasible, the second attempts to explain why a ‘just’ state is nonetheless prohibited from enforcing persons’ moral duties to others.

To take an example from the pages of Peter Singer,1 suppose we find a child drowning in a puddle of water. We could save her life at little cost to ourselves. Singer argues that we have a moral obligation to do so. However, Nozick claims that a moral obligation to help is consistent with the claim that the child does not have a right to be helped. So long as no one is to blame for her situation, then to use Moller-Okin’s characterization of Nozick’s position, her death would be “unfortunate, but not unjust.”2 How could this be?

On Nozick’s interpretation of rights violations, the girl’s rights would have been violated only if someone actively did things that led to her injury or death. Absent these active violations of her rights, no passive or omitted moral duty to help constitutes a violation of her rights. This distinction is weak, at best. There is certainly debate as to how to best distinguish action from omission and whether or not they generate different accounts of blame and culpability. However, this does not relieve those who omit actions from any responsibility, though their responsibility may be less than persons who knowingly commit an action. This would allow that a person who cast the girl into the puddle is highly blameworthy for her death, whereas the person who refrained from helping her out might be somewhat less blameworthy though they are both to blame, to greater or lesser degrees.

---


As we have seen, Nozick is willing to admit that refusal to help may be morally blameworthy – he accepts a person could have a moral duty to help. However, on his account, the only way to make sense of this blame is to say that our unhelpful citizen is blameworthy only of not fulfilling his moral duty to help her, and not blameworthy of violating the girl’s state-protected rights by letting her die. This is walking a thin line, indeed. Such a claim cannot stand alone as a sufficiently strong argument against coercive assistance. In order to reinforce his position against coercive assistance, Nozick and other libertarians must appeal more directly to the self-ownership thesis.

Nozick’s second implicit principle claims that coercive assistance violates self-ownership. If you force me to help someone who needs my help, you have used me and my labor against my will. I alone have full and exclusive property rights in my person and labor. By forcing me to use my labor or other personal resource (such as, for example, Rapunzel’s hair), you have made me slave to both you and the person I end up helping. Thus, the self-ownership thesis appears to be at the core of what many find to be a repugnant conclusion: though we may have moral duties to help, justice does not hold any person accountable for refusing help, no matter how small the cost of giving assistance.

The following constitutes what I take to be the libertarian argument against helping (mentioned above) and one possible way of rejecting the argument. Most libertarians hold that (1) coercive assistance violates one’s rights of self-ownership and (2) moral duties to help do not generate enforceable political or legal duties to help. It is based on these two claims that we arrive at the traditional conclusion regarding the drowning girl: Though I may be morally obligated to help the drowning girl, it is unjust for you or anyone else (including the state) to force me to help her. This conclusion can be and is generalized to a conclusion against coercive assistance in all cases. Libertarians must reject any strong connection between moral and political duties (2) in order for their positions to be coherent. Otherwise, the door is left open for individuals to have a legal or political right to be helped, even while this right is unenforceable by the state (or anyone else), given libertarians’ claim that coercive assistance violates rights of self-ownership (1). The intuition leading to (2) is not a bad one, and we should not deny that at least in some cases, a moral duty to do $x$ does not entail a legal obligation to do $x$, nor does it generate a legitimate legal/political claim by others that we do $x$. For example, I may have a moral obligation to be kind to strangers (or children, or grandmothers, if strangers don’t garner your moral sympathies), but we do not want to say that strangers (or children or grandmothers) have a right to my kindness – at best they have only a right to be left alone. However, we should reject the claim that no moral obligations entail political ones; libertarians must reject this as well. Our political obligations or duties to others are reciprocal in nature; A’s political obligation to Z entails Z’s claim against A that she fulfill her obligation to Z. So, while libertarians must claim that (some) moral duties to help do not generate enforceable political or legal duties to help, they must also argue that coercive assistance violates rights of self-ownership in order for their position against coercive assistance to work. My criticism of their position focuses on weakening both claims.

Singer’s own argument for moral obligations to assist the needy includes the caveat that our obligation to help depends on not having to give up something of equal or greater
significance. What Singer has in mind, of course, is that we would not be obligated to save the drowning girl if we were on our way to deliver the security code which would stop the clock on an atomic bomb underneath Times Square, or if the water in which the girl was drowning was equally threatening to our own safety. In our case, however, saving the girl would not require us to give up anything equal or of greater substantive significance. Instead, we would be asked to suspend our desire to keep on walking, or if you will, the macabre desire to watch as the girl drowns in front of us. Libertarians hold that we have the right to do both things, even if we are morally despicable for doing so. I think that within the scope of Singer’s argument though, this judgment stands. On the view I am challenging, though we may have a moral obligation to help (for which Singer successfully argues) we may not have an enforceable obligation to help – our moral obligation is not a political one. This is precisely the conclusion of traditional libertarians. As such, analyzing the problem in terms of Singer’s original argument will not generate a sufficiently strong case for having a political obligation to help. However, I intend to use the spirit of Singer’s original argument to develop the case for such an obligation. Below is such an argument.

Suppose Robert is faced with Singer’s case of the drowning girl. He finds himself in a position to help her, at little cost to himself. Though he foresees that not helping will result in her drowning, he does not help. Predictably, the girl drowns.

I argue that insofar as Robert could have prevented harm befalling the girl, he is, at least in part, responsible for the girl’s being harmed. ‘Responsibility’ here is used generally, in the sense that a free and rational person is ‘responsible’ for their actions. Whether this responsibility becomes moral or not depends on the relevant context. However, I take it that insofar as we are talking about knowingly and willingly harming someone or letting them be harmed, that the responsibility in question is moral in nature. This is to say that insofar as Robert is responsible for harm, then he is, ceteris paribus, blameworthy for the girl being harmed. Now, in a political context grounded in self-ownership, I take it to be unproblematic to claim that if a person is to blame for another’s harm, then that person has violated the others’ rights; after all, we have a right not to be harmed by others in such a state. So, by not helping the girl, Robert has violated her rights. Rights violations are prohibited and the state can and ought to enforce citizens to respect each others' rights. In this case, since Robert would have violated the girl’s rights by not helping her, he has an enforceable, political obligation to help her.

I take the above argument and its premises to be, commonsense, analytic or at the very least, unobjectionable to persons operating within the liberal tradition. However, a bit more might be said regarding some of my claims. It may seem as if I am playing fast and loose with the highly scrutinized concepts of blame and responsibility. While I acknowledge the debates over borderline cases of blameworthiness and responsibility, I do not think my argument trades on any ambiguities; it is designed to elicit common intuitions about blame and responsibility. However, to indulge opponents of coercive assistance, my

---

3 The ceteris paribus clause accounts for cases where blameworthiness for harm is overridden by circumstantial factors, such as in the case of x harming y in self-defense against y (or even if y is an innocent aggressor).
'sleight of hand' so to speak, enters into the argument in two places. The first deals with the relationship between preventing harm and responsibility and the second with the connection between responsibility for harm and blameworthiness for harm. For the sake of clarity, I shall spell out these two concepts:

1. **Principle of responsibility**: For any x and y, if x could have prevented harm befalling y and fails to do so, then *ceteris paribus*, x bears at least some responsibility for y's harm.

2. **Principle of Blame**: For any x and y, if x is responsible for y's being harmed, then *ceteris paribus*, x is to blame for y's being harmed.

Nozick and other libertarians will likely object to one or both of the consequents of the above 'principles.' This is (again) because they draw a strong distinction between an act of commission and an act of omission. Opponents of my argument may claim that Robert is not responsible for harm that befalls the girl because he did not do anything which directly led to the girl’s being harmed. He did not push her into the puddle, hold her down, place the puddle where he knew her to walk, etc. On such a view, Robert’s helping the girl would be strictly *supererogatory*. However, this denies that ‘responsibility’ can have any reasonable interpretation other than direct, strong causation.

Under this view of causation, x causes y if and only if there is a physical connection between x and y such that y is explained in virtue of x’s taking place. For example:

‘My swinging of the bat in such and such a manner is the direct cause of the ball being hit over the fence.’

Libertarians may weaken this notion of causation to some extent (i.e. not require a physical connection, per se), but the relevant claim remains: x counts as a cause of y if and only if x is an intentional *action* which directly leads to y. On a libertarian account, Robert does not cause the girl’s drowning; the water in her lungs causes that. Nor would the libertarian want to say that anything Robert *did* led to the girl’s drowning as a result of being in the puddle, taking in a lungful of air, etc. However, there is another sense of causation which I am utilizing in this case. On this account, x causes y if x is the most salient factor in an explanation linking events x and y. For example:

‘The manager’s absence at the fourth straight staff meeting caused a mutiny among his staff.’

It is this sense of causation at play in the *principle of responsibility*. The girl’s drowning is best explained by Robert’s *not* pulling her out of the water. ‘Responsibility’ substitutes for ‘causation’ as we are speaking not of arbitrary events but of persons with wills. Robert’s causing the girl’s harm is here a case of Robert’s being responsible for her harm, as he reasonably knew that not helping would have likely resulted in her harm and freely chose not to help.
It may be objected that the *principle of blame* makes a substantial claim in need of support. I do not deny that a more general claim along the same lines such as ‘R’ (below) is in need of further support:

R: For any x and y, if x is responsible for y, x is blameworthy for y

‘R’ is, I think, false on any account. However, the *principle of blame* specifically relates to *responsibility for harm*. Though blame is a distinctly moral concept, responsibility is only sometimes moral in nature. However, I take it to be uncontroversial that when speaking of harm to someone or something to which we owe some moral responsibility, that being responsible for harming that thing or person is moral in nature.

**Conclusion**

What this argument demonstrates is that the distinction between *moral duties* and *enforceable, political obligations* many libertarians make is not a sharp one. It is true that not all moral obligations are enforceable by the state. Moral obligations take a diversity of forms and various consequences await their fulfillment or lack thereof. What my argument demonstrates is that *foreseeable harm* arising from a *failure to help* can be viewed as a rights violation, and a limited[^4] duty to help can therefore be justly enforced by the state without rejecting, or violating the self-ownership thesis. Moreover, we need not take as our starting point a pre-political, *moral* obligation to help. The obligation to help can be generated as a purely political one on the self-ownership thesis with an exactly similar argument to the one detailed above. Once we see that knowing omissions of assistance can stand in a causal relationship to harm suffered by persons in need of assistance, generating a legitimate, *political* claim to assistance is not difficult, nor does it involve any violation of individuals’ self-ownership.

My claim is not an overly ambitious one; the conclusion I argue for, that a just state mandate help be given to persons in danger of harm, is not radical. What I take to be important is that I have reached these conclusions from within the scope of the self-ownership thesis.

[^4]: The mere existence of a causal relationship between a person and an event cannot be sufficient to generate moral responsibility or blame. We are all minimally ‘responsible’ for a great many harms that take place in the world. However, no one person can *practically* be held accountable for their small role, whether knowing or not, in bringing about harm to other persons. The example of the drowning girl is meant to stand as a paradigm case of blameworthiness for failure to help. Other cases will be much less compelling. My failure to donate my life savings to domestic aid organizations does not count as a violation of needy persons’ rights, though my failure to contribute equitably to this cause through a redistributive taxation system might. Though conceptually related, there is a distinction between *individual* responsibility for action or omission, and *social* responsibility. Society as a whole may have duties that no individual could reasonably have. Rather, it is an individual’s responsibility to contribute his or her share to their society’s duty.
ownership thesis, a concept utilized by libertarians to argue vehemently against coercive assistance. In so doing, I hope to have laid the groundwork for a more thoroughgoing defense of self-ownership within the liberal tradition. Furthermore, the conclusions I have drawn here do not carry over to coercive assistance outside the scope of foreseeable harm. This particular argument argues merely that the self-ownership thesis justifies coercive assistance in cases where we are avoiding harms, not seeking benefits. It is a different matter to show how failure to benefit someone counts as a rights violation. However, failure to help avoid foreseeable harm is a more than reasonable candidate.

---

5 For example, my neighbors may need help with their yard waste, and the state could provide a tax-funded leaf and limb pick-up service. However, nothing about the above argument rejects the libertarian position that forcing taxpayers to fund such a service (regardless of whether or not they themselves benefit in some way) is a violation of property rights derivative of self-ownership.
Scott Syzmanski – Justice: Glaucon’s View and Possible Objections

The question of justice is something that has been debated for centuries. It is a topic which begs to be understood, since its presence is at the forefront of most of our decisions. We are guided by some form of the concept of justice throughout our lives, and it is almost inseparable from human interaction. Within any society, justice exists and manages how people relate to one another, how they conduct their business and go about their lives. It is only natural to want to question this guiding principle of fairness and benefit. In this paper, I will discuss one particular theory of justice that was brought about through Plato’s Republic via a character named Glaucon. I will explain the reasons for his views, its constituents, and how this idea is applicable to the discourse of justice. His view will be supplemented by his brother Adeimantus’ opinion on the matter. In addition, I will attempt to answer the question “Is justice good in itself or only for its consequences?” Ultimately, although I do agree with some of Glaucon’s assertions, I will explain why justice can be defended in terms of his argument and why it is a better way to live despite its possible infringement on indulgence in this life.

In order to spark a discussion with Socrates, Glaucon represents Devil’s Advocate when he raises the idea that justice is a social construct, something which is only created out of its necessity in a functioning social environment. According to him, it is natural to be unjust and live for oneself and by one’s own means. The goal of a person is to get as far as he or she can in his or her own life and gain as much as possible along the way by any means necessary. Glaucon explains this thought by way of suffering. In his words: “to do injustice is naturally good and to suffer injustice [is] bad” (Plato 1992, 359e). By this quote, it is clear for one to commit injustice for his or her personal gain is warranted, since one’s fundamental goal in life is to achieve the best life possible. It follows that for one to suffer injustice—that is, for one to be taken advantage of, or stripped of property without consent, etc.—is something which does not coincide with the best life for everyone. Because individuals of the general masses—those who lack impunity and are weaker than the strong—saw that in the ways of injustice one benefits wrongfully from another while the other suffers unnecessarily, laws and covenants were created to prevent either case from occurring. For society to prosper and function efficiently and respectfully, each citizen must follow the laws that have been enacted. Glaucon asserts these laws were only made via the context of fear (359a); that is, those who vouch for justice and fairness and equality among all individuals within a society only do so out of the fear of suffering injustice. This is not to say these same people condemn those being unjust: only that they oppose themselves having to suffer. One would not like injustice done unto him or her, since this would cause suffering and pain. And one also would despise injustice if the unjust person committing it did not suffer punishment, which is something unavoidable by the masses. Justice, as it seems, is engaged because of its consequences, namely that individuals do not have to suffer nor do those who execute injustice be relieved of punishment. According to Glaucon, the just only support justice because they cannot commit injustice with impunity. Therefore, if one were to have impunity, he or she would partake in injustice, illustrating
the just only believe what they do unwillingly, as justice is a constraint on each person’s own desires through fear.

Although justice is something which is created through social interaction, this does not mandate all must follow its rules and affections. One group—the authority—is exempt from the stipulations of the contract. Ruling with impunity, the authority may commit injustice without consequence. Glaucopon discusses this idea with the example of the Ring of Gyges (359d). As the story goes, a shepherd in the service of the ruler of Lydia witnesses a great storm and subsequent earthquake in the pasture on which he tends his sheep. The cavern created by the earthquake reveals a hollow bronze horse in which a corpse with only a ring on its finger resides. The shepherd takes the ring, and later discovers that by turning it on his finger he can become invisible. With this power, he seduces the queen of Lydia, kills the king, and takes over the kingdom. The question arising from this tale is ‘would anyone who has this impunity—that is, freedom from punishment—commit injustice necessarily?' Glaucopon asserts if one did have this god-like power, it would be imperative that one act unjustly, since there would be no reason to continue to act through justice. It is not possible for one to be so incorruptible that he or she remains steadfast with justice even though there is no punishment present for any unjust actions. In Glaucopon’s words: “...someone who didn’t want to do injustice, given this sort of opportunity...would be thought wretched and stupid by everyone aware of the situation” (360d). This further affirms justice is only followed when fear of suffering and punishment are present.

Glaucopon insists the unjust life is more worth living than the just. He claims that because justice prevents one from achieving one’s goals as efficiently as possible, and instead restrains each person by a set of rules and guidelines, it creates a lifeless fulfilling and less beneficial to an individual than one that is unjust. In order to explain this concept in detail, a comparison between a just and unjust person can be made (360e). A completely unjust person commits whatever is advantageous and helpful to him or her regardless of its implications on the rest of society. This person has the outward appearance to society as being just, since this would aid in one’s betterment. He or she would have the ability to be unjust without detection, since being discovered would only be detrimental to one’s well-being and reputation. It can also be seen—if reputations are considered—that an unjust person could have one that is just, since injustice would suffice for such a perception by society. If one is unjust though appearing otherwise, the person receives the graces of both sides: the person achieves all he or she wants through taking advantage of others and obtaining whatever he or she desires in opposition to fairness, and he or she is respected by other individuals and regarded in good character.

Taking the opposite stance, a completely just person would value personal identity as higher than merely appearing as such. A just one would want to be wholly just, not only seem to be, through action and respect. This person would be cognizant of others’ desires and would act accordingly through his or her own actions so to not take advantage of nor harm them. As with the unjust person, the reputation can be taken away to reveal the implications of justice itself, since the reputation of one who is just would lend him or her honor and bounty by fellow individuals, something which Glaucopon wanted to disregard in defense of justice. Once stripped of this, the just person does not have this credibility for
support, and he or she can be speculated by society as being opposite, namely, unjust. Accepting this, the just one would suffer throughout life, as society looks down upon one who appears to be unjust, since this side of a person is public and all that can be perceived by others. This is especially important to acknowledge because Glaucon desired a juxtaposition of the two extremes of life as just or unjust. Taking the characteristics of an unjust and just person, it seems almost apparent that in order to fulfill a meaningful life in which one obtains all one desires and lives happily, one must choose to be unjust while appearing to be just, rather than the opposite.

Alongside the debate of a just and unjust person, the gods must be considered, for many may say that one can achieve a great life on this earth through dishonest and unjust means, but will be punished for it in the afterlife. In the times of Glaucon, the gods were understood through the poets Homer and Hesiod as being more loving to those who offered greater prayer and sacrifice. In other words, the gods would favor one who showed more appreciation and respect to them. Glaucon’s brother, Adeimantus, proposes the notion that a rich man who obtained his wealth through injustice could reconcile the gods to his side by ritual and enchantment (364c). The gods, as it were, are influenced by repentance and sacrifice, leaving the ability of one who is unjust to commit his or her acts willfully and receive the best possible life here and in the afterlife. Glaucon, using this example, purports that the only one who suffers in the afterlife is the one who does not perform accepted rituals. To compare this conclusion to that of a just person, he or she might be just in this life, but is only so to prevent punishment in the afterlife. The just person might acknowledge that being just will earn respect from the gods, but he or she will only be doing so in defense. Thus, it can be seen that to be unjust yields the spoils of both this world and the afterlife.

It has been shown by Glaucon that injustice is more beneficial than justice for an individual in this life as well as the afterlife. One who is unjust can gain all spoils and profits from his or her actions as well as the enrichment of a positive afterlife by sacrifice and prayer. To reference a modern saying: an unjust person can have his or her cake and eat it, too. The arguments proposed by Glaucon are sound and illustrate the excellent point mentioned above effectively. A question that is important to the discussion of justice is if it is so great and is prized by all and beneficial to society as a whole, how can it not stand on its own, that is, stand without the support of its consequences? Unfortunately, Glaucon is right that justice has no real purpose except for the outcome its implementation within society brings. If each individual were just, fairness and respect would be the norm, and care for others’ wants and needs would be common in the minds of all of the citizens. In fact, this shared bond between the members of this society would permit scrutiny when it comes to those who go against it. The members would collectively condemn the actions of an unjust person, citing it does not benefit others as it causes suffering with the advantage only to one party in the altercation. The point of justice is equality through a tacit pact between all members of a city, and these members would uphold this agreement so each is allowed to prosper without being at the expense of others. The consequences of justice, if anything, are more important to it than justice itself. One can be mindful of others and act in the right manner all the time, but this is neither as important nor advantageous as what becomes of such actions.
Without the consequences of justice, it is obvious how unsatisfactory justice really is. If one were to act just for the sole purpose of doing so and not for what it could bring, he or she might act out of altruism for almost any circumstance, especially when denying altruism could hinder another person. The person might even go so far as to endanger his or her life so to not harm another. This, of course, would be exercised out of fear of punishment for performing the opposite. Unfortunately, a person might also harm someone by continuing to be just, as would be the case if he or she refused to commit a crime to obtain medicine for a terminally ill significant other. The consequences of justice are so intertwined that justice itself is hardly defensible without acknowledging them. For justice to work effectively, it must pertain to the outcome of an act, not just its doing. Giving money to charity in itself seems illogical because one is giving money away to others who did not earn it by working or others one may not even know. However, if the outcome of people being saved with shelter, clothing, and food is considered, it makes the action worthwhile and understandable. Likewise, giving a lost wallet to its owner in itself seems wrong since the implication of “lost” means it is, so to speak, “up for grabs”, allowing the finder to keep whatever is in it. But looking at the consequence of the owner being very thankful and the just person increasing his or her well-being, as well as realizing he or she would treat others as he or she is treated, it is apparent the action has purpose beyond its initial performance. Justice almost completely requires its outcomes for it to be a reasonable idea to follow.

Glauc occurs a reasonable challenge against justice that it is more profitable and worthy to live a life of injustice. This, of course, accepts the fact that an unjust person—being allowed all the benefits of such—would be presented to others as just and trustworthy. Although the argument Glauc occurs makes is reasonable, namely, that being perceived just is more important than being just, it is not foolproof. By only being perceived just when in fact being unjust lends itself to a sophistic view of the world: that is, what is most important is how one appears, not how one is. This, I believe, is somewhat shallow and careless, as it accepts demeanor as superior to true character. Under such terms, a person can be a conniving womanizer yet dress and speak like a respectful gentleman or a person can be an advocate against illegal drugs but use them every day in private. The sophistic view leaves too many loopholes for hypocrisy and untrustworthiness, not to mention blatant lies. If true character is not important, how can anyone know others really are their friends instead of just appearing to be so? Acting unjustly may also be detrimental to a person in that one may miss an opportunity to live a more fulfilling life simply because he or she decided to take advantage of another. Maybe a person could have gained more wealth if he or she was courteous and remained friends with a business tycoon instead of swindling money from him or her. A person could have gained knowledge at a university if he or she didn’t cheat on an entrance exam. This does not seem to be considered by Glauc occur. Living the life of an unjust person with a façade of justice may be beneficial, but it is not without its faults.

Justice, accompanied by its consequences, is something that can be held as satisfactory by all individuals. Though injustice can be helpful to one person, justice supplements all people, allowing each to prosper and be guided through life by a principle of moderation. Socrates discusses the idea that it would be easier to see a large surface that contained a set
of letters first before affixing eyes on a smaller, more strained version of the larger (368d).
This is analogous to the comparison of a society to an individual. It can in fact be possible to
notice an individual before a society, but is it reasonable to assume an individual
completely represents and stands for the beliefs of the society as a whole from which he or
she came? Surely if all the citizens are lumped together, the justice of the group as a
composite will be greater than that of one person. Justice works in this way: it is more
beneficial to the general population, to all members of society than to the individual.
Equality and moderation is what is crucial to justice. The Golden Rule can be used here,
namely that by each person respecting everyone else’s goals and possessions, no one loses
what is rightfully his or hers, and no one gains by hurting another. It is, in fact, as part of a
whole, to the benefit of the individual that the general population follows justice. If one
were to live on his or her own, injustice might be reasonable, if not necessary, since
survival is most crucial and top priority. However, if one operates within society, being just
is beneficial to all, relieving anyone from suffering unnecessarily at the advantage of
another.

Besides society being the catalyst for justice’s service, the gods value justice as more
important and favorable than injustice. Socrates considers this idea by accepting that the
gods are perfect and all powerful. If a person were godlike, there seems to be no reason
sacrifices or prayers from mortals would have a strong impact on the gods. These prayers
and sacrifices are inherently less perfect, and the gods cannot become any better than they
already are. The gods would remain on the side of the just because these people attempt to
create the best possible life for everyone, permitting anyone any achievement and benefit
as long as he or she does it honestly and kindly. Glacon’s main argument regarding a just
person being viewed as unjust can also be refuted. The just who are viewed as unjust still
do live a decent life—they know in their hearts they are just, and for this they can be happy
and content—since they act within the realm of justice. Although others do not see this
side, they can rest assured there is no better way they could live (better, meaning more
honest and humble). Even so, the just might suffer because society would look down upon
those who appear unjust. But if they believe they are doing the right thing, and what they
do benefits society as a whole, there should be no question living this way is best. I believe
justice can at least stand toe-to-toe with Glacon’s argument for injustice.

Glacon’s theory that injustice is more beneficial than justice bases itself on the individual
without considering consequence. He focuses most of his argument on comparing justice
and injustice independent of their outcomes. However, since all actions have consequences,
it seems unfair to judge them solely by themselves. Repercussions happen from any action,
and whether it be just or unjust, the effect will always follow. It cannot be fathomed
entirely that the action itself is more important than the outcome of such a thing. Being just
is beneficial to all, not just to one person. It is not fair if one were to gain upon another’s
loss, hence injustice is inherently not fair. To live a life full of indulgence and greed,
injustice is a viable option and one which almost guarantees every desire to be fulfilled.
However, when individuals are viewed as a part of a larger system rather than independent
pieces, it becomes clear that for a society to prosper and flourish, justice must monitor the
actions of its citizens in order to assure all are benefited in life.
Michael Uhall – A Principle Objection: H. P. Grice and the Failure of Cooperation

Introduction

In the pages that follow, I will contend that Grice’s conversational maxims do not and cannot, in fact, apply meaningfully to discourse. First, let me briefly summarize Grice’s account of conversational implicature, paying particularly close attention to what are called Grice’s maxims. I wish to critically examine these maxims in detail because they, along with a certain principle that Grice asserts, provide the foundation on which he constructs a wider account.

Furthermore, Grice claims that his maxims are “general features of discourse;” in other words, they regulate meaningful talk exchanges. Without them, Grice implies, a talk exchange would consist of nothing more than an abnormal succession of disconnected remarks. As such, Grice in fact provides what is in essence an ethical account.¹ In response, I will attempt to show how his maxims do not, in fact, apply meaningfully to discourse. And, finally, I will conclude by providing my principle objection to his account, which also functions as a counterexample to the primary assumptions upon which he proceeds.

Grice’s Account

In “Logic and Conversation,” Grice argues that conversation can only occur on the basis of a certain set of maxims, and that when conversation does not proceed according to these maxims, it is only possible for successful communication to occur via what Grice calls conversational implicature, i.e., a codified failure to fulfill a maxim in a variety of ways. To clarify, the meaning of a conversational implicature may be distinct from what a speaker says, but it is implied by how a speaker says it, i.e., the context of that speaker’s failure to fulfill the maxims. But it is only if Grice’s maxims usually do govern conversation that we can determine what a speaker means when those maxims go unfulfilled. In other words, failing to observe Grice’s maxims implies something about the intent of a speaker beyond the literal meaning of what that speaker says.² According to Grice, the specifics of this failure provide a crooked route to what is behind what is said – that is to say, implied.

So Grice provides his maxims primarily to prepare for his discussion of conversational implicature, but he also describes these maxims as “the general conditions that, in one way

¹In this, Grice’s account resembles, e.g., Jürgen Habermas’s “discourse ethics” more than anything else.

²Here I make no commitment to the existence of “literal meaning” as a substantive postulate, but only in the limited sense necessary to contrast with “implied meaning.”
or another, apply to conversation... irrespective of its subject matter.”\textsuperscript{3} As such, his maxims – in addition to what he calls the Cooperative Principle – provide the groundwork for his work with implicatures. Furthermore, his entire account necessarily depends upon them, since without his list of maxims standing firm, there is no orderly system from which to generate analyzable implicatures, which are, after all, supposed to be meaningful deviations. As such, in the following pages, I will attempt to undermine the status of his maxims, thus casting doubt on his account in its entirety.

\textbf{Grice’s Maxims}

Since I have so far only referenced Grice’s maxims, let me now provide full summaries of each, as follows: (M1) under \textit{Quantity}, "Make your contribution as informative as is required (for the current purposes of the exchange)" and no more; (M2) under \textit{Quality}, “Try to make your contribution one that is true,” subdivided into proscriptions against lying and making claims without evidence to support them; (M3) under \textit{Relation}, “Be relevant;” and, finally, (M4) under \textit{Manner}, “Be perspicuous,” subdivided into “avoid obscurity,” “avoid ambiguity,” “be brief,” and “be orderly.”\textsuperscript{4} In short, the trouble with these maxims is, put simply, that they cannot be usefully applied to discourse. For example, let me examine M1 in detail, before more briefly examining M2, M3, and M4.

M1. This first maxim takes for granted that the “current purposes” of a talk exchange are both clear to the participants and intentionally made so by those participants. Neither point necessarily holds. Sometimes the purposes of a particular talk exchange are anything but clear to the participants. For example, any exchange in which the utterances of a participant are confusing or disorderly, such as in ambiguous or unclear speech, disruptive or unwelcome small talk, interruptions, misinterpretation. I would argue that a good portion of everyday discourse consists of the quest for clarification. Implicit in the accounts of philosophers of language like Donaldson and Quine is the partial view that underlying discourse in general is a need to determine the purpose of any talk exchange in which one becomes involved. Obviously, this precludes a given talk exchange from, a priori, embodying a clear and present purpose.

My contention with the second point is even more forceful. The purpose of an exchange exists only insofar as each participant contributes an interest to be advanced. After all, if there is no interest of any kind invested in a talk exchange, then there is no point to speaking at all. These interests \textit{might} agree, but nothing about discourse itself predicts that they will do so necessarily. In fact, the interests being advanced and contended in an exchange often do conflict with one another. For example, consider the following: arguments, economics (i.e., appraisals, bartering, price discovery, trade or contract negotiations), formal and informal debates, diplomatic negotiations, legal presentations,


\textsuperscript{4}Ibid., 176.
marital disputes, policymaking, political disagreements, treatymaking. Each of these examples provides an instance in which multiple interests conflict. As such, each interested participant will contribute only as much information as is necessary to advance her interests, for that participant will see the very purpose of the exchange as precisely to advance those interests. This does not mean that exchanges (even the ones in which interests do conflict) cannot be amicable, but only that many exchanges will resemble fistfights or poker games more than they will resemble other, more cooperative enterprises.

In summary, this maxim (M1) cannot do the work Grice needs it to do. It is either empty, because it cannot provide the cooperative glue Grice needs, or it says that participants will make their contributions to a talk exchange only as informative as is necessary to advance their own interests. This does not necessarily presuppose any mutuality of purpose, and that it fails to do so conflicts with Grice’s claim that “I have stated my maxims as if this purpose were a maximally effective exchange of information.” As such, the maxim cannot pack the ethical punch that Grice wishes it to have without begging the question.

Now, let me turn to M2, perhaps the most explicitly ethical of Grice’s maxims: “Try to make your contribution one that is true.” I must first point out the many situations in which this prescription would be useless. “White lies” often do not harm discourse, and they can even provide necessary and valued social lubrication. There also may be many situations in which the truth or falsity of what is said is irrelevant (e.g., compliments, small talk). In more extreme cases, Grice’s prescription might even be harmful; for example, in the well-worn example of someone who betrays to a murderer the hiding place of his would-be victim in the name of truthfulness.

Finally, both M3 and M4 may both be addressed very briefly, and through the lens of my argument against M1. Both the relevance and the perspicuity of a contribution to a talk exchange can be measured only against the interests of the participants. If those interests conflict, then that which may be relevant to the advancement of a particular interest can be completely irrelevant to the advancement of another interest.

Principle Objection

Now I arrive at my principle objection. Before providing his maxims, Grice details what he calls the Cooperative Principle, which he describes as “a rough general principle which participants will be expected to observe, namely: Make your conversational contribution such as is required, at the stage at which it occurs, by the accepted purpose or direction of the talk exchange in which you are engaged.” Grice arrives at this principle by arguing that conversation is a cooperative, mutual effort. Furthermore, “each participant recognizes in [talk exchanges], to some extent, a common purpose or set of purposes, or at least a

5Ibid., 174.

6Ibid., 173.
mutually accepted direction.”

Where the normative force of this principle comes from is unclear, and while many of my previous objections to the individual maxims illuminate certain defects in his reasoning, I will proceed to my objection to this principle.

When he first introduces his Cooperative Principle, Grice provides the aforementioned conversational maxims “on the assumption that some general principle as this [that is, the Cooperative Principle] is acceptable.” He attempts to justify this assumption, however, by claiming that “it is just a well-recognized empirical fact that people do behave in these ways [that is, according to these maxims and this principle].” Support for this claim is scanty, but he does provide one such instance: “It is much easier, for example, to tell the truth than to invent lies.” Grice here appears to sever his account from common moral intuitions. For example, children are often encouraged to tell the truth, even when lying might allow them to escape punishment, and legal threats are required to (at least, try to) prevent perjury in a court of law. If it is easier to tell the truth, then we should expect lying to be far less common than it is. In contrast to Grice, I am reminded of George Eliot, who wrote “Falsehood is easy, truth so difficult.”

It seems that Grice knows that the justification he provides for his fundamental assumption amounts to little, so he examines and discards the possibility of parsing his account in quasi-contractual terms before arriving at the following conclusion: “Anyone who cares about the goals that are central to conversation/communication (e.g., giving and receiving information, influencing and being influenced by others) must be expected to have an interest... in participating in talk exchanges that will be profitable only on the assumption that they are conducted in general accordance with the Cooperative Principle and the maxims.” This can be interpreted in one of two possible ways.

On the first interpretation, Grice is shown to beg the question. After all, according to his account, the Cooperative Principle and the maxims are regulatory of discourse. At the same time, however, he assumes that discourse assumes cooperation as a pre-condition. The circularity here cuts both ways. Discourse is cooperative, because the Cooperative Principle tells us so; and the Cooperative Principle exists, because discourse is cooperative.

---

7 Ibid., 173.
8 Ibid., 173.
9 Ibid., 175.
10 Ibid., 175.
11 Note that I do not wish to encourage my readers to lie, but merely to point out that the skilful liar will often benefit consequentially. It all depends on the sense of “easier” to which Grice appeals, for if a liar is caught in the midst of a lie, after all, the situation may very well go much worse.
12 Ibid., 175.
The second interpretation exploits an ambiguity in the second central goal which he provides, namely, “influencing and being influenced by others.” He appears to disregard the variety of ways in which a speaker might attempt to influence his audience that do not accord with either the Cooperative Principle or the maxims previously discussed. Examples that come to mind include corrupt politicians, predatory marketers, propagandists, and liars of all stripes. Exemplaries of these classes will care about influencing others – according to Grice, a goal central to conversation – yet none of them will act in accordance with the Cooperative Principle or Grice’s maxims. Furthermore, expecting them to act as such could only be justified on quasi-contractual grounds, i.e., some set of restrictions that come into effect the moment a speaker begins to speak.

Conclusion

In conclusion, it may appear that I have avoided discussing Grice’s account of conversational implicatures, but it should be restated once more that his account of implicatures relies entirely on his presumption of both the Cooperative Principle and the conversational maxims. And it is upon these that I have attempted to at least cast some shadow of a doubt. And, if I have been successful, the status of conversational implicature both as a concept and as an object of study is no longer clear.
Stijn Van Impe – The Importance of “Communication” in Kant Reconsidered

Abstract

In contemporary assessments of ethics, increasing attention is being given to the importance of “community” and “communication” for grounding and evaluating moral practices. Champions of discourse ethics as Habermas and Benhabib have focused on the importance of argumentative praxis and practical discourse for grounding an intersubjective moral practice and for establishing a moral communication community. Habermas and Benhabib claim that the Kantian solipsism of a subjectively centered reason must be dissolved in the intersubjective structure of communicative reason. Is this accusation of solipsism and monological reasoning as accurate as proponents of discourse ethics hold? Is it really the case that Kant’s philosophy offers no possibilities for actively engaging to enter into communication with others? Pace Habermas and Benhabib, I will argue that their criticisms are due to a narrow reading of Kant’s philosophy. Therefore, I will examine the communicative dimension in Kant’s philosophy by analyzing his – somewhat neglected – discussions of the sensus communis logicus, the maxims of the common human understanding and the interaction between the freedom to think and the freedom of public communication. On the basis of this analysis, I will conclude that Kant’s moral philosophy does not get entangled in mere thought experiments of monological reasoning and that it does not bypass the possibility of genuine communication.

I

In contemporary assessments of ethics, increasing attention is being given to the importance of “community” and “communication” for grounding and evaluating moral practices. Champions of discourse ethics as Jürgen Habermas and Seyla Benhabib have focused on the importance of argumentative praxis and practical discourse for grounding an intersubjective moral practice and for establishing a moral communication community. Although Habermas offers a quite benevolent evaluation of Kant’s moral community or realms of ends, i.e. “a whole of all ends in systematic connection (a whole both of rational beings as ends in themselves and the ends of his own that each may set himself)” (G 4:433), he nonetheless argues for the need of remodeling Kant’s underlying metaphysical

1In case of references to works of Kant, all references are given in the text itself and are to Immanuel Kant, Gesammelte Werke, edited by the Royal Prussian (later German) Academy of Sciences (Berlin: Walter de Gruyter, 1990–). Volume numbers are followed by page numbers. Translations are taken from Immanuel Kant, The Cambridge Edition of the Works of Immanuel Kant, general editors Paul Guyer and Allen W. Wood (Cambridge: Cambridge University Press, 1992–). Used abbreviations: Ant: Anthropology from a
frameworks. One of the crucial aims of Habermas’s reformulation of Kant’s moral theory concerns the shift from a monological to a dialogical account of moral justification rejecting Kant’s so-called assumption that “the individual tests his maxims of action in foro interno”. In similar fashion, Benhabib argues that in discourse ethics “the procedural model of an argumentative praxis replaces the silent thought-experiment enjoined by the Kantian universalizability test”. In the end, Habermas and Benhabib hold the view that the Kantian solipsism of a subjectively centered reason must be dissolved in the intersubjective structure of communicative reason. Is this accusation of solipsism and monological reasoning as accurate as proponents of discourse ethics hold? Is it really the case that Kant’s philosophy offers no possibilities for actively engaging to enter into communication with others? Pace criticisms à la Habermas and Benhabib, I shall argue to the contrary. I will show that such criticisms are largely due to a narrow reading of Kant’s philosophy. Both Habermas and Benhabib seem to neglect the communicative dimension in Kant’s philosophy by bypassing his discussions of the sensus communis logicus, the maxims of the common human understanding and the interaction between the freedom to think and the freedom of public communication.

II

In What Does it Mean to Orient Oneself in Thinking? Kant makes the explicit connection between the freedom to think and the freedom of public communication. After stressing that the freedom to think is opposed to all external compulsion, he writes, “it is said that the freedom to speak or to write could be taken from us by a superior power, but the freedom to think cannot be. Yet how much and how correctly would we think if we did not think [...] in community with others to whom we communicate [mittheilen] our thoughts, and who communicate theirs with us! Thus one can very well say that this external power which wrenches away people’s freedom publicly to communicate their thoughts also takes from them the freedom to think.” (WO 8:144). Kant thus explicitly argues for seeing the freedom of public communication as a kind of condition or guarantee for the freedom to think. As Katerina Deligiorgi notices, “the freedom to communicate with real interlocutors is essential, for without it we lose our capacity to think freely” and “communication is not merely an empirical test of the ideas and the rules we adopt in our thinking, but rather a

\[Pragmatic Point of View (1798); CJ: Critique of the Power of Judgment (1790); G: Groundwork of the Metaphysics of Morals (1785); JL: Jäsche Logic (1800); WE: What is Enlightenment? (1784); WO: What Does it Mean to Orient Oneself in Thinking? (1785).\]


distinctive condition of autonomous reasoning”. In this regard, Onora O’Neill additionally explains that the main reason for Kant’s omission of giving an account for the material and social requirements for exercising intellectual freedom under various historical circumstances lies in the fact that “he is concerned with a more fundamental requirement for a communication to be public”. If the freedom to think presupposes the freedom of public communication, Kant’s main interest does not concern the specific requirements for exercising intellectual freedom as such, but the more fundamental issue of public communication and of the requirements that have to be met for communication to be public. For my present purposes, it is first and foremost crucial to see that – as made clear in the passage from What Does it Mean to Orient Oneself in Thinking? quoted above – the freedom to think is but possible – or safeguarded – on the condition of there being a genuine, public communication with others. Hence, the act of free and autonomous reasoning cannot be accounted for as being a merely solipsistic act: from the perspective outlined above, there can be even no such thing as merely monological thought experiments. As O’Neill aptly puts it, “Intellectual freedom is from the start not merely freedom to engage in inward or solitary reflection.” Kant thus holds the view that without – at least – the possibility of there being genuine public communication, in which we actually share our thoughts with others, we could not even think freely at all. Hence, it rules out the accusation of Kant’s conception of reasoning as being merely “in foro interno” as Habermas and Benhabib tend to hold.

III

Another approach to refuting criticisms à la Habermas and Benhabib is to stress the irreducible social and communicative dimension of critical reflection. Therefore, I turn to Kant’s idea of the sensus communis and his maxims of the common human understanding. In the Critique of Judgment Kant argues that “By ‘sensus communis’ must be understood the idea of a communal sense [gemeinschaftliche Sinn], i.e., a faculty for judging that in its reflection takes account of everyone else’s way of representing in thought, in order as it were to hold its judgment up to human reason as a whole [gesamte Menschenvernunft]” (CJ 5:293-294). As Stuart Dalton rightly notices, the sensus communis is a capacity that is “common to everyone”, and at the same time it is a capacity “to discern what is common to everyone: it is not just a sense that is common, it is also a sense of the common.” It falls beyond the scope of this paper to deal with all the subtleties and intricacies of Kant’s


6Ibid.

discussions of the sensus communis throughout the third Critique, where he assesses the sensus communis alternatively in terms of “the free play of our cognitive powers” (CJ 5:238), “a merely ideal norm” (CJ 5:239), “a constitutive principle of the possibility of experience” (CJ 5:240), or “a faculty for judging” (CJ 5:293). For my present purposes, however, it is necessary to notice that Kant makes a sharp distinction between “sensus communis aestheticus” to designate taste, viz. to designate the “subjective principle, which determines what pleases or displeases only through feeling and not through concepts” (CJ 5:238), on the one hand, and “sensus communis logicus” to designate the common human understanding, which “judges not by feeling but always by concepts” (CJ 5:238), on the other hand. Hence, as Kant holds, the sensus communis aestheticus refers to the aesthetic power of judgment, while the sensus communis logicus with its entailed maxims of the common human understanding (gemeiner Menschenverstand) refers to the intellectual power of judgment (CJ 5:295).8

In his Anthropology from a Pragmatic Point of View Kant interprets the sensus communis logicus as a “subjectively necessary touchstone of the correctness of our judgments generally, and consequently also of the soundness of our understanding” by the fact that, as such, it “restrains our understanding by the understanding of others” (Ant 7:219), which is another way of saying – as he did in the third Critique – that it spurs us to hold our judgment up to “human reason as a whole” (CJ 5:294), i.e. to the judgments of others. Only by holding up one’s judgment to the judgments of others and by putting oneself in the position of everyone else, can one avoid that subjective private conditions could be mistakenly taken as being objective, which would yield erroneous judgments. Additionally, Kant explicitly stresses that this operation is to be seen in opposition to “isolating ourselves with our own understanding and judging publicly with our private representations” (Ant 7:219). This clearly hints that the operation mode of the sensus communis logicus is no mere solipsistic act of monological reasoning. This is made even further evident as Kant emphasizes that the sensus communis logicus stands in fundamental opposition to the sensus privatus logicus and that in relying on the latter one gets locked up in “a play of thoughts in which he sees, acts, and judges, not in a common world, but rather in his own world (as in dreaming)” (Ant 7:219). Finally, Kant connects the exercising of the sensus communis logicus as the “greatest and most useful means of correcting our own thoughts” to the fact “that we advance them in public in order to see whether they also agree with the understanding of others; for otherwise something merely subjective [...] would easily be taken for something objective” (Ant 7:219). Once again, Kant thus stresses the importance of rendering our thoughts publicly and of actually confronting them to the understanding of others. Hence, it becomes evident that claims arguing that Kant’s sensus communis logicus comes down to a mere solipsistic and monological thought experiment locked up and executed in the mind of the solitary thinker, are by no means valid.

---

8See also Hannah Arendt, Lectures on Kant’s Political Philosophy (Chicago: University of Chicago Press, 1989), 70-72; Gundula Felten, Die Funktion des Sensus Communis in Kants Theorie des ästhetischen Urteils (Paderborn: Fink, 2004), 129; Tamar Japaridze, The Kantian Subject. Sensus Communis, Mimesis, Work of Mourning (New York: SUNY Press, 2000), 55. Although these authors do mention the difference between the sensus communis aestheticus and the sensus communis logicus, they focus exclusively on the former.
That Kant considers active engagement in interaction and genuine communication, is not only evidenced by the considerations above, but also by taking a closer look at the maxims of the common human understanding. In the third critique these maxims are stated as: “1. To think for oneself [Selbstdenken]; 2. To think in the position of everyone else [an der Stelle jedes anderen denken]; 3. Always to think in accord with oneself [jederzeit mit sich selbst einstimmig denken]” (CJ 5:294; see also Ant 7:228; JL 9:57). These maxims are further referred to as the unprejudiced, the broad-minded and the consistent way of thinking respectively (CJ 5:294). For my present purposes, I shall only discuss the first two maxims. The first maxim guiding our common human understanding is “to think for oneself”, which Kant further characterizes as “the maxim of the unprejudiced way of thinking” or “the maxim of a reason that is never passive” (CJ 5:294). This maxim can be directly related to Kant's famous creed of enlightenment, “Have courage to make use of your own understanding!” (WE 8:35), which means that one should not adhere unquestioningly to external authorities such as the state or the church, but solely to the internal authority of one's own reasoning in forming one's opinions and judgments. In this sense, the first maxim is to be understood as the maxim of a reason that is not passive but active, that is not heteronomous but autonomous, and that does not rely on other sources of authority than on its own. As such, this maxim spurs us to emancipate ourselves from the “need to be led by others” in our thoughts and judgments (CJ 5:295) and is therefore also called “the principle of freedom from constraint” (Ant 7:228).

The question remains, however, how the first maxim may contribute to genuine and public communication. An interesting perspective on this issue is developed by O'Neill who argues that “if there is to be genuine communication and debate, all parties must be guided by such a maxim, for otherwise understanding and agreement will be spurious, mere echoing of what the other or many assert. Genuine communication occurs only between beings who are at least partially separate.” What O'Neill seems to have in mind here, is the following. Genuine communication implies a kind of independency in thinking of the interlocutors involved and hence a kind of distinctiveness between the speakers and the audiences, which is only possible if all involved parties think for themselves so that a plurality of opinions and judgments can prevail. If this were not to be the case, genuine communication and debate would not even be possible at all: one could only expect a mere repetition or

---


10 The maxims of the common human understanding are not only discussed in Kant's officially published works, but are also mentioned in his so-called “Reflexionen” or the unpublished notes from his handwritten legacy. See Kant, Gesammelte Werke, 15:715; 15:820; 15:822.

echoing of only one or very similar opinions or judgments, which does away even any possibility of disagreement or dissent and hence yields a conflation of speakers and audiences. In such case, there would be even nothing to be debated about or to be communicated on. So, if and only if everyone thinks for himself, can genuine communication and debate be possible. For this reason, O’Neill rightly concludes that where nobody thinks for himself there is no “plurality of viewpoints” to be voiced, heard and debated so that we can no longer participate in any discussion. In this way, the first maxim has great importance in establishing genuine communication and dialogue as it safeguards a plurality of debatable standpoints, opinions and judgments, which is a condition sine qua non for taking part in any kind of communication and dialogue.

V

Let us now turn to the second maxim of common human understanding. In the third critique this maxim is stated as “to think in the position of everyone else” (CJ 5:294) and is referred to as the “broad-minded way of thinking” or the “maxim of enlarged thought” (CJ 5:294). This is achieved if the individual “sets himself apart from the subjective private conditions of the judgment [...] and reflects on his own judgment from a universal standpoint (which he can only determine by putting himself into the standpoint of others)” (CJ 5:295). What is characteristic of this maxim, is that it requires us to broaden our horizon of thinking by abstracting from subjective private conditions that could come into play into our reasoning and judging and that could lock us up in the narrow constraints of what we – from our own limited point of view – assume to be correct. An antidote to this is provided by placing oneself in the position of others and by reflecting critically whether one’s judgment could hold for others as well. The maxim of enlarged thought thus spurs us to critically examine – and re-examine – our own judgments by taking into account others’ judgments and by looking at our own judgments from the perspective of others. In exercising this maxim, it is not a question of taking up a so-called view from nowhere or absolutely neutral and transcendent standpoint. On the contrary, it is a question of actively engaging oneself in and moving between a plurality of different points of view. As O’Neill aptly remarks, “taking a universal or detached view cannot be a matter of adopting any neutral, Archimedean standpoint, but only one of seeking to see one’s own initial judgment from the standpoint of others.”13 As such, the second maxim is a dynamic principle in that it prescribes interaction with others.

Critics à la Habermas and Benhabib might argue at this stage, that, while it may be true that this maxim invites us to interact with others to critically examine our judgments, the interaction in question ultimately boils down to another thought experiment “in foro interno”: the interaction is merely carried out in thought, but not in any actual practice of communication, and, as such, it only entails merely possible but not real communicative interaction. However, they are mistaken for in reformulating the maxim of enlarged

12Ibid., 543-544.

13Ibid., 544.
thought in the Anthropology Kant does explicitly take notice of the fact that this mode of enlarged thought is to be established in and through communication with others. As he makes clear, the maxim of the broad-minded way of thinking or “the principle of liberals who adapt to the principles of others” is “to think oneself (in communication [Mittheilung] with human beings) into the place of every other person” (italics mine) (Ant 7:228). The second maxim thus entails much more than a mere thought experiment. On the contrary, it entails that taking up the perspectives of others has to be effected in and through actual communication. O'Neill is thus right in arguing that “one who adopts a maxim of enlarged thought must therefore listen to what others are actually judging and communicating”¹⁴. Hence, Kant clearly attributed to communication the function of being the medium through which we can take notice of other people’s judgments and opinions, and through which we can hence broaden our horizon of thinking.

VI

In this paper, I have argued that – contrary to what certain critics of Kant claim – “communication” does play a major role in Kant’s philosophy, although it has to be admitted that Kant does not develop a systematic philosophy of communication as such. For instance, he does not discuss in any detail how the process of communication precisely proceeds. Neither does he set up any specific, structural rules or procedural standards upon which the process of communication ought to be based, as e.g. Habermas does in The Theory of Communicative Action or in Moral Consciousness and Communicative Action.¹⁵ However, by focusing on Kant’s sensus communis logicus and the maxims of common human understanding, I have argued that criticisms à la Habermas and Benhabib who claim that Kant’s moral philosophy gets entangled in mere thought experiments of monological reasoning and that it bypasses the possibility of genuine communication, have no valid

¹⁴Ibid.

¹⁵See e.g. Habermas’s structural rules for reaching mutual agreement in discourse, which can be summarized as follows. The rules at the logical-semantic level of products are “(1.1) No speaker may contradict himself. (1.2) Every speaker who applies predicate $F$ to object $A$ must be prepared to apply $F$ to all other objects resembling $A$ in all relevant aspects. (1.3) Different speakers may not use the same expression with different meanings.” The rules at the dialectical level of procedures are: “(2.1) Every speaker may assert only what he really believes. (2.2) A person who disputes a proposition or norm not under discussion must provide a reason for wanting to do so.” And the rules at the rhetorical level of processes are: “(3.1) Every subject with the competence to speak and act is allowed to take part in a discourse. (3.2.) (a) Everyone is allowed to question any assertion whatever. (b) Everyone is allowed to introduce any assertion whatever into the discourse. (c) Everyone is allowed to express his attitudes, desires and needs. (3.3) No speaker may be prevented, by internal or external coercion, from exercising his rights as laid down in (3.1) and (3.2).” See Habermas, Moral Consciousness and Communicative Action, 87-89.
grounds. If Kant attributed a thoroughly important role to communication via the sensus communis logicus and the maxims of common human understanding, as I have shown, it would be harsh to exclude these considerations from his discussions of the categorical imperative and the realm of ends. Although Kant does not treat these issues in his moral philosophy as such, there is no sufficient reason to assume that, when Kant deals with the categorical imperative and asks us to consider whether a maxim of action could hold as a universal principle for all action, he necessarily restricts himself to mere thought experiments. Likewise, when Kant asks us to reflect on a maxim for moral action by taking up the standpoint of everyone else, thereby creating a realm of ends by taking into account the morally legitimate ends of all possible rational beings, it would be utterly unfair to argue that he takes refuge to solipsistic mind games and that he is not interested in nor concerned with the need of at least a minimalist form of genuine communication for bringing about the realm of ends. On the contrary, I here concur with Scott Stroud and argue that the maxims of common human understanding can be seen as “a structure for the general behavior of communicators” in the realm of ends.\footnote{Scott R. Stroud, “Rhetoric and Moral Progress in Kant’s Ethical Community”, \textit{Philosophy and Rhetoric} 38:4 (2005): 345.} To accept Habermas’s and Benhabib’s claims would be contradictory to the communicative spirit of Kant’s philosophy as worked out in his discussions on the sensus communis logicus, the maxims of the common human understanding and the interaction between the freedom to think and the freedom of public communication.\footnote{I wish to thank Bart Vandenabeele for his generous support with regard to my research and for inspiring discussions on the topic of this paper. I also wish to thank the Scientific Research Foundation Flanders for the financial support.}


\footnote{I wish to thank Bart Vandenabeele for his generous support with regard to my research and for inspiring discussions on the topic of this paper. I also wish to thank the Scientific Research Foundation Flanders for the financial support.}


