METEORITE

The University of Michigan
Ann Arbor

Contents

A Critique of Cartwright's Entity Realism: Redundancy of Causal Explanations

Nicholas M.G. Friedman
Skidmore College

Abstract

Nancy Cartwright attempts to justify the existence of the unobservable entities postulated by our best scientific theories without appealing to fundamental scientific theories themselves. According to Cartwright, we can identify the most probable cause of our empirical findings through carefully controlled experiments. Once we have devised a causal explanation of our empirical findings, we are licensed to infer the existence of the entities invoked in the given explanation. Cartwright believes that her approach avoids the problems that Bas van Fraassen claims plague inference to the best explanation. Here, I will argue that Cartwright's reliance on causal explanation still succumbs to the problem of redundancy: There will always be numerous conceivable causal explanations that are consistent with empirical findings. Thus, we cannot identify the most probable cause of our empirical findings as Cartwright insists. I will explore a possible modification of Cartwright's entity realism that does overcome the problem of redundancy, but I will ultimately argue that the modified position is unsatisfactory.

Introduction

Nancy Cartwright attempts to carve out a middle ground between realism and antirealism known as entity realism. The stance – also advocated for by Ian Hacking – conjoins realism about unobservable theoretical entities with antirealism about scientific theories. In justifying the existence of theoretical entities, Cartwright intends to avoid the use of inference to the best explanation (IBE), a form of inference that Bas van Fraassen argues is plagued by two problems, both of which I will later discuss in detail: the problem of metaphysical frugality and the problem of redundancy. Instead of relying on IBE, Cartwright relies on causal explanation and inference to the most probable cause (IPC). According to Cartwright, IPC can be employed to justify the existence of theoretical entities without justifying the truth of scientific theories, all while overcoming the problems associated with IBE (Cartwright 1983; Clarke 2001). Here I will argue that Cartwright's invocation of IPC fails to overcome the problem of redundancy. I suggest that there are two potential routes that Cartwright might take to overcome the problem, both of which I will argue are ultimately unsatisfactory. Cartwright might rely on the use of pragmatic virtues, but there is no clear connection between pragmatic virtues and truth. In addition, the use of pragmatic virtues licenses IBE for theoretical explanations – something Cartwright needs to avoid. Alternatively, the problem of redundancy can be overcome if we are willing to define entities in terms of their observational consequences. Although defining entities as such might provide a justification for entity realism,

the resulting version of entity realism is very weak. Thus, Cartwright's attempt to carve out of a middle ground between realism and antirealism is ultimately unsuccessful.

The Scientific Realism/Antirealism Debate

The scientific realism/antirealism debate concerns the extent to which science describes and reflects the world as it really is. A scientific realist will typically subscribe to one or more of the following claims: (1) theories are truth-apt, and their truth status does not depend on our beliefs; (2) the truth or falsity of theories depends on the mind-independent state of the world; (3) scientific theories aim at the truth; (4) some or most of our currently accepted scientific theories are true or approximately true; or (5) the unobservable entities postulated by scientific theories do exist. Antirealists will deny some or all of the above claims (Resnik 1994). Notice that claim (5) concerns postulated entities whereas claims (1) through (4) concern scientific theories. Hence, a distinction may be invoked between theory realism and entity realism, the importance of which will be revealed below.

Historically, scientific realists have attempted to justify their position by appealing to the success of science. According to Hilary Putman's "No Miracle Argument," the truth of scientific theories and the reality of postulated entities may be inferred from the tremendous predictive, explanatory, and instrumental success of the scientific endeavor. It would be a miraculous coincidence after all if our scientific theories do not describe the world as it really is, and yet somehow still help us to predict the motion of planets, explain the behavior of chemical reactions, and invent life-saving medicines. In brief, scientific realism is our best explanation of the success of science (Putnam 1975).

Whereas realists see the glass as half-full, antirealists see the glass as half-empty. Instead of appealing to the success of science, antirealists have historically appealed to the failures of science. In formulating his "pessimistic induction," Larry Laudan notes that countless theories which previously enjoyed predictive, explanatory, and instrumental success have since been proven, in some sense, to be false. It seems that success might not be the most sensitive indicator of truth (Laudan 1981). Moreover, it seems that we ought to infer by enumerative induction that our currently accepted and seemingly successful theories will one day be rejected as well.

Hopefully these brief descriptions of the "No Miracles Argument" and the "pessimistic induction" shed some light on the intuitions behind the realist and antirealist positions, respectively. Although the scientific realism/antirealism debate is certainly richer than I have let on, it should be clear that no consensus has been reached in the field. Given the difficulties in maintaining either the realist or antirealist position, attempts have been made to develop a middle ground stance between the two. Although our scientific theories might not be true, as the antirealists hold, perhaps there are specific elements of our most successful scientific theories that are deserving of our epistemic commitment. In formulating structural realism, John Worrall (1989) has notably claimed that only the mathematical and structural content of our most successful theories deserve our commitment. Alternatively, in formulating entity realism, Ian Hacking (1982) and Nancy Cartwright (1983) have argued that we are warranted to believe in the existence of the unobservable entities postulated by our most successful theories without believing in the truth of the

theories themselves.

Entity Realism: A Promising Middle Ground

If tenable, entity realism offers advantages over both [theory] realism and antirealism. Whereas the antirealist struggles to explain the success of science, the entity realist can claim that the success of science is explained by the fact that our theories invoke entities that do actually exist. And whereas the theory realist struggles to account for the high turnover rate of scientific theories, as noted in the above discussion of the "pessimistic induction," the entity realist can claim that unobservable entities are typically retained even as theories are replaced. Although entity realism might *arguably* be safe from the "No Miracles Argument" and the "pessimistic induction," we must still investigate whether the position enjoys independent support.

Hacking argues that experimental practices provide the best justification for belief in the existence of unobservable scientific entities, or theoretical entities as he refers to them. The best evidence for the existence of theoretical entities like electrons is that they can be manipulated during experimentation: *"We are completely convinced of the reality of electrons when we regularly set out to build – and often enough succeed in building – new kinds of devices that use various well understood causal properties of electrons..."* (Hacking 1982, 71, Hacking's emphasis). Although a substantial argument behind Hacking's claim is difficult to pinpoint in his works, Hacking has been widely interpreted as offering an inference to the best explanation. Hacking seems to be arguing that the best explanation of our ability to reliably generate experimental effects is the existence of our theoretical entities. Interpreted as such, Hacking's argument has been subjected to a range of devastating criticisms (Resnik 1994; Reiner and Pierson 1995; Sankey 2012; Miller 2016). Although I am sympathetic to Hacking's endeavor, I take Cartwright's attempt to justify the existence of theoretical entities to be more promising. But alas, I will ultimately argue that Cartwright also fails to carve out a middle ground between realism and antirealism.

Two Problems with Inference to the Best Explanation (IBE)

In attempting to justify the existence of theoretical entities, Cartwright attempts to avoid two problems that Bas van Fraassen claims plague inference to the best explanation (IBE): the problem of redundancy and the problem of metaphysical frugality.[1] It will be worthwhile to describe these problems before delving into Cartwright's argument. As discussed above, the scientific realist often relies on IBE as a means to justify the truth of our scientific theories. Bas van Fraassen (1980) criticizes IBE, however, because there is not an immediately obvious connection between explanatory power and truth. As Cartwright (1983) puts it, an argument is needed to show that "if x explains y and y is true, then x should be true as well" (4). The scientific realist needs to provide a strong argument as to why strong scientific theories are likely to be true just because their truth explains the success of science. Un-

---

[1]Although the problem of metaphysical frugality is described by van Fraassen, the use of the term "metaphysical frugality" in this context may be attributed to Steve Clarke (2001).

fortunately for the scientific realist, the problem of redundancy and the problem of metaphysical frugality create a barrier for drawing a connection between explanatory power and truth (van Fraassen 1980).

First, there is the problem of redundancy. There are usually multiple scientific theories that are consistent with our observations and empirical findings. Typically, scientists rely on what are known as "pragmatic virtues" like simplicity or elegance to select amongst empirically equivalent theories. We might have multiple potential theories that are consistent with observed planetary motion, but we're likely to select the mathematically simplest theory, for example, as the best explanation of planetary motion. Reliance on pragmatic virtues is problematic to van Fraassen. It is not at all clear how pragmatic virtues connect us to truth. We might prefer simpler or more elegant theories, but why should such theories be more likely to be true? If we cannot rely on pragmatic virtues to select amongst theories, we are left with multiple theories all equally consistent with empirical findings (van Fraassen 1980). We are left with, what Cartwright (1983) refers to as, a redundancy of theories. And as Clarke (2001) notes, if we cannot decide which scientific theory is the best explanation of empirical findings, we certainly cannot infer the truth of a particular scientific theory.

The second problem is that of metaphysical frugality. Even if we are able to identify the best explanation of our empirical findings, why do we need to conclude that the best explanation is true? We seem to be making an unwarranted and unnecessary assumption in jumping from explanatory power to truth. Nothing demands that we make this jump. If we are metaphysically frugal, we will be content with merely claiming that a theory is the best explanation of empirical findings without needing to also claim that the theory is true (van Fraassen 1980). In fact, according to Cartwright (1983), it would be coherent to say that we have strong theoretical explanations that are not true, i.e. that do not describe the world as it really is.

Van Fraassen's critique of IBE, endorsed by Cartwright, deals a blow to the realist. Pragmatic virtues must be employed to narrow in on the best explanation, but there is no obvious connection between pragmatic virtues and truth. Even if the realist uses pragmatic virtues to "find" the best explanation, the principle of metaphysical frugality demands that we cannot accept the metaphysically-rich claim that the best explanation is true. To the extent that IBE is plagued with the problem of redundancy and of metaphysical frugality, IBE is insufficient to ground the truth of our scientific theories.

Cartwright's Justification for Entity Realism

Cartwright attempts to justify the existence of theoretical entities without justifying the truth of scientific theories in such a way that avoids the two problems that van Fraassen identifies as plaguing IBE.[2] According to Cartwright, scientists are able to perform well-designed, carefully-controlled experiments to isolate the most probable cause of empirical findings. Through manipulation of theoretical entities during experimentation, scientists are able to eliminate most other potential causes:

---

[2]Cartwright distinguishes between phenomenological theories and fundamental theories. For Cartwright, phenomenological theories are descriptive, whereas fundamental theories are explanatory. Cartwright is an antirealist about fundamental theories, but is an acknowledged realist about phenomenological theories (Cartwright 1983).

> We make our best causal inferences in very special situations—situations where our general view of the world makes us insist that a known phenomenon has a cause; where the cause we cite is the kind of thing that could bring about the effect and there is an appropriate process connecting the cause and the effect; and where the likelihood of other causes is ruled out...Seldom outside of the controlled conditions of an experiment are we in a situation where a cause can legitimately be inferred. (Cartwright 1983, 5)

Once we have identified a casual explanation of empirical findings, we are licensed to infer that the entities invoked in a given causal explanation do in fact exist. According to Cartwright (1983), this inference to the most probable cause (IPC) is justified in a way that IBE is not because IPC avoids the two problems that plague IBE: "[van Fraassen's] challenge [to show how explanatory power and truth are connected] has an answer in the case of *causal explanation*, but *only* in the case of causal explanation" (4, Cartwright's emphasis).

Cartwright suggests that IPC avoids the problem of metaphysical frugality. Causal explanations seem to have an existential component built into them. It does not seem to make any sense to provide a causal explanation without automatically positing the existence of entities invoked in the causal explanation. It seems meaningless to say that a pathogenic microbe is the cause of a patient's disease without holding that the pathogenic microbe exists. Unlike theoretical explanations, where we might be able to get away with saying that a good theory invokes non-existent entities and yet provides good explanations and predictions, we cannot really say the same thing about causal explanations. Insofar as we can identify a strong causal explanation, we have good evidence for the existence of the purported causal entities (Cartwright 1983). Questions have been raised about whether or not causal explanation really does involve an existential component (Clarke 2001), but I wish to focus in this paper on the question of whether causal explanation helps us overcome the problem of redundancy.

Cartwright (1983) insists that causal explanation helps us overcome the problem of redundancy because we can infer the most probable cause of empirical findings without needing to pick one fundamental theory about the purported causal entity: "I infer to the most probable cause, and that cause is a specific item, what we call a theoretical entity. But note that the electron is not an entity of any particular theory, it is the electron, about which we have a large number of incomplete and sometimes conflicting theories" (92). Our commitment to the existence of the electron is compatible with numerous fundamental theories that invoke the electron. We seem to have overcome the problem of redundancy by avoiding the need to select particular theoretical explanations for our empirical findings altogether.

Redundancy of Theoretical Entities and Causal Explanations

I grant that Cartwright's IPC prevents the need to pick amongst rival theories. But I argue that we still encounter the problem of redundancy. Although Cartwright might have shifted the problem away from a redundancy of theories, there is still a problem of redundancy with respect to entities and causal explanations. Cartwright claims that through carefully controlled experiments, we can identify the most prob-

able causal entity responsible for observed results. But aren't there all sorts of viable alternatives that are equally consistent with the empirical observations? Suppose, through carefully controlled experiments, we seem to isolate the electron as the cause of a detectable trail in a cloud chamber particle detector. We can still debate about the properties that the electron might have. For example, I might claim that the electron has mini spikes on it, whereas you might deny this arbitrary claim. It seems that we now have two possible causal entities responsible for our empirical findings – namely, the electron with and without spikes. And of course, we could conceive of a seeming infinite number of arbitrary properties that the electron might possess. If we propose that these various conceivable alternatives are equally consistent with the empirical findings, then we would have to rely on pragmatic virtues to select the most probable cause.[3] But again, we should avoid a reliance on pragmatic virtues because there is not an immediately obvious connection between their use and truth (van Fraassen 1980). And moreover, if we license the use of pragmatic virtues to identify the most probable cause of our empirical findings, what would stop us from licensing its use in the inference to the best theoretical explanation? A reliance on pragmatic virtues would seem to shift us from entity realism towards theory realism (Clarke 2001).

I have argued that Cartwright is mistaken in thinking that we can single in on the most probable cause, because we can conceive of many potential causal entities each with different properties from one another that are all equally consistent with empirical findings. Steve Clarke rightly notes, however, that possible causal entities that are seemingly different might actually be the same entity. In conceiving of different properties that electrons might have, in the above example, we are still pointing to one single possible cause – namely, the electron (Clarke 2001). Perhaps there is no redundancy in causal entities, as I had argued, but just a redundancy in the properties of causal entities. So, perhaps Cartwright is right that we can single in on the most probable cause, even if we might justifiably disagree about the properties that can be attributed to the most probable cause. Although we might have to rely on pragmatic virtues to select the properties that our causal entities might have, Cartwright seems to be right that we can identify the causal entity itself without reliance on pragmatic virtues by performing carefully controlled experiments. That being said, even if we grant that there is no redundancy in causal entities, a redundancy in the properties of casual entities is itself problematic for Cartwright. If we cannot identify what properties a causal entity has, because of the problem of redundancy, then how can we devise causal explanations that invoke the given entity? How could we possibly explain how an electron causes a trail in a cloud chamber particle detector if we are agnostic about the properties that the electron has? Insofar as the existence of theoretical entities is justified through causal explanation, as it is on Cartwright's account, we need to be able to concretely identify the properties of causal entities.

Here I must admit that I have neglected to make a distinction between arbitrary entity properties and causally impactful ones. Even if defenders of Cartwright grant my claim that there is redundancy in arbitrary entity properties, like spikes or no spikes on an electron, Cartwright's defenders might argue that there is only redun-

---

[3]I owe my gratitude to Dr. Larry M. Jorgensen for noting that the principle of metaphysical frugality might also caution us against identifying entity properties, unless we are demanded to by the empirical findings. It would be metaphysically frugal to remain agnostic about the properties than an entity might have.

dancy in the case of arbitrary properties. If we consider non-arbitrary, causally influential properties, like the charge or mass of an electron as opposed to whether it has mini spikes, then there will be no redundancy. And we can devise causal explanations that incorporate casually influential properties without incorporating arbitrary properties. Whereas a seemingly infinite set of arbitrary entity properties might be consistent with empirical findings, carefully controlled experiments should be able to isolate the specific causally influential properties that our causal entities possess. For example, the accepted charge and mass of an electron presumably fits into our causal explanation of how the electron might produce an observable trail in a cloud chamber particle detector. If the electron had a different charge or a different mass, there would be clear detectable consequences that we could identify through experimentation. Within our given causal explanations, the possible causally influential properties that our causal entities might possess are fixed; there is no redundancy with respect to causally influential properties of entities.

This defense of Cartwright is reasonable, but we could also conceive of different causal explanations that invoke different mechanisms and entities that are all consistent with empirical findings. Instead of invoking electrons in our causal explanations of the observable trail in a cloud chamber, we could conceive of altogether different particles or unobservable minions running around. It might take some creativity, but we should be able to conceive of alternate explanations that are consistent with the empirical findings. Now, of course, pragmatic virtues will drive us to select the electron explanation over a silly explanation that involves minions. But again we are committed to avoiding a reliance on pragmatic virtues.

It might be objected that while our many conceivable alternate causal explanations are currently consistent with empirical findings, we will be able to eliminate all but the true causal explanation as we develop better experimental techniques. As we develop our detection methods and, say, build an exceptionally precise microscope, we will be able to see if there are electrons or minions in the cloud chamber. Improved detection methods and experimental techniques will certainly allow us to eliminate more and more causal explanations, but with a bit of ingenuity we should always be able to conceive of multiple causal explanations that are consistent with empirical findings.

It might be further objected that if our many conceived of causal explanations, invoking different entities, all lead to the same observable consequences, then aren't we really talking about the same entities? This objection requires that we define entities in terms of their observable consequences. An electron must not be defined by its structure or unobservable behaviors, but by its observable consequences. An electron is simply that entity that causes a trail in a cloud chamber. In a sense, I think this objection offers the best defense of Cartwright's entity realism. Defining entities in this way overcomes the problem of redundancy. We cannot conceive of multiple potential causal entities that are consistent with our observations if our entities just are – and are nothing more than – the very things that cause our observations. We can be entity realists, but it is a very thin version of entity realism and one that I think Cartwright would reject. We could be realists about entities insofar as entities are simply defined by their functions, but we could not be realists about atomic structure, the unobservable behavior of genes, etc. I am unfortunately reminded of Alan Musgrave's caricature of the entity realist position:

> To believe in an entity, while believing nothing further about that entity,
> is to believe nothing. I tell you that I believe in hobgoblins. So, you reply,

> you think there are little people who creep into houses at night and do the housework. Oh no, say I, I do not believe that hobgoblins do that. Actually I have no beliefs at all about what hobgoblins do or what they are like. I just believe in them. (Musgrave 1996, 20)

I do not think that the modified form of entity realism that I have presented as an escape for Cartwright is quite as amusing. Nonetheless, insofar as what we want to be realists about is the unobservable structures and behaviors of theoretical entities, this modification of Cartwright's position is deeply unsatisfying.

Conclusion

Cartwright has attempted to stake out a realism about entities conjoined to an antirealism about theories. To justify the existence of theoretical entities without appealing to fundamental theories, Cartwright has relied on inference to the most probable cause (IPC). According to Cartwright, IPC avoids the problem of metaphysical frugality and the problem of redundancy that van Fraassen claims have plagued inference to the best explanation (IBE). I have argued, in contrast to Cartwright, that IPC does not overcome the problem of redundancy. Whereas Cartwright believes that we can isolate the most probable cause through careful experimentation, I have argued that we can always conceive of alternate causal explanations that are consistent with the experimental facts. The only way to select amongst the alternative explanations would be to rely on pragmatic virtues.

For van Fraassen, reliance on pragmatic virtues is problematic because there is no obvious connection between the pragmatic virtues and truth. An argument is needed to show that causal explanations that are simpler or more elegant are more likely to be true (van Fraassen 1980). Moreover, relying on pragmatic virtues in IPC might license their use in IBE as an argument for theory realism (Clarke 2001). For pragmatic virtues to be of use in staking out the middle ground of entity realism, an argument is needed to show that pragmatically virtuous causal explanations are more likely to be true, whereas pragmatically virtuous theoretical explanations are not necessarily more likely to be true. Or perhaps there are *particular* pragmatic virtues than can be used to select amongst competing causal explanations which could not be used to select amongst competing theoretical explanations. More work needs to be done, but I am not confident that pragmatic virtues can be successfully employed to stake out the middle ground of entity realism.

Maintaining the middle ground therefore seems to require that we provide a functional definition of entities in which we define entities merely in terms of their observational consequences. While this might allow us to hold on to entity realism, we lose the most exciting features of entity realism. On this modified version of entity realism, we do not get to be realists about the unobservable structures and behaviors of theoretical entities. It is not at all clear that this weak version of entity realism amounts to anything more than a realism about causes. If we are realists about entities, insofar as entities are merely the causes of observed phenomena, are we not just realists about observed phenomena having causes? Further work is needed to determine whether this version of entity realism provides us with justification for actually meaningful and exciting claims about theoretical entities.

References

Cartwright, Nancy. *How the Laws of Physics Lie*. Oxford: Oxford University Press, 1983.

Clarke, Steve. "Defensible Territory for Entity Realism." *The British Journal for the Philosophy of Science* 52, no.4 (2001): 701-722.

Hacking, Ian. "Experimentation and Scientific Realism." *Philosophical Topics* 13, no.1 (1982): 71-87.

Laudan, Larry. "A Confutation of Convergent Realism." *Philosophy of Science* 48, no.1 (1981): 19–48.

Miller, Boaz. "What is Hacking's Argument for Entity Realism?" *Synthese* 193, no.3 (2016): 991-1006.

Musgrave, Allan. "Realism, Truth, and Objectivity." *In Realism and Anti-Realism in the Philosophy of Science*, edited by Robert S. Cohen, Risto Hilpenen, and Qui Ren-Zong, 19-44. Boston: Kluwer, 1996.

Putnam, Hilary. *Mathematics, Matter and Method*, Cambridge: Cambridge University Press, 1975.

Reiner, Richard and Pierson, Robert. "Hacking's Experimental Realism: An Untenable Middle Ground." *Philosophy of Science* 62, no.1 (1995): 60-69.

Resnik, David. B. "Hacking's Experimental Realism." *Canadian Journal of Philosophy* 24, no.3 (1994): 395-412.

Sankey, Howard. "Reference, Success and Entity Realism." *Kairos* 5, no.1 (2012): 31-42.

van Fraassen, Bas. *The Scientific Image*. Oxford: Oxford University Press, 1980.

Worrall, John. "Structural realism: The best of both worlds?" *Dialectica* 43, no.1-2 (1989): 99–124. 198

Trust, Reliance and Normativity; Explaining the Normative Dimension of Trust in Terms of Interpersonal Relationships

Johnny Hugill

Sidney Sussex College, Univeristy of Cambridge

Abstract

Trust is fundamental to our social lives. Our most important relationships involve rich and deep trust, and without trust, our relationships lose all meaning. Trust also seems to be a moral relation; trust obligations are included within the subset of obligations that we would properly characterise as moral, in some sense. More specifically, trust relations seem to be morally normative; they give us *moral* reasons for action. We have moral reasons not to betray the trust of a devoted friend or lover; when we do betray them, we are committing a (quite severe) moral transgression. This paper gives an explanation of the unique moral normative profile of trusting. It does so, specifically, in terms of the moral demands derivative of specific social relationships. That is, I claim that trust is constitutive of interpersonal relationships, and the moral demands of trust are themselves constitutive of the demands of these relationships. In this way, it moves away from reliance accounts of moral normativity, and towards an account that properly grounds the moral reasons of trust as interpersonal social reasons. Trusting is a social phenomenon; and its normativity is grounded quite clearly in the moral importance of social relations.

Introduction

Trust is fundamental to our social lives. Our most important relationships involve rich and deep trust; trust allows us to establish intimacy, warmth and love. Conversely, when our trust is betrayed, it hurts us so deeply and so profoundly. Without trust, we cannot have meaningful interpersonal relationships. It is a familiar story; when a friend or lover betrays our trust, it leaves us questioning the entire relationship itself. This is what leads the crestfallen lover to ask: 'Did I ever really know him?' A relationship without trust is no relationship at all; destroyed trust results in a destroyed relationship.

Given the importance of trust to our social lives, it is hardly surprising that trust and trust-betrayals are couched in moral terms. That is, there are moral norms governing how we should behave when others trust us, and when we trust others in return. If someone trusts in us, we think that we morally *ought* not to betray that trust, and we, ourselves, feel morally *wronged* when others betray our trust.

Moral concepts such as trusting are not always easy to understand. It is not immediately clear *why* betraying someone's trust is so morally problematic. This paper explores the nature and source of the normativity of trust. The first part of this

paper attempts to explain the *nature* of the normativity; in particular, it considers whether trust is a genuinely moral relation. Karen Jones (2017), for instance, has argued trust gives us non-moral reasons for action. Therefore, while trust is an important social relation, it should not be understood in moral terms. I argue that this approach is mistaken; I argue we have *moral* reasons to honour trust.

The rest of the paper is dedicated to attempting to properly *ground* this moral normativity. It attempts to give a systematic explanation as to why the particular social relation of trust is morally normative. I first consider attempts to ground this moral normativity in terms of a theory of reliance. I show, however, that such attempts fail. First, I claim that reliance does not generate the right kind of 'reactive attitudes' with respect to trust-betrayals. Second, I show that reliance, in fact, seems to be neither necessary nor sufficient for trusting.

Instead, I argue (following Domenicucci and Holton 2017) that trust should be thought of as, fundamentally, a two-place relation. This two-place relational structure of trust is important. The formal structure of trusting is, like love, two-place. This reveals a lot about why and how trusting is a moral social relation. The final section of my paper argues that the moral normativity of trusting is grounded in this formal two-place relational structure. Specifically, it is my claim that trust is constitutive of interpersonal relationships; such that the moral normativity of trust is grounded *in virtue of* these specific interpersonal relationships.

This is a theory of the moral normativity of trust that takes interpersonal relationships to be morally reason-giving. It is my claim that the moral demands of trusting are derivative of what we owe to each other in interpersonal relationships. Trust and its moral demands are specifically social interpersonal demands.

The Normative Dimension of Trust

In this section, I will examine the nature of the normative reasons given by trusting. First, I will clarify what is meant by the 'normative dimension' of trusting. Trust has a normative dimension insofar as the existence of trust relations give us reason for action. Other social relations work in the same way; canonically, promises and contracts give us reason to fulfil certain defined contractual obligations. This can be generalised to all sorts of social or interpersonal relations. Parenthood, friendship, and love all give us reason to fulfil certain particular tasks. To that extent, these distinct social relations have distinct normative profiles.

The first thing to note is that the normativity of trusting is not well evaluated as a contractual obligation. First, most trust relationships seem to resist being characterised in this way. A spouse's trust in her partner to be faithful is surely not grounded in any binding contractual agreement (it would be a damning indictment of the relationship if it were). Second, many trust relations are not chosen (but are, instead, entirely un-chosen) by at least one of the parties. This is not to say that trusting does not incur obligations; the point is simply that these obligations are not grounded contractually, as is the case with genuinely promissory obligations, or obligations of consent.

Indeed, the existence of a specific trust relation nonetheless gives us *reason* for action. Imagine you trust your friend with a secret; and, later, she divulges the information to another friend. When our trust is betrayed in this way, we tend to appeal to our friend: 'But I trusted you!'. The existence of a trust relation seems to be a

reason for action; the fact that you trusted your friend is *itself* a prescriptive reason for her behaving in a certain way (viz., keeping your secret). As such, when someone betrays our trust, what is at stake is an improper sensitivity to this reason. Hence, the implied content of the reminder: 'But I trusted you (and *that's why* you should not have done it)!'

Further, trust is a *pro tanto* reason insofar as we can have competing reasons for acting in accordance with someone's trust. If you trust your friend to tell a small lie on your behalf, to excuse your absence from a university seminar, the existence of this trust is, to that extent, a justificatory reason for her doing so; even if there are some other competing reasons against her lying (such as the general moral norm against lying, or some university-internal norm prohibiting giving false information to professors). As such, if our friend fails to lie on our behalf, while she might have avoided transgressing other moral (or non-moral) norms, it seems that she has *wronged* us in a distinctive way. This wrongness suggests that trusting involves a normative dimension.

Moral Normativity

In this section, I will clarify what *kind* of normativity is involved in trusting. Specifically, I hope to show that trusting is *morally* normative. I argue for this claim with reference to the 'reactive attitudes' that characterise trust, which are distinctly moral. This claim, as I outlined above, seems to be in line with our pre-theoretic intuitions. It does, however, need defending from a plausible explanation in terms of non-moral reasons.

Karen Jones (2017) argues in favour of a 'non-moralized account of the normative force' of trust; such that 'morally bad people can be genuinely trustworthy'.[1] The normative force of trust, according to Jones, is grounded in the non-moral utility of practice-internal trust norms. That is, trust norms are grounded in our interest in having cooperative social lives, insofar as trust serves the important social mechanism of allowing us 'to recruit the agency of others'.[2] Jones argues that trust norms must be non-moral because 'even the amoralist is able sincerely to make the complaint: "but I was counting on you!"'.[3] The 'textbook amoralists, the Mafioso', for instance, still take trust to be a *reason* for action; but insofar as they are amoral, the reason itself must be non-moral.[4]

Amoralist cases of this kind can be misleading. It is much more natural to describe even 'textbook amoralists' such as the Mafioso as immoral, rather than amoral. That is, they are immoral insofar as they typically put their own greedy self-interests above moral considerations. Perhaps trust is a moral reason that the Mafioso take seriously only because it tends to be conducive with their self-interests (of course, they only take trust to be normatively binding *within* the Mafioso). There is no mystery in immoral people incurring morally normative reasons for action. Immoral promises (or promises made by immoral people), for instance, are made frequently; if someone fails to fulfil such a promise, this is (in one way at least), a moral

---

[1] Karen Jones, "'But I Was Counting on You!'", in The Philosophy of Trust, eds. Paul Faulkner and Thomas Simpson (Oxford: OUP, 2017), 107.

[2] Jones, "'But I Was Counting on You!'", 102.

[3] Ibid, 107.

[4] Ibid, 107.

transgression. The same can be said for trust; if you are trusted to do something immoral, the trust itself (as I outlined above) is still a *pro tanto* moral reason for action.

More importantly, our attitudes towards trusting seem to show, quite decisively, that trust gives us specifically *moral* reasons for action. Many philosophers (following Holton 1994) have noted the importance of Strawsonian 'reactive attitudes' to relations of trust.[5] Usually the relevant reactive attitudes are feelings of blame, indignation or resentment. When we trust someone to do X, we are ready to feel certain reactive attitudes, should we discover that they have failed to do X. These reactive attitudes are distinctly moral; they are comparable with the attitudes that arise from other significant moral transgressions (such as lying, promise-breaking and stealing). Specifically, we reserve these attitudes for when we feel we have been *morally wronged*.

Indeed, these reactive attitudes are quite distinct from the feelings accompanying the transgression of non-moral norms. Take, for instance, the important non-moral norm of using the English language correctly. The practice-internal rules of English are important, in order to establish some shared basis for communication. If someone fails to cohere to the norms of proper English usage (even such that we could not communicate with them), our feeling is unlikely to be one of blame or indignation. We might, instead, feel frustrated that we cannot properly communicate, or perhaps disappointed that they failed to cohere to a norm that we, ourselves, faithfully respect.

Further, not only are the reactive attitudes of trusting distinctly *moral*, they also seem to be specifically *interpersonal*. That is, when our trust is broken, not only do we feel blame and resentment, we also feel a deep sense of betrayal. In this way, trust seems to have a distinct reactive profile that is most closely associated with profound emotional hurt. When someone, for instance, commits financial fraud, we might be morally indignant (and we certainly will blame them). When someone exploits our trust, however, we also feel personally hurt and betrayed. The reactive attitudes associated with trust reveal that trust gives us *moral* reasons for action; and more specifically, it gives us *interpersonal* moral reasons. The emotional attitudes associated with trust, therefore, are those that accompany betrayed personal relationships. This is why betrayals of trust hurt on such a personal level; because they concern specific people with whom we felt we shared a social connection.

I have characterised the normativity of trust as a type of interpersonal moral normativity. In the following sections, I will attempt to give a grounding for this morally normative dimension of trust. I begin by considering accounts that ground the normativity in terms of reliance.

Trust and Reliance

In this section, I will outline attempts to ground the moral normativity of trust in terms of reliance. I argue, however that such attempts fail.

Before proceeding, I will present a standard version of the reliance account of trusting. This account is more sophisticated than some reliance accounts (which imply that, *inter alia*, there is no difference between trusting an alarm clock and

---

[5]Richard Holton, "Deciding to Trust, Coming to Believe," Australasian Journal of Philosophy, 72(1) (1994): 63–76.

trusting one's friend). According to more sophisticated reliance accounts, the moral normativity of trust is grounded in terms of a specific reliance; such that: A (reasonably) relies on B with respect to task X, and B is aware of this reliance.

The condition of reasonability is clearly doing explanatory work; first, there seems to be an epistemic condition to reasonability; such that, it is epistemically reasonable for B to rely on A. Second, there is a practicality condition to reasonability; reliance cannot incur an obligation to perform a task, if that task cannot reasonably be performed.

How might reliance explain the moral normativity of trust? One way is by invoking some broadly contractualist framework (similarly to how Scanlon (1998) uses his 'contractually irresistible' 'Principle F' to ground the moral normativity of promissory obligation).[6] That is, we might appeal to some external contractualist moral principle grounding reliance, and then claim that trusting is analysable in terms of reliance; and, therefore, that its normativity is also grounded in this external moral principle. Further, we can imagine a possible rule-consequentialist grounding; that failed reliances (tend to) generate negative consequences, and so, there is a consequentially-justified moral principle prohibiting such failures.

Of course, the project of grounding the moral normativity in terms of reliance is convincing only insofar as the reliance account of trust seems feasible. I present two distinct objections to the reliance account of trusting: (i) reliance generates the wrong kind of reactive attitudes; and (ii) reliance is neither necessary nor sufficient for trusting. From these objections, I move (in §5) to the positive claim that trust is, fundamentally, a two-place relation.

The first objection, again, is due to the reactive attitudes of trusting. I agree with those (following Baier 1986 and Holton 1994) who think that the difference between trust and reliance is that trust involves certain reactive attitudes (outlined above) towards the person you are trusting.[7] Specifically, when you trust someone, you have a readiness to feel betrayal should you be disappointed. Betrayal, on the other hand, does not seem to accompany instances of failed reliance. Consider, for instance, the fact that we rely on online websites for information all the time; and that the authors of such websites are fully aware of this reliance. Notwithstanding, we hardly feel personally *betrayed* in cases of websites giving us false information. More likely, again, we might feel disappointed or frustrated.

The second objection is that reliance seems to be neither necessary nor sufficient for trusting. First, it does not seem necessary; imagine you have a spouse who you deeply trust (and on whom you can rely in many important ways). If she were to suddenly fall ill, such that she lost many of the abilities that informed this reliance, it could hardly be said that you would, thereby, stop trusting her. Rather, you would cease to rely on her, but continue, quite separately, to trust her.

Second, reliance does not seem sufficient; that is, we can rely on a person with respect to some specific task, without it being said that we *trust* them. Imagine the following example (modified from a case in D'Cruz 2015): you have employed an irresponsible babysitter, who you do not trust at all.[8] As a result, you decide to install

---

[6]'Contractually irresistible' is a phrase from Pettit's discussion of Scanlon's contractualism: Philip Pettit, "Two Construals of Scanlon's Contractualism," The Journal of Philosophy, 97 (2000): 149.

[7]Holton, "Deciding to Trust, Coming to Believe". Annette Baier, "Trust and Antitrust," Ethics, 96(2) (1999): 231–60.

[8]Jason D'Cruz, "Trust, Trustworthiness, and the Moral Consequence of Consistency," Jour-

a video camera in the house, and you tell her that you will observe how she behaves. In such a case, you can rely on the babysitter not to shirk her babysitting responsibilities, as she knows that, if she does, she will be caught. Despite this reliance, you certainly do not trust the babysitter; in fact, this is the exact reason why the camera is necessary.

Trust, I have argued, is more complete than reliance; we can rely on people to perform certain tasks without trusting them, and we can trust them even in the absence of a specific reliance. Trusting someone requires trusting them *qua* person; that is, trusting a specific person for who she is, rather than relying on her to perform a certain task. In the following section, I claim that trusting is, in fact, a two-place relation; and as such, any attempts to ground the normativity of trust must reflect this two-place structure.

Trust as a Two-Place Relation

The above considerations lead to the conclusion that what is important about trust is the interpersonal relationship between two people; that is, trusting a person *qua* person, and not in virtue of some task-dependence. In this section, I argue (with Domenicucci and Holton 2017) that trust is, fundamentally, a two-place relation.[9] As such, any attempt to locate the moral reason-giving power of trust must properly reflect trust's fundamental two-place structure.

A three-place relational analysis of trusting consists of two persons, A and B, standing in a three-place relation between themselves, and some task X, such that: A trusts B to do X, or A trusts B, with X, or perhaps, A trusts B, in the capacity of X. Instead, the two-place relational account, that I am proposing, holds that A trusts B *simpliciter*.

The first argument for a two-place relational account comes from how we typically use terms of 'growth', '(re)building' and 'establishing' to refer to trust relations. It is perfectly natural to say: 'It is important at this time to grow trust'. It is difficult to imagine the parallel construction with respect to reliance: 'It is important at this time to grow reliance'. Again, this is because the trust relation is fundamentally a two-place interpersonal relation; we surely do not *grow* trust with another person, with respect to some task X. Rather, we grow trust with them *simpliciter*. The reliance relation, however, cannot be analysed in such terms. We rely on another person with some task in mind; reliance is, fundamentally, three-place.

There are also strong linguistic arguments for the two-place relational account (for further examples, see Domenicucci and Holton).[10] One such argument comes from completeness; the trust relation 'A trusts B' is linguistically complete, in a way that 'A relies on B' is not. That is, it is perfectly natural to say: 'Anne trusts Bob', without any further questions being asked. The construction 'Anne relies on Bob', however, is incomplete; it crucially misses out *how* she relies on him. The point is, the construction 'A trusts B (*simpliciter*)' seems to be linguistically complete, while the construction 'A relies on B (*simpliciter*)' does not. This lends further *prima facie* support to the claim that trust is, fundamentally, a two-place relation (in a way that

nal of the American Philosophical Association, 1(3) (2015): 468.

[9]Jacopo Domenicucci and Richard Holton, "Trust as a Two-Place Relation," in The Philosophy of Trust, eds. Paul Faulkner and Thomas Simpson (Oxford: OUP, 2017), 149-160.

[10]Jacopo Domenicucci and Richard Holton, "Trust as a Two-Place Relation," 150-153.

reliance is not).

Trust and Relationships

So far, I have shown that; (i) reliance cannot explain the moral normativity of trust; and (ii) any attempt to locate the moral normativity of trust must do so in terms of trust's fundamental two-place structure.

In this section, I explain that trust gives moral reasons for action insofar as trust is constitutive of interpersonal relationships. That is, the normativity of trusting is grounded in the obligations that arise in virtue of being in particular interpersonal relationships. Such an account properly reflects trusting's fundamental two-place structure; as it captures the normativity of what specific people owe to each other qua persons, and not in terms of specific tasks. I offer two arguments for such an account of trusting, before showing how it might provide a grounding for the moral normativity of trust.

It is my claim that the moral normativity of trust is grounded in interpersonal relations. We have moral reasons not to do all kinds of things; to steal, to cheat, to lie, to damage property, etc. The moral reasons we have for trusting, however, are uniquely interpersonal insofar as they are sensitive to the particular interpersonal relationships we develop with others. They concern not just how we should treat people quite generally (as with most other moral reasons), but how we should treat people given the particular social relationships we have with them. I have moral reason not to betray the careful trust of my spouse *because* she is my spouse; and we share a committed interpersonal relationship. Trust is moral because of the relationships that ground it.

Aside from the conceptual neatness of such an account, there are three main arguments that support my case. First, my account explains why trust is 'rival' in a way that reliance is not (Markovits 2011 makes a similar point with respect to love and promising).[11] The formal pattern of reliance is nonrival; when someone relies on another person, it does not reduce her capacity for relying on others. Of course, there are constraints on how many people we can rely on (we can only keep track of a certain number of reliances). The important point, however, is that the formal structure of reliance is nonrival. Trusting, however, seems to be distinctly rival. Trusting is a discriminate relation; we choose who to trust for who they are *qua* persons. By trusting more people, we render our existing trust relations less discriminate, and, therefore, less important; to trust someone is to foreclose trusting others. On my account of trust, this is easily explained; trusting is constitutive of interpersonal relationships. Intimate relationships, of course, similarly depreciate in value as they increase in number. We cannot *properly* trust a large number of people, insofar as we cannot *properly* share intimate interpersonal relationships with a large number of people.

In this way, trust is like love. Both trust and love involve serious interpersonal commitments to another person. They both involve opening oneself up to some, and closing oneself off from others. You cannot properly and truly love an extremely large number of people; to love someone involves choosing *them* over others. Similarly, when we trust someone, we trust *them*; we share a uniquely intimate interpersonal

---

[11] Daniel Markovits, "Promise as an Arm's-Length Relation," in Promises and Agreements, ed. Hanoch Sheinman (Oxford: OUP, 2011), 295-326.

relationship with them.

Second (and, surely, relatedly), my account can easily explain the reactive attitudes associated with trusting. Above, we saw that trust-transgressions are characterised by feelings of deep betrayal. My account can explain this neatly. A failed reliance disappoints one's reasonable expectations; this can be enormously disruptive and frustrating. It does not, however, leave a person questioning the very relationship itself. Betrayed trust is different in this respect. It leaves the betrayed truster feeling as if she misunderstood the trustee, in a way that calls the relationship itself into question. When our trust is betrayed, we often feel that the relationship was an illusion, (hence the accompanying saying: 'Well, I guess I never really knew him'). The reason for such a response is that trust is constitutive of relationships, so failing to take someone's trust seriously constitutes failing to take the relationship seriously; this is what is so hurtful about trust betrayals. Betrayals of intimate interpersonal relationships hurt us; they leave us feeling, amongst other things, upset, embarrassed, self-doubting and confused. Allowing ourselves to vulnerable with another person is important for proper social relations. Of course, vulnerability can easily be betrayed. This is why trusting hurts so much when it goes wrong.

Third, it should be further added that betrayals of trust affect our ability to develop proper interpersonal relationships in the future. That is, once our trust has been badly betrayed (or, perhaps, systematically betrayed), it leaves us feeling like we cannot develop proper relationships with anyone. Imagine the scorned lover, who has had to ensure a series of trust betrayals over her life. With each betrayal, her ability to form new relationships becomes strained. Insofar as she cannot properly trust people as she once could, she can no longer develop the kind of meaningful, intimate relationships that she once could either. A capacity to properly trust is so central to developing proper interpersonal relations, insofar as trust is constitutive of these relations.

My account, therefore, holds that the moral normativity of trust is grounded in the specific interpersonal relationships we share with others. It is so morally wrong to betray the trust of another because we share an intimate and unique relationship with them. My account neatly explains trusting's rival formal structure, the unique reactive attitudes associated with trusting, and the connection between a capacity to trust and a capacity to develop social relationships.

It might be objected at this stage that we trust people even in the absence of any specific interpersonal relationship. And, further, that betrayals of trust, even when there is no such relationship, still constitute moral transgressions. I admit that such cases would, indeed, be problematic for my account of the moral normativity of trust. I am tempted to deny that there are any such cases. A canonical example of such a case would be how we trust strangers with small tasks in our daily lives; we might, for instance trust a stranger in coffee shop to watch our laptop while we go to the bathroom. Or we might trust the barista not to spit in our coffee when we are not looking. In both cases, we trust a stranger, in some specific domain, and they have purportedly moral reason not to betray this trust. I think that these kinds of cases are better analysed in terms of contractual agreements, rather than trust relations. We would not trust a stranger to watch our laptop unless we asked them to do so, and they confirmed that they would (viz., they promised, or agreed, that they would watch our laptop). In such a case, it is not trust doing the normative work, but rather a specific contractual agreement with another person. Again, our trust in the barista is better analysed in terms of other contractual agreements. In this case, the barista

is under a contractual obligation to do her job properly, which means, *inter alia*, not spitting in my coffee. It is the existence of this contractual obligation that gives her normative reason not to spit in my coffee, not any trust relation that we might share.

Indeed, it is true that we trust almost everyone we come across, at least in some limited way. This usually involves negative trust relations; trusting others *not* to do particular things. We trust strangers not to do all kinds of things that might inconvenience us in our daily lives. A lot of these domains will be covered by legal obligations. These, like the specific contractual obligations above, do not require much explanation. Our trust in strangers not to assault us, steal from us, abuse us (or commit any crime against us) is derivative of their normative obligation not to break the law. Those that are not covered by contractual obligations (legal, professional, promissory, or otherwise) will be derivative of specific interpersonal relations. How else could someone have moral reason for action, in the absence of such contractual obligations, other than what they owe to a specific person *qua* person?

The 'stranger' cases examined above, I claim, pose no particular threat to my theory of trust. We are, however, left with the following problem: how, exactly, does my account explain the moral reason-giving power of trust? We have seen how the trust relation is constitutive of relationships that are defined 'personally rather than juridically'.[12] This grounds a specific type of interpersonal moral normativity; it generates commitment rights and obligations between the trusting parties. That is, in virtue of standing in particular interpersonal relationships, we are conferred a status of special interpersonal intimacy. We have an obligation to show commitment to this intimacy, and a corresponding right to expect it in return.

We might explain this interpersonal moral normativity in terms of some moralised conception of 'normatively defined relationships'.[13] That is, we might think that we have a moral obligation to honour particular types of interpersonal relationships; and insofar as trust is constitutive of these relationships, betrayals of trust constitute betrayals of these relationships. Further, we could give an explanation in terms of 'normative interests'; that the moral normativity of trust is grounded in the fact that we have a particular normative interest in certain types of relationships (specifically, meaningful interpersonal relationships).[14] Alternatively, my analysis also seems to be amenable to a contractualist explanation; that trust norms are morally grounded in what we could reasonably expect of each other. That is, that there exist contractually-justified moral norms governing what we owe to each other in interpersonal relationships; trust-norms would be an example of such norms. This would be an interesting result, insofar as the reliance account (which I rejected above) is the typical way of explaining a contractualist grounding of trust.

It might appear disappointing that my paper cannot provide a decisive answer to these questions. Indeed, the project of this paper is to find out why trust relations are morally normative. I have provided an explanation for the specific moral profile of trusting. How moral reasons are properly normative requires an examination of moral normative frameworks quite generally. The benefit of my account is that it can be accommodated into (amongst others) consequentialist, contractualist and

---

[12]Stephen Darwall, "Trust as a Second-Personal Attitude (of the Heart)," in The Philosophy of Trust, eds. Paul Faulkner and Thomas Simpson (Oxford: OUP, 2017), 46.

[13]Cogley, Zac Cogley, "Trust and the Trickster Problem," Analytic Philosophy, 53(1) (2012), 31.

[14]Seana Shiffrin, "Promising, Conventialism and Intimate Relationships," Philosophical Review, 117 (2008): 481–524.

interest-based normative frameworks. The question of how exactly the uniquely interpersonal moral profile of trust gives moral reasons will then be settled according to whichever is the correct theory of moral normativity, quite generally.

Trust and Domain-Specificity

My account, as I have shown, properly captures many of our important pre-theoretic judgements about trust; it also properly explains why trust gives us moral reasons for action. Despite this, it seems to be susceptible to substantial objections from domain-specificity.

First, two-place accounts of trust (not just *my* particular account, but two-place accounts quite generally) seem to be susceptible to the following problem; that some trust relations can, in fact, be domain-specific. That is, it is clearly the case that we trust some people more than others in certain domains. Put another way, trust is not an all-or-nothing relation. The extension of things I would trust my spouse to do is greater than the extension of things I would trust my plumber to do. We typically trust our spouses with our secrets, our insecurities, our finances (that is, we trust them in highly vulnerable ways).

I argue that we should not explain the differential nature of trust in terms of domain-specificity. Rather, trust is differential insofar as we trust people *more* than others; that is, trust is two-place, but it is a stronger relation in some instances than others. We trust some people richly and deeply (usually our family, best friends and partners); those with whom we have the strongest interpersonal relationships. We also trust people with whom we have less intimate relationships (colleagues, professors, classmates); but we trust them, therefore, to a much lesser extent. Finally, we do not trust (or, perhaps, barely trust) strangers, or people with whom we have no established relationship. The most natural way of explaining this connection between trust and intimacy is that trust is constitutive of interpersonal relationships. Consider the following claim: 'Mary is much closer friend of mine than Anne, but I trust Anne more'. Such a claim strikes us as immediately unnatural; this is because, again, trust is *constitutive* of intimate relationships. We can explain the differential nature of my trusting the plumber and my trusting my spouse *because* we share different kinds of interpersonal relationships.

There is, however, a more problematic objection from domain-specificity; the phenomenon of specific exclusions of trust. It is very common to richly trust someone, but still exclude some specific domains from that trust relation. For instance, it is perfectly natural to say, of your friend: 'I trust her, but not with my secrets'. This seems to be a problem, because we might have friends with whom we are less close (and so, we trust less), who we, nonetheless, *do* trust more in this particular domain.

I argue that we can explain particular exclusions of trust by distinguishing between the commitment and competence conditions of trusting. That is, when we trust someone, we are invoking both their *commitment* to the relationship, and their *competence* to properly reflect this commitment. For instance, if we trust someone not to share a particular secret, we are invoking both their commitment to our specific relationship, but also their competence (or ability) to properly respect what my invoking demands. Indeed, it seems that commitment transgressions are more morally condemnable than competence transgressions. Above, I explained the reactive attitude of betrayal in terms of one's unilateral lack of commitment to a certain

interpersonal relationship. In instances of failed competence, however, we do not tend to have the corresponding reaction of betrayal (in fact, we only do when one's lack of competence becomes symptomatic of a lack of commitment – insofar as their inability or unwillingness to improve a given incompetence reveals a structural lack of commitment).

As such, in our secret-telling case, if we want to maintain that our friend is trustworthy, we need to make excuses for her unreliability in this domain. The excuse that we make in instances of particular exclusions is that their unreliability in this domain does not reflect a lack of commitment to the relationship, but rather, is merely an incompetence. With respect to a betrayal of trust, taking the competence stance saves the relationship, while taking the commitment stance, *pro tanto*, destroys the relationship. When taking the competence stance, we might say: 'Anne is really trustworthy, but (through no lack of commitment to our relationship), she has a habit of divulging secrets when she gets too excited at parties'. The more systematically isolated the exclusion is, the more likely it is to be a competence failure. That is, we would expect a structural lack of commitment to a relationship to come out in other ways than just particular, isolated exclusions.

I have argued that we can understand specific exclusions in trust in terms of competence, rather than commitment. It is lack of commitment, and not competence (unless this competence, itself, indicates a lack of commitment) that generates the reactive attitude of betrayal. As such, we can trust someone richly, and still allow for specific incompetences, without jeopardising the relationship.

Conclusion

I have presented an account that grounds the moral normativity of trust as social normativity, in terms of interpersonal relationships; that is, I have argued that trust is constitutive of interpersonal relationships, and so the demands of trust are derivative of the demands of these relationships. The moral profile of trusting, on my account, is distinctly interpersonal. I argue for this conclusion for a number of important reasons. First, I argue for this by showing that any proper account of trust must take seriously trust's fundamental two-place structure. My account of trust as constitutive of interpersonal relationships properly does this; and, further, it explains very neatly many of the conceptual and attitudinal features unique to trusting. Second, it explains the rival nature of trusting (as opposed to reliance); how trusting specific people forecloses us from trusting others. Third, it explains why betrayals of trust generate such a sense of profound and personal hurt; because they are, fundamentally, social dependencies. Finally, it explains why a lack of a capacity to trust results, necessarily, in a lack of a capacity to form proper relationships; insofar as trust is constitutive of interpersonal relationships. I have presented a theory of the moral normativity trust, therefore, that; (i) properly respects the formal structure of trust, (ii) reflects our pre-theoretic intuitions about trust, and (iii) gives an account of how the moral normativity of trusting could be accommodated into a general normative framework. Trust is an interpersonal relation; the moral normativity of trust is derivative of this interpersonality; and the importance of these social relations to our lives.

References

Baier, Annette. (1986). Trust and Antitrust. *Ethics*, 96(2), 231–60.

Cogley, Zac. (2012). Trust and the Trickster Problem. *Analytic Philosophy*, 53(1), 30–47.

Darwall, Stephen. (2017). Trust as a Second-Personal Attitude (of the Heart)'. In Paul Faulkner and Thomas Simpson (Eds.), *The Philosophy of Trust* (OUP), pp. 35-50.

D'Cruz, Jason. (2015). Trust, Trustworthiness, and the Moral Consequence of Consistency'. *Journal of the American Philosophical Association*, 1(3), 467-484.

Domenicucci, Jacopo and Holton, Richard. (2017). Trust as a Two-Place Relation'. In Paul Faulkner and Thomas Simpson (Eds.), *The Philosophy of Trust* (Oxford: OUP), pp. 149-160.

Holton, Richard. (1994). Deciding to Trust, Coming to Believe'. *Australasian Journal of Philosophy*, 72(1), 63–76.

Jones, Karen. (2017). But I Was Counting on You!'. In Paul Faulkner and Thomas Simpson (Eds.), *The Philosophy of Trust* (Oxford: OUP), pp. 90-108.

Markovitz, Daniel. (2011). Promise as an Arm's-Length Relation'. In Hanoch Sheinman (Ed.), *Promises and Agreements* (Oxford: OUP), pp. 295-326.

Pettit, Philip. (2000). Two Construals of Scanlon's Contractualism'. *The Journal of Philosophy*, 97, 148–64.

Scanlon, T.M. (1998). *What We Owe to Each Other* (Cambridge, MA: Harvard University Press).

Shiffrin, Seana (2008). Promising, Conventialism and Intimate Relationships'. *Philosophical Review*, 117, 481–524.

The Gender of Consent

Paula Keller
Univeristy of Cambridge

Abstract

I argue that common opinion about consent to sex has significant sexist tendencies. I illustrate this by drawing on typical speech about consent in consent workshops, philosophical literature and in the legal definition of rape. I argue further that this sexist opinion is not only objectionable in itself but harms: It establishes a norm of acquiescence to sex which in turn is made possible by a disparity inherent in the structure of how consent functions. I take this to suggest that consent to sex is unfit as a progressive moral standard.

Introduction

Mutual consent is regarded as the main criterion (and more specifically as a necessity) for morally and legally unobjectionable sex (Dougherty 2013, 719) and is contrasted with rape, sex in the absence of consent.[1] I will focus on whether consent really is a morally unobjectionable standard, particularly in regards to gender equality. I argue that consent is gendered, reinforces the oppression of women and that consent as a moral standard ought to be rejected. "Gendered" will mean this: consent entails specific roles and expectations for one gender, women, which it does not for others and so contributes to gender inequality.

I will first define valid consent and then will examine how and why such consent might be gendered: In section one I examine consent giving, where patriarchy and oppression influence consenting behaviour. In section two I look at how the structure of consent - someone giving and another receiving consent - perpetuates gender inequality. In section three I briefly explore alternative proposals and conclude.

By "gendered" I understand that gender matters during a particular action, for one gender there are expectations to conform to a specific role which in this particular fashion do not exist for other genders. If consent is gendered in patriarchal society, it comes with task divisions between genders where women occupy inferior and men dominant roles. I mean "gendered" in the sense of "care taking is gendered" where care taking is seen as something particularly feminine. I argue that consent - if important at all - is particularly a woman's job, consent is female.

Consent is "moral[ly] transformative" (Hund 1996, 124) making a previously wrong action right. Take Ann wanting to ride Paul's bike. If Ann simply takes Paul's bike to get to the supermarket, it is morally objectionable: Ann is stealing Paul's bike. If Paul grants Ann permission, Ann is allowed to ride the bike to the supermarket because Paul has consented. I will call the role Ann is in, demanding

---

[1]Note that this contrast can come apart: e.g. intercourse with children per se is rape (S. 5-8, Sexual Offences Act 2003).

something from Paul, consent receiver[2], and the part Paul plays, granting permission to Ann, consent giver.

Imagine three different cases: (1) Ann threatens Paul with a knife to give her the bike, (2) Paul is a small child, or (3) Paul falsely understands "Can I look at your bike". In these cases we would not say that Paul has consented to Ann riding Paul's bike to the supermarket. In (1) the consent is *forced* or given under threat, in (2) Paul as a toddler is not *competent* to consent and in (3) Paul is not *informed*, does not understand the meaning of Ann's sentence and consents to something different from what Ann wants. To call something consent we want it to be voluntary (not coerced or forced), competent and informed (Dougherty 2016b). When speaking of consent below, I will mean valid consent fulfilling these three criteria. There is much more to be said about these criteria, but for my discussion below it will suffice to have a rough, intuitive idea.

Women as Consent Givers

I argue that common opinion regards the man as the active consent-seeking and the woman as the passive consent-giving part. While claims about discourse might be difficult to make forceful, I draw on a range of examples to at least suggest that the conversations we have about consent are perpetuating heteronormativity and female submissiveness.

Consent giving, where "moral magic" (Hund 1996, 121) happens, that is where a morally impermissible act is rendered permissible, typically is seen, in the case of sex, as the woman's job. If women aren't objects incapable of consent, then they are consent givers, while men are constructed as demanders of consent. Men are seen as the more sexual gender while women have a lower sex drive: men are "unable to control their sexual urges" (Foley 1995, 8). Such myths, as Lorber asserts, are internalised: "boys learn their version of seduction and girls learn their version of romantic love", role divisions become normalised and assign the active and initiating role to men (Lorber 1994, 70). When consent itself becomes the subject of public opinion, it is viewed in generally heterosexual terms with a male consent receiver, the active part, and a female - passive - consent giver. The following examples illustrate this.

Rape law in the UK defines "rape" as a crime done by a person with a penis: "A person (A) commits an offence if— (a) he intentionally penetrates the vagina, anus or mouth of another person (B) with his penis, (b) B does not consent to the penetration, and (c) A does not reasonably believe that B consents" (S.1 (1), Sexual Offences Act 2003); all (a), (b) and (c) are necessary conditions. While the gendered pronoun "he" is a generic masculine, "rape" is qua definition an offence only open to those with a penis: women cannot rape (that is initiate sex and override the partner's refusal), only be raped by others.

Examples from philosophical literature on consent amplify these stereotypes. Heidi Hund considers reasons for why someone might consent to sex and equates women to victims and men to defendants:

---

[2]Ann is also "consent demander" in this case. I however use "consent receiver" since receiving is necessary so that the receiver is allowed to do something she was not allowed to do before.

> ...*he* need only intend to cause a contact. (...) *she* need only intend a
> contact.(...)—the fact that *she* hopes the contact will yield a promotion,
> or improve *her* singing voice, or cure *her* cancer (...). Since the defendant
> does not have to intend the consequences of *his* contact, the victim does
> not have to intend them either. (Hund 1996, 134, my italics)[3]

Conly in "Seduction, Rape and Coercion" uses heterosexual Alec d'Urberville raping
Tess as main example and continues to speak of a "woman's saying 'no' to sex" (2004,
101): "woman" equals "consent giver".

Consent workshops at university often suggest the same heteronormativity and
role division. Designed to teach the importance of consent they use examples such
as "[a] guy I was dancing with began dancing really close" or "one guy (...) pulled
down my pants and was touching me" (University of Bristol guide 2015). This guide
uses three examples, all of which involve men initiating sex, perpetuating the idea
of sexually active and demanding men. The common feature is: heterosexual sex,
man initiates, woman consents.

This wide-spread image affects how women consent and refuse consent and how
men perceive such consent or refusal. I will show that women's consent often is ac-
quiescence to power; the woman tolerates and consents to sex while she does not
want nor welcome it.

Let Paul be the person wanting sex. Ann consents. However, Ann's consent can-
not be equated with willingness, she would rather not have sex. Expectations and
implicit power inequalities make Ann consent when Ann would much rather watch a
movie.[4] It looks as though we can easily construct a similar case in which Paul is the
consenting part and Ann wants sex. Paul might consent because he wants pleasure
for Ann, because he generally likes sex with Ann even if he does not actively want
sex at that moment: Paul's reasons are personal. While Ann's reasons can plausibly
be the same, Ann has further reasons to consent which Paul has not: her subordinate
position in patriarchy, she does what is expected of her as a woman, following a norm
of passive submissiveness.[5] This is what MacKinnon calls consenting by "acquies-
cence" (MacKinnon 2016, 463). She argues that "women are socialized to passive
receptivity; [they] may have or perceive no alternative to acquiescence" (MacKinnon
1989b, 177). Patriarchal norms influence a woman's consent giving: she agrees to
sex in situations where she consciously and knowingly does not want sex.

One might object that such consent as described does not meet the valid consent
conditions specified above, specifically that this consent is not voluntary consent,
Ann is pressured into submissiveness or forced to comply by societal norms.

If one allows norms to count as prohibiting voluntariness in this case, the argu-
ment that we in fact never act voluntarily because norms constantly influence our
actions comes into sight and it seems rather difficult to argue why acting according

---

[3]I cite this only to highlight the gendered pronouns used by Hund, its content is secondary.

[4]Robin Morgan summarises this thought: "How many millions of times have women had
sex 'willingly' with men they didn't want to have sex with? ... How many times have women
wished just to sleep instead or read or watch the Late Show?" (Morgan 1977, 165-6)

[5]There seem to be three broad categories of reasons to consent: (1) personal reasons such as
Paul's, not influenced by power structures, (2) expectations of a woman's role in sex, the case
illustrated here, (3) instrumental considerations related to gender or power inequality such as
"to stop other abuse and do their best to make it sexy so that it will end quickly" (MacKinnon
2016, 465). I will leave (3) aside here as I take it to be a sub-case of (2) also involving power
inequalities.

to some norms makes my actions involuntary and acting according to others does not. It seems much more plausible to claim that Ann's actions are in fact voluntary, she assessed her possibilities and chose to act conforming to oppressive social norms. Such a minimal account of voluntariness (where no explicit force or threat is involved) preserves female agency under oppressive conditions and avoids a too radical conclusion that female consent can never be valid consent. The problem here is not that a condition for consent is not met, but rather that (valid) consent itself perpetuates female oppression.

Two examples clarify my point and answer possible objections:
(1) Paul wants to refuse but then consents to sex with Ann because she is his boss. He thinks (but cannot say for sure) that he might lose his job if he does not comply. We want to say that the sex between Paul and Ann was not morally unobjectionable. Does this not suggest that power inequalities, rather then gender, distort consent?

Power inequalities do distort consent but Ann's consent is furthermore normalised and eroticised. MacKinnon asserts the influence of gender norms: "submission eroticized defines femininity" (MacKinnon 1989a, 318). Submission is the sexual standard for women regardless of their partners' feminist attitude or attention to women's desires. What is more, female submissiveness is sexy. Paul's consent does not have this additional dimension.
(2) Sam is Paul's dentist. Sam tells Paul that Paul needs a tooth operation; knowing that such an operation will be painful Paul is hesitant. However, he eventually allows Sam to remove his tooth. His motivation is a higher order interest that in future his teeth be white and shiny which seems completely morally unobjectionable - not like the cases above.

So what makes the difference? First, the consent to the dentist (unlike Ann's consent) is independent of worries about gender inequality, Sam and Paul are both men. One might however argue that there still is a power inequality between the dentist and his patient which influences Paul's decision to have the operation. But Paul agrees to the removal of his tooth independently and because of his own rational considerations that the operation best leads to health. These considerations are supported by Sam's expertise and presuppose that Sam is trustworthy and speaks the truth. Paul therefore agrees to the operation relying on Sam's expertise, not because agreement or submissiveness is expected of him as in the case of Ann and sex.

I have shown that patriarchal society, particularly public opinion and norms about consent and sex, influences women's consenting to which allows me to conclude that consent is gendered. I used a cases in which a female's choice can be oppressed in ways in which a male's choice cannot: a woman's spoken "yes" - expected and eroticised - does not signal willingness but endurance or acquiescence.

One might argue that due to sex myths, heteronormativity and the novelty of discussion about consent, oppression generated by ways we talk about (or forget to talk about) consent can be overlooked. Consent as a concept is simply too novel to judge it by how misogyny influences it, we should not be too hard on it. The debate only started in the 1980s and 1990s, when consent did not legally matter in marriage (Kennedy 1993, 112) and rape required the presence of physical force (West 1996, 233).

One might further claim that a misogynist "tone" does not make consent itself objectionable, that patriarchy - not consent - is to be blamed for its misogynist connotations.

While both these objections might have something to be said for them, they are

irrelevant to the fact that ways we talk about consent harm: they perpetuate a standard of female submissiveness during sex. I argued that the typical picture of consent involves heterosexual partners, initiating man and consent giving woman. Even if this might well be an accurate representation of the majority of cases of sexual consent, it cements female submissiveness and inferiority in sex. As Langton asserts: "to say something is to do something" (Langton 1993, 295), to say "the woman consented to the man" is to create a standard of passive women consenting to active men.

I now argue that these objections are also mistaken since consent is in particularly suited to lend itself to gender inequality, more suited than other standards such as willingness or welcomeness. Consent therefore can be blamed for its misogynist connotations.

Gendered Consent

The arguments above suggest that consent is gendered in an unequal society: public opinion, if it does not forget the importance of consent, gives a woman the passive consent giving role in which her refusal might be eroticised or she might only tolerate sex. However, a general worry about the concept of consent arises independently of societal power structures. I here point to the unequal structure inherent in consent and explain that this structure perpetuates inequalities already present in the practices one consents to and so makes acquiescence possible.

Recall that Ann wanted to borrow Paul's bike: Ann wanted Paul to agree, Paul was the consent giver and Ann the consent receiver. Consent as a concept requires these two roles which in sex in patriarchy are distributed to different genders. While the gender dimension is due to structures of a particular society, a disparity between active, demanding role and passive, (dis)agreeing role is inherent in consent. This makes consent lend itself to inequality and leaves it unfit as a standard in sex if the desideratum is mutual agreement.

Consent is contrasted with assent: while consent describes an "agreement to" (Archard 1998, 5), a toleration of something, assent describes an "agreement with" (Archard 1998, 5), an independent willingness. "Seeking consent is essentially asking 'Is it all right if I do this *to* you?'" (MacKinnon 2016, 465, my italics); an interaction in which only one partner decides and the other agrees. This does not suggest that the concept of consent itself is gendered, but it shows that its structure is not set up for equality. Structural disparity is what makes consent gendered in a gendered society.

A defence of consent might be that we usually require mutual consent in sex: for consent to be valid both partners need to consent which would erase worries concerning inequality. Borrowing a bike is an unequal interaction, Ann has to actively demand something from Paul, Paul does not want anything from Ann. In sex consent is different: Both Ann and Paul want something which they can only get through collaboration. Consenting to consensual sex might be compared to consenting to a game of chess. Both partners need to consent to playing chess, otherwise neither Ann nor Paul will be able to play chess with each other.

However, even if we add this requirement for mutuality in consent, an inequality still exists due to a difference between consenting and assenting, willing. In the case of playing chess, someone needs to initiate the game of chess. Paul needs to come up

to Ann and say "Hey, how about some chess?", otherwise there is nothing Ann can consent to. For mutual consent Ann might then say: "Sure, I'm in - are you?" Paul: "Yes".[6] While consent also allows for the possibility that Ann and Paul initiate chess simultaneously, the structure of mutual consent makes it *possible* that only one initiates and both consent (which does not require willingness from both sides): it carries the *potential* for inequality. Consent allows for someone to agree to something that person does not want.

Paul expressed that he *wants* to play chess otherwise he would not have asked. Initiating implies wanting, so it seems. Surely Ann could equally be willing to play chess but Ann does not *have* to be willing. The chess example can be applied in just the same way to Ann and Paul having sex. Mutuality in consent does not make consent equal as the inequality arises out of initiating the conversation over the game of chess/sex act, asking the question and a connected willingness to play chess/have sex. Mutuality requires both Ann and Paul to be consent giver and consent receiver, but only one to initiate or want. "Consent must always be given *to* something" (Pateman 1980, 164) mutual consent must also be given to the person who initiated, consent therefore "is not an equal mode" (MacKinnon 2016, 465). It follows:

P1: (Mutual) consent is an unequal interaction which allows for an active and a passive part.
P2: In a patriarchal society men are the dominating gender and women the dominated.
C: In a patriarchal society, men tend to occupy the dominating, active role and women the passive in consent.

Does this suggest that in consenting to a game of chess, men tend to equally be the dominating, women the dominated in such a way as during sex? This seems absurd. In sex the unequal structure of consent makes consent unequal because sex already is: heterosexual sex is a gendered game. Chess is not socially constructed such that it entails gender roles: both women and men play chess following the same rules. C holds only for consent to gendered acts and practices such as sex or other acts with involve power inequality, e.g. master-slave relationships. This suggests that one can raise more general objections to consent as a moral standard: it might generally be problematic in scenarios that involve any sort of power inequality.

While in an unequal society the unequal standard of consent perpetuates, not counteracts, (gender) inequality (C), even in an equal society consent entails a disparity due to its "agreement to"-structure, its contrast with willingness and its requirement for an initiator (P1). However, this structure in itself is not objectionable or unjust. Between equals one's acquiescence to sex might well be unobjectionable, rather objectionable patriarchal structures that often explain why women consent make a consent standard in patriarchy objectionable. Sex is something we might call a "high-stakes" interaction (Dougherty 2016a), involving an infringement on another's personal, intimate sphere and it is therefore all the more alarming that present legal and moral standards tolerate that women are pressured into having sex.

---

[6]Such an explicit conversation does not occur before most sex, but seems implied in the partners' behaviour.

A Non-gendered Standard

I argued that consent is unsuited as the main standard for morally unobjectionable sex for a society with power inequalities, particularly gender inequalities. They impact heterosexual sex most prominently and disadvantage consent giving women due to the unequal structure of consent. I will now briefly explore proposals for other standards that could replace consent. I look at two suggestions MacKinnon makes and favourably discuss a welcomeness standard but object to defining rape in terms of coercion only.

MacKinnon proposes: "Rape Redefined" is "a physical invasion of a sexual nature under circumstances of threat or use of force, fraud, coercion, abduction, or of the abuse of power, trust, or a position of dependency or vulnerability" (MacKinnon 2016, 474). However, this new definition might be suitable for rape but seems to be a regress when wanting to define morally unobjectionable sex. It leaves out female agency entirely and does not solve the problem of acquiescence.

My interest is different than MacKinnon's: I am not after a legal definition of rape but rather explore what, instead of consent, could make sex morally permissible. For this purpose, especially in light of society's tendency to objectify of women, emphasising the willingness of both seems more fit to reach a standard that promotes (gender) equality. MacKinnon's definition with a focus on coercion looks at one agent only, the one who coerces. This approach might be suitable in legal investigations into rape as it lowers the burden of proof. Proving something that was present is easier than to prove something that was not (MacKinnon 2016, 452). However, this approach seems to give a counter-productive signal for gender equality in sex.

MacKinnon sets out analysing acquiescence (see 2016, 440). However, her proposal falls short of tackling this problem. MacKinnon specifies: "for a criminal conviction, it would be necessary to show the exploitation of inequalities—their direct use—not merely the fact that they contextually existed" (2016, 470). Ann in the example above who acquiesces to power is not coerced, nor is power abused by Paul when she consents. Inequalities are not *directly used* by Paul but they are present, even though Paul might be unaware of them. MacKinnon might be right that Paul did not rape in this case, but female acquiescence to sex remains morally objectionable as it harms women, pressuring them to do something they would otherwise (e.g. in an equal society) not do. The absence of coercion as a standard for morally unobjectionable sex does therefore not seem to be much better in ensuring gender equality than consent.

MacKinnon also speaks of desire, mutuality or willingness as moral and legal standards in sex. However, take Lucy and Paul: Lucy would like to watch a movie but then agrees to have sex with Paul even though she does not desire or want to have sex but rather because she likes Paul a lot and she is happy to give Paul pleasure. We might call her want to have sex with Paul a second-order willingness that overrides the absence of her desire or her first-order want not to have sex. This scenario seems rather common, is (by stipulation) independent of gender or other power inequalities and is nothing we would want to object to on moral grounds. Desire, mutuality or willingness are therefore too stringent a standard in sex.

We are left with welcomeness, a standard used in sexual harassment law in the United States. MacKinnon claims that welcomeness is more consistent with equality than consent: The woman who acquiesces to sex might voluntarily consent, but

she does not welcome sex. Her yes is generated by power inequalities and is voluntary only "'in the sense that the complainant was not forced to participate against her will'" (MacKinnon 2016, 451). Lucy from the example above on the other hand does welcome sex, she is fine with it, even if she does not enthusiastically await it. Welcomeness as a standard therefore seems neither too exclusionary nor too broad. While a resilient definition of what exactly counts as welcome sex still needs to be specified, welcomeness, at least from this exploratory, brief consideration, looks as though it faces less gender related obstacles than consent does.

Problems might however arise when we ask about sex that is not welcome but where consent was given for other reasons than power inequalities or when we think more generally about what standard we should adopt when power inequalities are absent. Should we still demand a welcomeness standard even though the main argument for such a standard was motivated by worries about power inequality or should we demand a welcomeness standard only in cases in which power inequality plays a role and how should these cases be singled out?

It goes beyond the scope of this essay to further explore positive proposals. I flagged up some which however all come with more or less severe obstacles. However, I conclusively argued that severe objections can be found against the current main standard for morally permissible sex: Consent entails and perpetuates gender inequality and in itself is grounded in disparity. This influences women's consent giving or refusing in morally objectionable ways. I conclude that consent is not fit to make sex morally unobjectionable. If we value a "standard of mutuality" in sex (MacKinnon 2016, 467), my arguments above show that consent is never a suitable standard in a society with massive power inequalities. The often unquestionable standard of consent to sex therefore needs to be questioned. While consent might be progress compared to normalising rape, the implications it bears for gender equality require that further progress be made.

References

Archard, David (1998), *Sexual Consent*, especially ch.1.

Austin, John, (1962), *How to Do Things with Words* (Oxford University Press)

Conly, Sarah (2004), "Seduction, Rape, and Coercion", in *Ethics*, 115, 1, 96-121.

Dougherty, Tom (2013), "Sex, Lies and Consent", in *Ethics* (*123*(4): 717-744.

——— (2016a), "Communicating Consent", Lecture on Consent at Amoral Sciences Club (Cambridge).

——— (2016b),"Intoxication and Consent", Lecture at Trinity Hall Philosophy Society (Cambridge).

Dripps, Donald (1992), 'Beyond Rape: An Essay on the Difference Between the Presence of Force and the Absence of Consent', in *Columbia Law Review*, 92, 1780–1806.

Foley, Linda, et al. (1995), 'Date Rape: Effects of Race of Assailant and Victim and Gender of Sub- jects on Perceptions', *Journal of Black Psychology*, 21, 6-18. Hund, Heidi (1996), 'The moral magic of consent', in *Legal Theory*, 2, 121-146. Kennedy, Helena (1992), *Eve Was Framed* (London: Chatto and Windus), ch. 5 'Asking for it'.

Langton, Rae (1993), 'Speech Acts and Unspeakable Acts', in *Philosophy and Public Affairs*, 22, 305-330.

——— (1995), 'Sexual Solipsism', in *Philosophical Topics*, 23, 181-219.

——— (1998), 'Subordination, Silence and Pornography's Authority', in ed. R. Post, *Censorship and Silencing*, pp.261-83.

Lorber, Judith (1994), *Paradoxes of Gender* (Yale University Press).

MacKinnon, Catharine (1987), *Feminism Unmodified: Discourses on Life and Law* (Harvard University Press).

——— (1989a), 'Sexuality, Pornography, and Method: "Pleasure Under Patriarchy"', in *Ethics*, 99, 314–346.

——— (1989b), *Toward a Feminist Theory of the State* (Cambridge, MA: Harvard University Press).

——— (2016), 'Rape Redefined', in *Harvard Law and Policy Review*, 10, 431-477.

Maitra, Ishani and McGowan, Mary Kate (2012), *Speech and Harm: Controversies over free speech* (Oxford University Press).

Morgan Robin (1977), *Going Too Far: The Personal Chronicle of a Feminist* (New York: Random House).

Pateman, Carole (1980), 'Women and Consent', in *Political Theory*, 8:2, 149–68.

Punts, gowns, and sexual consent: introducing consent workshops to freshers week, CUSU Women's Campaign guide, 2014

Sexual Consent Workshop, Participant Handbook, Welcome Week 2015, University of Bristol

Sexual Offences Act 2003, British Stature Law

Schulhofer, Stephen (1998), 'Unwanted Sex', in *Atlantic Monthly*, 282, 4, 55-66.

West, Robin (1996), 'A Comment on Consent, Sex, and Rape', in *Legal Theory*, 2, 233–251.

——— (2010), 'Sex, Law and Consent', in F.G. Miller and A. Wertheimer, eds., *The Ethics of Consent: Theory and Practice*, pp. 221- 250.

Beyond a Reasonable Doubt: An Analytic Reconstruction

Jared I. Mayer

Johns Hopkins Univeristy

Abstract

In the Anglo-American legal tradition, defendants in criminal trials are convicted of a crime if and only if the evidence of their performing the crime is proven "beyond a reasonable doubt." While this standard is certainly banal – anyone who has watched an episode of *Law and Order* knows about the standard – its meaning remains notably elusive. After all, how can one clearly differentiate between those sets of evidence which establish guilt "beyond a reasonable doubt" and those that do not? More fundamentally, what does "reasonable" even mean?

In this essay, I attempt to investigate some possible answers to these questions and more. In Part I, I present a background of the "beyond a reasonable doubt" standard and where the main questions posed above arise. Part II attempts to construct some linguistic and epistemic explanations for the standard. I argue that "reasonable doubt" is highly contextually sensitive such that we are more likely than not to confuse and frustrate ourselves in attempting to define the standard *outside* of the context of a particular case. Addressing some other concerns relevant to the standard leads me to Part III, where I provide a Wittgensteinian account of relevance that immediately relates to the "beyond a reasonable doubt" standard. Part IV discusses how the standard interplays with the structure of courtroom deliberation. Part V summarizes and reflects on the findings in this essay.

I. The Criminal Trial[1]

The jury's decision must be the product of a deliberation that considers whether the defendant is guilty "beyond a reasonable doubt."[2] This standard has strong and historical roots in the Anglo-American legal tradition; indeed, Sir William Blackstone famously declared that "it is better that ten guilty persons escape, than that one innocent suffer."[3] And while the United States Supreme Court has remained

---

[1]It may seem at first that the discussion of the criminal trial procedures is an unnecessary digression from the important philosophical matters at hand. But as well shall see later on, the *highly formalized* environment in which the "beyond a reasonable doubt" standard takes place makes all the difference.

[2]My discussion here about the standard is heavily indebted to Ronald J. Allen, Richard B. Kuhns, and Elenor Swift, *Evidence: Text, Cases, and Problems*, Second ed. (New York: Aspen Publishers, Inc., 1997), 885-890.

[3]Blackstone, William. "On Trial, and Conviction" Bk. 4, Ch. 27, *Commentaries on the Laws of England*. Lonang.com. Lonang Institute, 1765. Accessed. 17 Dec. 2016.

largely silent on the matter, "beyond a reasonable doubt" has retained its well-entrenched and unquestioned status as the sole standard that satisfies the Due Process Clauses of the Fifth and Fourteenth Amendments.[4]

Yet ever since the "beyond a reasonable doubt" standard has been employed, especially in the American context, there has been a great debate over just what the standard means. This debate is further augmented by the fact that courts must convey this standard to jury members. So dogged was the Supreme Court with the question of the standard's meaning that it flatly proclaimed, "Attempts to explain the term 'reasonable doubt' do not usually result in making it any clearer to the minds of the jury."[5]

Consider, for example, the following jury instruction defining "beyond a reasonable doubt":

> 'Reasonable doubt' is such a doubt as would cause a reasonable and prudent person, in one of the graver and more important transactions of life, to pause and hesitate before taking the represented facts as true and relying and acting thereon. It is such a doubt as will not permit you, after full, fair, and impartial consideration of all the evidence, to have an abiding conviction, to a moral certainty, of the guilt of the accused. At the same time, absolute or mathematical certainty is not required. You may be convinced of the truth of a fact beyond a reasonable doubt and yet be fully aware that possibly you may be mistaken. You may find an accused guilty upon the strong probabilities of the case, provided such probabilities are strong enough to exclude any doubt of his guilt that is reasonable. A reasonable doubt is an actual and substantial doubt arising from the evidence, from the facts or circumstances shown by the evidence, or from the lack of evidence on the part of the state, as distinguished from a doubt arising from mere possibility, from bare imagination or from fanciful conjecture.[6]

All that we have is that a "reasonable doubt" is one that arises in relation to the evidence (or the lack thereof) and nothing else.[7] This is what we may call the "relevant

---

[4]The earliest case in the U.S. Supreme Court discussing the standard was in 1880, where the Court with unquestioning precision declared: "The evidence upon which a jury is justified in returning a verdict of guilty must be sufficient to produce a conviction of guilt, to the exclusion of all reasonable doubt." *Miles v. United States*, 103 U.S. 304, 312 (1880) (per Woods, J.). Later on, the Court provided three reasons for upholding this standard of conviction in criminal cases: a) possible loss of liberty (which we are unwilling to take lightly), b) maintenance of the individual's sense of respect from the community of citizens in which he lives, and c) maintenance of the sense of respect the community itself has for the rule of law and the criminal justice system. *In re Winship*, 397 U.S. 358, 363-364 (1970) (per Brennan, J.).

[5]*Miles, supra*, at 312. This sense of frustration with the standard has yet to be abandoned by the legal community, as we shall soon see.

[6]Allen, Kuhns, and Swift, ibid. At 887-888. This standard was upheld by the Supreme Court as constitutional under the demands of the Due Process Clause. See *Victor v. Nebraska*, 511 U.S. 1 (1994) (per O'Connor, J.).

[7]In this way, the standard as presented here is reminiscent of Newton's Fourth Rule in his discussion of how to test for physical phenomena: "In experimental philosophy we are to look upon propositions inferred by general induction from phenomena as accurately or very nearly true, not withstanding any contrary hypothesis that may be imagined, till such time as other phenomena occur, by which they may either be made more accurate, or liable to exceptions."

factors" portion of the jury instruction. But there are other considerations in the jury instructions that need further elaboration. Take, for example, what we may call the "disposition" portion of the jury instruction: "'Reasonable doubt' is such a doubt as would cause a reasonable and prudent person, in one of the graver and more important transactions of life, to pause and hesitate before taking the represented facts as true and relying and acting thereon." The purpose of this passage, one may presume, is to remind the jurors that they must adopt a particular, solemn disposition when adjudicating cases, one that is disinterested in the outcome of the case. In this way, jurors are supposed to conduct themselves much like Adam Smith's "impartial spectator."[8] But even here, reasonable doubts are those that *reasonable* people, given a sober, deliberate disposition, would find troubling given the evidence. What is more, the jury instructions also state that one may convict the defendant on a probability, provided that it is strong enough to "exclude any doubt of his guilt that is reasonable." This is circular, plain and simple. We have not yet escaped the indescribability of "reasonable."

And can we escape it? "Beyond a reasonable doubt," of course, remains incredibly vague, even when contrasted with the notion of "mathematical certainty." When one achieves a mathematical level of certainty, it is logically *impossible* for someone to be incorrect. But with reasonableness, it is not immediately clear what would prove sufficient; the addition of the term "moral certainty" does nothing to clarify the distinction, if not further obfuscates the line between reasonable and mathematical doubts.[9] These reflective hesitations regarding the "beyond a reasonable doubt" standard might just bring us to an epistemic standstill.

II. "Beyond a Reasonable Doubt" and Its Defenders

In reviewing the "beyond a reasonable doubt" jury instruction, we noted that despite attempting to further clarify the notion of a "reasonable doubt," the standard could not help but use "reasonable" (and corollaries thereof) in its definition. And even when contrasted with a standard like mathematical certainty, it was not clear what would be sufficient to meet the "beyond a reasonable doubt" standard. But this inability to further express what we mean by "reasonable" is not a defect of the standard, but indeed one of its primary features.

To see this, let us consider a distinction that J.L. Austin makes between words and sentences on the one hand, and statements on the other. Words and sentences *without more* cannot be said to be true or false. This is because they are uttered without any context.[10] So if I were to ask, "Is "red" true?" we cannot say whether it

---

Isaac Newton, "Principia Mathematica Book 3" in *Science Rules: A Historical Introduction to Scientific Methods*, ed. Peter Achinstein (Baltimore: Johns Hopkins University Press, 2004), 80.

[8]Adam Smith, *The Theory of Moral Sentiments*, ed. Ryan Patrick Hanley (London: Penguin Books, 2009), 133-136.

[9]As we shall see later, this contrast will prove critical in further understanding "reasonableness."

[10]J.L. Austin, "Truth," in *Philosophical Papers*, Third Ed., ed. by J.O. Urmson and G.J. Warnock (New York: Oxford University Press, 1979), 119. I understand "context" in a broad sense; that is, it refers to all of the conditions to which a sentence refers and in which it is made.

is true or false simply because there is no context to which the question is attached. The same holds true with sentences. Suppose I were to say, "Evaluate the truth of the following: "The gravitational constant is 9.81 meters per second squared." As a sentence, this, too, would defy any label of either truth or falsity, for a sentence – a mere arrangement of words in a given sequence – can only make sense given some understanding of the conditions under which it is said. Indeed, it is questionable whether these kinds of questions are even intelligible without such a context. After all, is the red *what* true? The Gravitational Constant *where*?

What can be evaluated as true or false, however, are statements. Statements are "the utterance[s] by a certain speaker or writer of certain words (a sentence) to an audience with reference to an historical situation, event or what not."[11] Statements are, of course, comprised of sentences and words. But they are also comprised of, or perhaps better understood as "tethered by," the context in which they are made. This means that the sentences and words used to make the statements must relate to those factors that are formally external to the statement itself.[12] To further understand the implications of this, we must follow Austin in making a distinction between descriptive and demonstrative conventions. Descriptive conventions correlate "the words (=sentences) with the *types* of situation, thing, event, c., to be found in the world."[13] Demonstrative conventions, by contrast, correlate "the words (=statements) with the *historic* situations, c., to be found in the world." Austin then provides a clear understanding of what makes a statement true: "A statement is said to be true when the historic state of affairs to which it is correlated by the demonstrative conventions. . . is of a type with which the sentence used in making it is correlated by the descriptive conventions."[14] So for a statement to be true, the statement in question must sufficiently be like other, similar states of affairs such that the same description can be used for both this statement and the other states of affairs.[15]

We may now return to the notion of "beyond a reasonable doubt." Recall that the trouble we had with this standard was twofold: first, we could not help but avoid using "reasonable" in our definition or explication of the standard, moving nowhere with respect to understanding just what "reasonableness" demands of us; and second, we could not avoid the ambiguity that arose between mathematical and reasonable (or moral) certainty. Austin's analysis suggests a solution for both problems. Austin's heavy emphasis on context suggests that by insisting on defining "reasonable doubt" in the abstract, we have stripped ourselves of the very tools we need to make it intelligible. In this sense, "reasonable doubt" would only really make sense – indeed, we can only know what would constitute a doubt – in the context of a case

---

[11]Ibid. At 119-120.

[12]By formally external, I mean external to the very semantics of the statement. So the context in which the question is being asked is *formally* external to the statement, despite being indispensible for the intelligibility and truth evaluation of the statement.

[13]Ibid. At 121-122 (emphases in original). As Austin admits, the sentences and words we use often times serve in both kinds of conventions, even sometimes simultaneously. See Austin, ibid. At 122n.

[14]Ibid. At 122.

[15]Ibid. At 122nn. So to say the statement must sufficiently relate the item in question with a set of other red things such that all of these items can, by the descriptive convention, be called red. Ibid. General laws of nature (e.g. laws and constants in physics), too, seem to pass this bar with flying colors. Not only are the laws of nature sufficiently like other states of affairs, but rather they are (nearly) exactly alike. 9.81 meters per second2 is the same here and there, today and tomorrow.

wherein a whole variety of doubts may arise.[16] So we cannot simply ask "What are reasonable doubts?," but rather, "What are reasonable doubts *within this particular case*?"[17] And indeed, in evaluating the truth of this statement, we must compare this case with like cases and determine whether this case is sufficiently similar to previous cases such that it merits (or perhaps does not merit) the claim of truthfulness. Thus, in order for any "reasonable doubt" claim to have meaning, it must be within a particular context, which itself must be contextualized, as it were, in a constellation of like cases.

But if the standard is so case-dependent and contextually bound, why ought we create a standard that presses us, as we have seen with Austin above, to compare it to cases elsewhere? H.L.A. Hart answers this question by noting that in cases like "reasonable doubt," "it is impossible to identify a class of specific actions to be uniformly done or forborne and to make them the subject of a simple rule, yet the range of circumstances, though very varied, covers familiar features of common experience."[18] Hart here understands certain standards, such as "reasonable," as something akin to range properties. That is, "reasonable" is a single property or attribute that we want to allow for the inevitable, particular differences that will arise in individual cases. While it may be tempting for us to abandon the "reasonable doubt" standard in favor of a *rule* – that is, in favor of a standard whose sufficient conditions are easily and plainly satisfied – this is to content oneself with the strong possibility that one will restrict the law to those particular cases for which the rule is calibrated. As such, the law would be unable to accommodate future, unforeseen cases.[19] The aim, therefore, is to provide a single, regulatory principle under which particular aims can be achieved while allowing for the inevitable plurality and diversity of future cases.

One more consideration must be addressed before we can proceed. While we have partly given an answer to the question of the distinction between "moral" and "mathematical" certainty, more needs to be said with respect to what is needed to distinguish between them. At the heart of this distinction lies the notion of possibility. Philosophically speaking, possibility can be taken in a variety of different senses: logical, metaphysical, real, and so forth. What mathematical certainty wants to secure is a level of epistemic security such that it precludes not only real, but also logical and metaphysical possibility. When spelled out in this fashion, it becomes clear that such a standard is somewhat than chimerical. But as Stephen Toulmin points out, we rarely – perhaps only in particular, mathematically and formally-logically oriented areas of inquiry – do we hold ourselves to that standard. Rather, "to speak of a particular suggestion as a *possibility* is to concede that it has a *right*

---

[16]Note, once again, that skeptical doubts are wholly removed here because they are not doubts that originate from within the context of the case. The jury instruction we saw said as much. See, *supra*, at 2.

[17]"Within this particular case" would of course be a placeholder, as it were, for a whole array of different conditions, facts, and so on that would give texture to the question itself.

[18]H.L.A. Hart, *The Concept of Law*, Third Ed., ed. Leslie Green, Penelope A. Bulloch, and Joseph Raz (Oxford: Oxford University Press, 2012), 132.

[19]Ibid. At 129-130. An example of a rule, in this sense, is the age restriction provided for in the following statute: "No person under the age of 21 may purchase alcoholic beverages." Even here, a rule may break down with the advent of some new technology or circumstance that is not clearly encapsulated by the law, e.g. a form of vaporized alcohol that can be consumed but is not in any sense a "beverage."

to be considered."[20] Certain possibilities, by virtue of the facts of the cases, can, as it were, vouch for our consideration of them by virtue of being self-evidently relevant; others require more discerning minds to notice their (perhaps) more concealed, and (perhaps) important, relevance to a particular case. Either way, not all possibilities have this *right* to command our attention. Rather, only select possibilities, ones that are shaped by the evidence at hand, need to be considered.

But rather than solving a problem, I have just introduced a new one. How can we possibly discern, with any sense of objectivity or consistency, which possibilities are worth our consideration and which ones may be dismissed? And if this cannot be solved, won't this jeopardize not only the "reasonable doubt" standard, but criminal trials more generally? As I shall next argue, the "reasonable doubt" standard heavily interplays with the other elements of the criminal trial. Many, if not most, of these elements are included in the formal trial institutions, but others are simply part of our ordinary linguistic and epistemic frameworks. As a combined, coherent framework of inquiry, these institutions provide us with the answer to the above question as well as a full picture of how the "reasonable doubt" operates.

III. Riverbeds and Courtrooms

As we noted above, one of the problems we now face is the question of which doubts we ought to exclude from or include in our deliberations. It is important to note, however, that this question is at one and the same time gripping and trivial. For it grips us when it is pointed out to us that, right now, we have no principled way of discerning which possibilities we should admit as relevant and which ones we should antecedently dismiss. Concurrently, however, we feel that the question is ridiculous, perhaps even bordering on the idiotic – of course we know how to discern which possibilities are the proper ones and which ones are not! We find ourselves in something of a dialectic, running between complete certainty and crushing skepticism.

Fortunately, there is an avenue for escape. Ludwig Wittgenstein claims that certain doubts are not merely unwarranted (though they are that, too), but also *unintelligible*. He says in a famous passage:

> From [a thing's] seeming to me – or to everyone – to be so, it doesn't follow that it is so. What we can ask is whether it makes sense to doubt it. If e.g. someone says "I don't know if there's a hand here" he might be told "Look closer". – This possibility of satisfying oneself is part of the language-game. Is one of its essential features.[21]

To take part in a collective endeavor, such as inquiring, comes with terms and conditions. Among them is adherence to what Wittgenstein calls a "language-game." Such a game is not like the rules antecedently laid out in a written constitution or charter. Rather, it is a set of practices that we, as inquirers and respondents who interact with one another, adopt in our epistemic-linguistic activities. These practices themselves determine the limits of our epistemic-linguistic activities and the kinds

---

[20]Stephen E. Toulmin, *The Uses of Argument*, Updated Ed. (New York: Cambridge University Press, 2008), 17 (emphasis added).

[21]Ludwig Wittgenstein, *On Certainty*, ed. G.E.M. Anscombe and G.H. von Wright, trans. Denis Paul and G.E.M. Anscombe (New York: Harper Row Publishers, Inc., 1972), 2e (§2-3) (emphases in original).

of inquiry are properly included within those practices. Among those excluded kinds of inquiry are those that ask us questions about possibility without providing sufficient reasons for how we could have gone wrong.[22] Notice how Wittgenstein does not declare these questions to be impossible in any sense. Rather, he says that in our language game, one need not respect skeptical questions that do not point to any concrete error (or possible source of error). And since inquiry and response only work within a language game, the skeptic cannot escape from our grasp.[23] The skeptical question is thus rendered null and void.[24]

But perhaps not. Perhaps we can say that the skeptic operates using his own language game, one that we fail to properly adhere to. So, in fact, it is *we* who are, relative to his language game, unable to live up to his standards. Wittgenstein defuses this situation with the following example. Suppose you have people who have claimed that they have gone to the moon.[25] In responding to them, we can point out how such a feat is not possible; we can employ the laws of physics, we can ask them questions relating to *just how* they achieved this feat, and so forth. Since they are unable to sufficiently answer these questions according to the demands of our language game, we need not take their claims seriously. In other words, to adhere to their beliefs requires them to ignore much of what we know to be true and for them to instead, at best, adopt *ad hoc* hypotheses and explanations. Even worse, they may attempt to counterattack with a blanket response, such as "you can't explain everything." Since they have failed to adhere to our language game, we need not take them as epistemic-linguistic partners or interlocutors. They have simply failed the test.[26]

These considerations have begun to illustrate for us the confines within which reasoned, intelligible argumentation can take place. Our language game is such that we do not investigate, indeed we do not learn, in an atomistic fashion, whereby each proposition is investigated or learned one after the other. Rather, "[w]hen we first begin to *believe* anything, what we believe is not a single proposition, it is a whole system of propositions. (Light dawns gradually over the whole system)."[27] Each link of the system, then, is deeply interconnected with the other links; you cannot have one without the others. But what justifies the system as a whole? Why I am entitled to use it? Wittgenstein responds, "... I did not get my picture of the world by satisfying myself of its correctness; nor do I have it because I am satisfied of its correctness. No: it is the inherited background against which I distinguish between true and false."[28] We cannot shed our language game for the sake of adopting another one *precisely* because it is this background on which all of our other epistemic and linguistic commitments rely. And even if we imagined for a moment that could do just that, since we cannot discard one commitment without making changes to the others, our entire system of understanding the world would have to be surrendered. Such a Herculean effort is impossible, not to mention unintelligible.

Then how do questions arise? If our game is such that we can exclude certain

---

[22]Michael Williams, "Wittgenstein and Skepticism: Illusory Doubts," (Paper Draft, Johns Hopkins University, 2016), 9-10.

[23]Wittgenstein, ibid. At 5e (§24).

[24]And so Wittgenstein muses, "What is 'learning a rule'? – *This*. What is 'making a mistake in applying it'? – *This*. . . ." Ibid. At 6e (§28). See also ibid. At 15e (§95).

[25]Wittgenstein wrote this well before the Apollo missions took place.

[26]Williams, ibid. At 32, 37.

[27]Wittgenstein, ibid. At 21e (§141) (emphasis in original).

[28]Ibid. At 15e (§95).

kinds of interlocutors and questions that they ask, by virtue of what may we say that other interlocutors are valid ones, ones with which our language game is willing to accommodate? To this, Wittgenstein responds in his famous and somewhat lengthy "river bed" passage:

> It might be imagined that some propositions, of the form of empirical propositions, were hardened and functioned as channels for such empirical propositions as were not hardened but fluid; and that this relation altered with time, in that fluid propositions hardened, and hard ones became fluid. The mythology may change back into a state of flux, the river-bed of thoughts may shift. But I distinguish between the movement of the waters on the river-bed and the shift of the bed itself; though there is not a sharp division of the one from the other.[29]

So the inherited background that we mentioned just above can, in fact, change. But it doesn't just change willy-nilly; there must rather be some notable force that, as it were, causes the river-bed to turn over. The process of altering our fundamental commitments requires us, as we mentioned earlier, not only to alter the single commitment in question, but also a whole host of auxiliary fundamental commitments, too. And even when this does occur, never does it occur to the *entire* river-bed; certain fundamental commitments remain. So the process of alteration and refitting of fundamental commitments is a slow and gradual one.[30] Non-fundamental commitments, however, can be altered or discarded far more easily. They are, as it were, the movement of the waters on the river-bed. But notice how these commitments only make sense in light of our fundamental ones; the waters of the river can only flow provided there is a river-bed beneath them. In this way, less fundamental commitments can be changed rather quickly, for to alter them imposes little alteration cost on our other commitments.

Now we may return to our discussion of "beyond a reasonable doubt." Recall that in order for any proposition that we may have to come into question, we must be able to find a way to reckon with the shift that will occur in the absence of those changes. Since the shifts that will occur with changes in our non-fundamental commitments are easier to reckon with than with those that will occur after changes in our fundamental commitments, our non-fundamental commitments are (relatively more) expendable. But notice that whatever the change may be, there must be something else upon which our commitment shifting relies. That is to say, in altering or abandoning a commitment, we must rely upon the set of other commitments that we have. Wittgenstein makes this point clear with another metaphor, the famous "hinge passage":

> [T]he *questions* we raise and our *doubts* depend upon the fact that some propositions are exempt from doubt, are as it were like hinges on which those turn. That is to say, it belongs to the logic of our scientific investigations that certain propositions are *in deed* not doubted. But it isn't that the situation is like this: We just *can't* investigate everything, and for that reason we are forced to rest content with assumption. If I want

---

[29]Ibid. At 15e (§96-97)

[30]As Wittgenstein himself notes, the line between fundamental and non-fundamental commitments is certainly not a bright one. There may even be some gradation between these two "kinds" of commitments.

the door to turn, the hinges must stay put.[31]

If indeed we wish to know anything new, or in fact to merely alter prior commitments, then we must take a whole set of commitments as givens. Much like the door without the hinge, we will be left simply unable to perform our task at hand without some epistemic-linguistic items held constant. So going back to the courtroom, if we wish to investigate the particular crime in question, we cannot let everything we possibly know be "up for grabs." In fact, very little of what we know *can* be put "up for grabs." This is not because we will become distracted or unfocused on the indictment at hand (though that may be so as well); rather, if every epistemic-linguistic commitment were jeopardized, we would simply not be having a trial at all. At best, we would be having an epistemic-linguistic circus. Such a prospect makes Joseph K.'s trial in Kafka's *The Trial* look like a paradigm of the rule of law.

Such considerations aren't merely the inventions of some very clever philosophers of the 20th century. Courts themselves recognize, and heavily enforce, this standard. At nearly every stage of presenting an argument, particular arguments, pieces of evidence, or even witnesses can be challenged and/or dismissed as irrelevant.[32] Indeed, witnesses perhaps receive the most stringent treatment with respect to relevancy.[33] Juries, too, are heavily regulated with respect to the content they hear such that no inadmissible content – that is, no content that may tamper with the jury's perception of one side or the other. This is because this information is squarely irrelevant for the case at hand; criminal trials are not popularity contests. So when jurors are instructed to deliberate whether the defendant's guilt is proven "beyond a reasonable doubt," part of what the court is doing (albeit implicitly) is ensuring that the only issue before the jury and the court is the matter of the indictment. Surely, a motion to dismiss the case on the grounds that counsel's client is a product of the evil deceiver would be overruled, and counsel disbarred. In this way, then, courts ensure that far-off doubts cannot even *arise* within the courtroom setting.[34]

IV. The Structure of Jury Deliberation

Now that we have settled the issues that arose with the "beyond a reasonable doubt" standard as such, we may now proceed to discuss the nature of jury reasoning. Some, like John Wisdom and Neil MacCormick, have argued as follows:

---

[31]Wittgenstein, 44e (§341-343) (emphases in original).

[32]For evidence generally, see Fed. R. Evid. 401; for witnesses, see Fed. R. Evid. 611(b).

[33]Given our discussion thus far, it is difficult to provide a formal account for why certain witnesses will be relevant rather than others. And while it is tempting to invoke it, possible world semantics will prove to exclude worlds that we find to be quite relevant while including those we think are easily dismissable. See Edward Craig, *Knowledge and the State of Nature* (New York: Clarendon Press of Oxford University Press, 1990), 18-23. It may be better to simply posit some "Property X," as Craig puts it, that, given the particulars of a case, makes the witnesses relevant. Ibid. At 24-26.

[34]"Far-off" doubts do not always remain "far-off." Over time, such doubts may become increasingly salient by virtue of their evidence coming to light. So to invoke the smudge of one's fingers as evidence of a crime in the 17th century would indeed be far off; obviously not so today. But many years of biological, physiological, and criminological research was devoted to the study of fingerprints. While the riverbed may shift, it only does so gradually.

> In [legal] cases we notice that the process of argument is not a *chain* of demonstrative reasoning. It is a presenting and re-presenting of those features of the case which *severally co-operate* in favour of the conclusion.... The reasons are like the legs of a chair, not the links of a chain.... [T]he procedure resembles scientific argument in that the reasoning is not *vertically* extensive but *horizontally* extensive – it is a matter of the cumulative effect of several independent premises, not of the repeated transformation of one or two. And because the premises are severally inconclusive the process of deciding the issue becomes a matter of weighing the cumulative effect of one group of severally inconclusive items against the cumulative effect of another group of severally inconclusive items.[35]

This account presses us to accept the notion that each piece of evidence is to be admitted independently, followed by the jury, with some kind of calculation, deducing its verdict based on balancing and weighing the *separate* factors from the two different accounts of what occurred. But this may not be so. Wisdom notably downplays the importance of the connection between pieces of evidence. To be sure, each piece of evidence plays an important role in any given account of what transpired in a case. But these pieces of evidence only have their epistemic and narrative force by virtue of *cohering* with other pieces of information. If they did not cohere properly, then the pieces of evidence would serve no use for the counsel that employs them. Wisdom's account, to use William Whewell's metaphor, provides the pearls yet is missing (or underplays) the string that connects them.[36]

To illustrate this point, let us take an example. Suppose you have a defendant who is indicted for murder. The prosecution attempts to prove, beyond a reasonable doubt, that the defendant indeed committed the murder. So they introduce to the court a whole series of testimonial evidence (witness who heard the gunshot, ballistics expert who knows about the operation of firearms and their ammunition, and so on) and non-testimonial evidence (the gun found at the crime scene, the victim's splattered blood on the defendant's jeans, etc.). The very force behind these presentations is not in the items and witnesses themselves but rather in how the attorneys unite the pieces of evidence into a single, coherent, and compelling account of what occurred. The defense, of course, can form its own set of evidence that attempts to directly compete with that of the prosecution. But this is not necessary, for the minimum the defense (hypothetically) needs to do is undermine the coherence of – that is, increase the level of doubt surrounding – the prosecution's account. By cutting the string that unites the prosecution's pearls, the defense can have those pearls come tumbling to the ground without having to produce a pearl necklace of its own. This, to be sure, is a juncture at which criminal trials are divorced from ordinary epistemic contexts. In ordinary epistemic contexts, the ability to reduce the coherence of an opponent's account in no way entails victory. Given the adversarial criminal justice system and formal bias in favor of defendants, however, all the defense needs to do is undermine the prosecution's account such that there exists a "reasonable

---

[35]John Wisdom, "Gods," *Proceedings of the Aristotelian Society* 45 (1944-1945): 194 (emphases in original). See also Neil MacCormick, "Reasonableness and Objectivity," Notre Dame Law Review 74, no. 5 (1999): 1596.

[36]William Whewell, "The Philosophy of Inductive Sciences," in *Science Rules: A Historical Introduction to Scientific Methods*, ed. Peter Achinstein (Baltimore: Johns Hopkins University Press, 2004), 150-151.

doubt." "Reasonable doubt," then, interplays not only with the particular facts of the case (as we mentioned), but also with how counsel from each side presents, bolsters, or undermines the coherence of the story that is presented by the prosecution.

V. Summary and Conclusion

Let us now bring it all together. "Beyond a reasonable doubt," as we have seen, is only really intelligible in the context of a particular case. This itself is not a drawback of the standard, but rather one of its virtues; it allows for there to be a rough standard by which all cases are adjudicated while accounting for the complexities and unforeseen circumstances that will arise in new cases. A "reasonable doubt" is one whose possibility has a right to be considered. And indeed, we saw that this, too, is highly contextually bound. Even within these proper contexts, not every epistemic item is up for grabs. If we wish to have any semblance of a trial, many things – arguably infinitely many – must be taken as givens, if only implicitly. So a reasonable doubt is one whose possibility has a right to be considered in light of the interplay of arguments between the prosecution and defense, one that, again, will be highly contextually bound.

This analysis is not, at least on my reading of it, a normative argument, one that would instruct courts how they ought to act. That would require much more analysis than I have presented here and thus is work for another day. Hopefully, by clearing the obfuscation and conceptual brush, as it were, this task will soon become possible.

References

Allen, Ronald J., Kuhns, Richard B., and Elanor Swift. *Evidence: Text, Cases, and Problems*. Second Edition. New York: Aspen Publishers, Inc., 1997.

Austin, J.L. "Truth." In *Philosophical Papers*. Third Edition. Edited by J.O. Urmson and G.J. Warnock. New York: Clarendon Press of Oxford University Press, 1979.

Craig, Edward. *Knowledge and the State of Nature*. New York: Clarendon Press of Oxford University Press, 1990.

Hart, H.L.A. *The Concept of Law*. Third Edition. Edited by Leslie Green, Penelope A. Bulloch, and Joseph Raz. Oxford: Clarendon Press of Oxford University Press, 2012.

In re Winship, 397 U.S. 358 (1970) (per Brennan, J.)

MacCormick, Neil. "Reasonableness and Objectivity." Notre Dame Law Review 74, no. 5 (1999): 1575-1603. Accessed December 18th, 2016.

Miles v. United States, 103 U.S. 304 (1880) (per Woods, J.)

Newton, Isaac. "Principia Mathematica, Book 3." In *Science Rules: A Historical Introduction to Scientific Methods*, edited by Peter Achinstein, 78-80. Baltimore: Johns Hopkins University Press, 2004.

Smith, Adam. *A Theory of Moral Sentiments*. Edited by Ryan Patrick Hanley. London: Penguin Books, 2009.

Toulmin, Stephen E. *The Uses of Argument*. Updated Edition. New York: Cambridge University Press, 2003.

United States House of Representatives Committee on the Judiciary. *Federal Rules of Criminal Procedure*. Washington, D.C.: U.S. Government Publishing Office, 2015.

United States House of Representatives Committee on the Judiciary. *Federal Rules of Evidence*. Washington, D.C.: U.S. Government Publishing Office, 2015.

Victor v. Nebraska, 511 U.S. 1 (1994) (per O'Connor, J.)

Whewell, William. "The Philosophy of Inductive Sciences." In *Science Rules: A Historical Introduction to Scientific Methods*, edited by Peter Achinstein, 150-169. Baltimore: Johns Hopkins University Press, 2004.

Williams, Michael. "Wittgenstein and Skepticism: Illusory Doubts." Paper Draft, Johns Hopkins University, 2016.

Wisdom, John. "Gods." *Proceedings of the Aristotelian Society* 45 (1944-1945): 185-206.

Wittgenstein, Ludwig. *On Certainty*, Edited by G.E.M. Anscombe G.H. von Wright. Translated by Denis Paul and G.E.M. Anscombe. New York: Harper Row Publishers, 1972.

When Some Speak Too Loudly: Reimagining the Relationship between Political Equality and Free Speech

Mohit Mookim
Stanford Univeristy

Abstract

This paper wrestles with the tension between two political ideals we deeply value—political equality and free speech. Observing vastly unequal influence in our political system, contemporary egalitarian political philosophers like Joshua Cohen have proposed a principle of equal opportunity for political influence. Commentators like Ryan Pevnick expose a dilemma that these egalitarian philosophers are caught in: either the principle is not demanding enough to sufficiently curb inequalities, or the principle is strong enough to quash our civil liberties. In this paper, I carve out a space for a comprise principle that balances our commitments to both equality and liberty. I argue that this principle should be satisfying to both egalitarians and free speech advocates.

Introduction

It is difficult to conceive of a democracy without public officials—sometimes elected and sometimes not—who make authoritative decisions that structure life in that society. The hallmark of a democracy is its openness to hear everyone's voice before deciding who becomes a public official and what those officials end up doing. Only by being inclusive in this way can a democracy live up to its promise of respecting each of its citizens as free and equal. Yet, many democracies fail to live up to this promise. It is widely acknowledged that wealthy individuals wield immense political power in the United States. But if the problem is that wealthy people's voices are too loud, the solution seems like it must be to silencing them.

The problem becomes: How can we coherently object to the state of affairs without undermining basic civil liberties? Political philosophers like Josh Cohen responds with a principle of equal opportunity of political influence. After discussing this principle in "Money, Politics, Political Equality" (2001), Cohen maintains that the demands of the principle would not undermine First Amendment guarantees of free speech. But Cohen is only able to rein in the principle's demands by limiting its scope to *political* influence rather than *public* influence. Cohen's narrow conception of political influence concerns influence over decisions made only by formal political institutions. On the other hand, his more capacious notion of public influence concerns influence over "the formation of opinion in the wider, informal public sphere, as well as decisions taken by formal political institutions."[1]

---

[1] Cohen, Joshua, "Money, Politics, Political Equality," in *Fact and value: Essays on ethics and metaphysics for Judith Jarvis Thomson*, edited by Robert Stalnaker, Alex Byrne, and Ralph Wedgwood, (Cambridge, Mass: MIT Press, 2001), 54.

For convenience's sake, I will refer to the principle concerning the *wide* public sphere as the wide principle of political equality. The principle concerning the narrow political sphere will then be the *narrow* principle of political equality. In this paper, I will argue for a principle of political equality that splits the difference between the narrow and wide principles. I will proceed in several parts. In the first section, I will lay out Cohen's explication of the narrow principle and expound on his dismissal of the wide principle. In the second section, I will introduce Ryan Pevnick's objection that Cohen faces a dilemma of embracing a specious narrow principle and a tyrannical wide principle. Finally, in the third section, I will turn to defend a compromise principle by fleshing out its details and fending off various objections. Elizabeth Anderson's approach in "What is the Point of Equality?" (1999) will aid my arguments for this final section.

I

Cohen defends the principle of equal opportunity of political influence. He makes three distinctions about what his principle is *not*. First, his principle does not demand that each citizen have equal political influence. Cohen points out that such an alternative principle would not be able to accommodate differences in how much a citizen cares about politics. Cohen favors a principle of equal *opportunity* for influence over simply equal influence. Second, Cohen differentiates his principle of equal opportunity from a principle of sufficient or maximin opportunity. Cohen rejects any proposal to give some individuals more political influence than others. He argues that if some had greater say in an authoritative system of making collective decisions, then the public bases of mutual respect would be threatened.

Cohen's third distinction is the one mentioned above between political and public. Recall that the political sphere includes formal political institutions, while the public sphere includes the background culture or "the dispersed networks of political-cultural discussion, founded on the associational life of civil society."[2] Cohen reiterates that he does not advocate for equal opportunity for *public* influence, just equal opportunity for *political* influence—he supports the narrow principle, not the wide one. He offers three reasons for this move. First, Cohen argues that the requirements of the wide principle are obscure: "The informal process of opinion-formation is not at all well defined or bounded: it extends throughout life, spreads through all its spheres, and the processes involved are not at all well understood."[3] Instead of requiring equal influence in this ill-defined sphere, Cohen insists that our civil liberties ensure that citizens have a *sufficient* chance to participate in this political-cultural discussion. Equalizing public influence would undermine our basic civil liberties, at least according to Cohen. Second, he claims that we cannot cite the same reasons for demanding the wide principle as we can for demanding the narrow principle. Our commitment to democracy grounds the narrow principle, because a democracy must regard each of its citizens as equals in contributing to authoritative collective decisions. On the contrary, Cohen does not see democracy as issuing any further command to equalize everyone's contribution to the public sphere. Cohen's third reason seems almost to combine his first two reasons. Compounded by the obscurity he has already defended, Cohen does not see the wide principles as being necessary

---

[2]Ibid., 49.

[3]Ibid., 54.

for establishing mutual respect among citizens.

II

In "The Anatomy of Debate about Campaign Finance" (2016), Pevnick claims that Cohen face a dilemma between speciousness and tyranny. If Cohen embraces the narrow principle, he ends up being specious in his commitment to equality. But, if Cohen opts for the wide principle, he ends up advocating for tyranny.

Let us first focus on the speciousness leg of Pevnick's argument. His crucial move against Cohen reconfigures the public sphere as inextricably part of the political sphere. Since the background culture and the processes of opinion formation certainly influence how citizens vote, interventions by the wealthy to influence the background culture will thus end up affecting election outcomes. So, we are left with a system in which the wealthy have a disproportionate ability to influence decisions made by formal political institutions. Since Cohen argues decisively that citizens must exercise equal opportunities to influence such decisions, we can see Cohen as being speciously committed to equality: his principle does not actually solve the problem that motivates its creation.

Before turning to the second leg of Pevnick's dilemma related to tyranny, we can assess Pevnick's objections in light of the three reasons Cohen provides for rejecting the wide principle. We can summarize Cohen's three reasons as amounting to the following: The public sphere is too obscurely defined to feasibly constrain. Given that equalizing public influence would undermine our civil liberties, we can rest assured that citizens have sufficient chances at influencing the public sphere. Plus, failing to advocate for the wide principle will not erode democracy or equal respect in any meaningful way. Pevnick's main intervention is to counter Cohen's final point. Failing to commit to the wide principle will allow our democracy to be compromised in the way explicated above. Simultaneously, Pevnick goes on to agree with Cohen that the wide principle would undermine civil liberties.

As for the other aspects of Cohen's defense, Pevnick does not address them directly but we can try to respond on his behalf. Pevnick would likely be unsatisfied with Cohen's assurance that citizens will have good enough access to the public sphere otherwise. It seems plausible that the extraordinary public influence of the wealthy may render other voices powerless in this domain. We want a principle that is strong enough to render such outcomes impermissible. It is apparent that the narrow principle is not a good candidate for such a principle. As for Cohen's remark about the obscurity of what precisely counts as the public sphere, Pevnick may try to take that on board as a point in his favor. We are left with a wide principle that may deeply harm civil liberties and may do so in an ambiguous and uncertain way.

Pevnick goes on to contend that the wide principle is tyrannical in its rejection of civil liberties. The wide principle would recommend that everybody would have an equal opportunity to influence the background culture. Pevnick mentions some examples of activity that would then be restricted: subsidizing the publication of books, funding scholarships, and starting magazines to promote a certain viewpoint. Pevnick puts this point even more forcefully by suggesting that the wide principle demands that the government monitor every citizen's use of private resources to ensure that they do *not* stand up for what they believe. Pevnick summarizes his dismissal of the wide principle with the following: "This is a leveling approach: rather

than protecting our liberty as a way of respecting our equality, it seeks to establish equality by dramatically limiting our freedom."[4]

III

In this section, I will try to mediate the disagreement between Cohen and Pevnick by proposing a principle that splits the difference between the wide and narrow principles. For convenience, I will call it the just-right principle. After arguing for the initial plausibility of the principle, I will take on the hefty task of dismantling Cohen's and Pevnick's overlapping concerns that any principle resembling the wide principle will compromise civil liberties in an undesirable way. Borrowing argumentation from Elizabeth Anderson will be helpful in fending off these concerns.

To motivate the just-right principle of political equality, let us recap the dialectic so far. On the one hand, Pevnick argues that the narrow principle is not strong enough as an egalitarian ideals, since it leaves unchecked certain activity in the public sphere. On the other hand, Cohen argues that a democracy *qua* democracy has no obligation to follow the wide principle in equalizing influence in the public sphere. What can explain these differing accounts of the public sphere? I propose a division within the public sphere that will resolve these paradoxical features. Plus, this division will aid us in building a new principle of political equality.

The basic division is between elements of the public sphere that influence formal political institutions and elements that do not. The public sphere or the background culture encompasses a variety of subject-matters that have nothing to do with politics: art, science, sports, entertainment, religion, cultural experience, quotidian experience, etc. Perhaps every item on this list may overlap in some areas with politics, but I suggest only that there are some elements of the discussion about these subject-matters that remain uncontaminated by politics. We can refer to the amalgamation of elements in the public sphere that do not influence political outcomes as the *non-political-public sphere*. To be sure, the public sphere includes discussions relevant to political outcomes. Pevnick identifies elements of the public sphere that heavily influence political opinion, including think-tanks, scholarships, yard signs, news media, and advertising more broadly. We can refer to the *political-public sphere* as the amalgamation of elements in the public sphere that influence political outcomes.

Thus, I think we can split the difference along these lines without too much difficulty. The just-right principle would then aspire to everyone having equal opportunities to influence both the political sphere and the political-public sphere—i.e. equal opportunity of political influence and political-public influence. This principle would be responsive to both Pevnick's and Cohen's points: Pevnick can be satisfied that inequalities in political-public influence will be alleviated, while Cohen can rightfully claim that democracies have no interest in equalizing non- political-public influence.

We can now turn our attention to whether the just-right principle undermines civil liberties. For the sake of the discussion, I will assume that Cohen's argument is successful that the narrow principle can accommodate adequate civil liberties. We will therefore focus on whether we can justify restrictions to politically efficacious speech that occurs in the public sphere. In the previous two sections, we have referenced a few different arguments and examples that suggest we value liberty over

---

[4]Pevnick, Ryan, "The Anatomy of Debate about Campaign Finance," *The Journal of Politics* 78.4 (2016b), 1192.

equality when it comes to this political-public influence. Pevnick offers a few examples of this type of activity that we would not want to restrict: publishing partisan books, funding partisan scholarships, and starting partisan magazines. In "Does the Egalitarian Rationale for Campaign Finance Reform Succeed?", Pevnick warns of the consequences of embracing a principle that would equalize political-public influence: "Doing so would prevent Al Gore from traveling on his own dime to give talks about the importance of climate change... It would prevent citizens from using their money to print leaflets or distribute yard signs arguing for the importance of their preferred position on abortion."[5] Our intuitions tell us that attempts to regulate such activity is not permissible in a democracy. Recall the close government surveillance that Pevnick thinks such a principle would be committed to. We can also resurrect Cohen's suggestion that we can rest assured that every citizen has a chance to speak *at some volume*, because our civil liberties combined with the narrow principle will ensure this. Finally, we may be worried about First Amendment arguments against this principle that we will flesh out later.

My first point in responding to these objections is to insist on the parallels between political influence and political-public influence. We can again combine Pevnick and Cohen. Pevnick argues persuasively that certain activities within the public sphere will be politically efficacious in just the way political influence is. How directly the speech affects political institutions makes no morally relevant difference. At the end of the day, we object to discrepancies in political-public influence for the same reasons we object to discrepancies in political influence. Cohen contributes that we must insist on equal political influence as opposed to just sufficient or maximin political influence. If we ought to treat political influence just the same as political-public influence, then Cohen's reasoning seems to commit himself to equal opportunity of political-public influence as well. Thus, so long as we take on board Cohen's narrow principle, we can use Pevnick's objections to further commit ourselves to the just-right principle. Finally, if we think that equality trumps concerns about civil liberties in the case of the narrow principle, we seem plausibly committed to this priority relation in the case of the just- right principle as well. I will defend this thought in greater detail in the following.

How can we overcome our intuitions about the various examples Pevnick gives? These examples in general struck me as a little misguided about what the just-right principle would require. For example, I found his example of constant government surveillance rather baffling. The just-right principle may call for somewhat radical proposals about equalizing institutions that apportion opportunities for political-public influence. But, regulating yard signs in addition to constant government surveillance are by no means necessary. For a harder case, consider what we are to do about Al Gore. Being unsure what exactly Pevnick has in mind, suppose that Al Gore spends millions of dollars on literature, movement-building, promotional materials, and traveling/living expenses in order to advocate for the importance of climate change. To some degree, this example probably appeals to the liberal sensibilities of Pevnick's audience. Besides the content of Gore's speech, there is no difference between his activities and those undertaken by the Koch brothers to make the country significantly more conservative. I would venture that Pevnick's audience may not have stepped in as quickly to defend Charles Koch's liberties. Moving away from this specific concern about the example being used, we can suggest why neither

---

[5]Pevnick, Ryan, "Does the Egalitarian Rationale for Campaign Finance Reform Succeed?" *Philosophy Public Affairs* 44.1 (2016a), 65.

of these shows of wealth appear legitimate in a democratic society. This is not to say that we should abandon Gore's or Koch's liberties altogether. Rather, it is not permissible for either of them to wield tremendously more influence on the political system than the rest of us. The fundamental principle is that we want everybody's voice to be heard, but we do not want arbitrary characteristics like wealth to be able to amplify a person's voice to whatever degree her money can by.

There is a tweak to the Gore example that is worth exploring. Suppose instead that there was a government program in which everyone would have the option to travel around the country for a few weeks and give un-promoted talks in free venues like public parks. It is true that Gore would probably attract a much bigger crowd than you or I would. Thus, Gore would have greater political-public influence. We might even say that Gore has greater opportunities for political-public influence. Is this inequality unjust? I would argue that it is not, according to the just-right principle. The idea is that some inequalities in political-public influence are earned. In this case, Gore earned throughout his career the respect and trust of many fellow citizens. I would resist that Gore has greater *opportunities* for influence. Gore was simply able to make the most of the opportunities for political-public influence given to him. At some point in Pevnick 2016a, Pevnick worries about whether Jon Stewart's or Bill O'Reilly's extraordinary influence would be allowed to exist under the just-right principle. I think these comments speak to this concern; assuming these personality's greater influence was legitimately earned, it is unproblematic. This is the essence of democratic rule that motivates elections in the first place: a person can only gain the political influence of being a public official if enough of the public agrees that she is worthy of it.

In order to more fully address our intuitions about Pevnick's examples, we must address the First Amendment worries. Drawing on Cohen's discussion of this topic, we find that the Supreme Court uses a core Millian doctrine in deciding cases about campaign finance regulations. The thought is that reducing the quantity of speech always hampers collective decision-making. Restricting my ability to speak does not merely harm me, but it also harms society's ability to be as informed as possible. Thus, challengers may charge the just-right principle of advocating for the government to restrict the flow of information in certain nefarious ways. We want as much speech as possible in order to ensure the best democratic decisions.

My response to this objection is that it is overly simplistic in insisting on maximizing speech. We can hear echoes of Rawls' concern that utilitarianism does not take seriously the distinction between persons. Likewise, this basic formulation of a Millian principle too does not seem to take seriously the distinction between speakers. Of course, the principle does not sanction anyone's voice to be ignored, since the stipulation is that everyone's political voice should be equally loud. On the contrary, I would in turn challenge these First Amendment advocates of effectively silencing certain voices. By permitting the very rich to do most of the talking in our political discourse, we can observe the drowning out effect that emerges from allowing individuals to speak as loudly as they wanted—the loudest drown out the voices of the quietest. I believe that the failure to recognize the serious drowning out effect in this domain stems from a failure to see influence as a deeply positional good. While I cannot argue in great detail for this, it seems at least plausible that increasing influence for one person always ends up decreasing everyone else's influence. The other major grievance Pevnick may have against the just-right principle is that of leveling down: if we implemented the just-right principle, we would end up in a soci-

ety in which every citizen has almost no opportunities for political-public influence. In response, I will defend that this objection rests on empirical predictions on the feasibility of certain institutional arrangements. As alluded to in the next section, certain changes could be made that move us closer to the just-right principle without leveling down to minimal opportunities for everyone. It is certainly a non-trivial enterprise to show decisively that the just- right principle could not be implemented without severe leveling down.

I will now gesture at restrictions that may be proposed on the basis of the just-right principle. Consider a policy that would forbid attaching partisan conditions to donations to universities and think-tanks. This policy would prohibit wealth individuals to push around on partisan grounds these crucial knowledge-producers and educators. Or, consider a policy that would forbid the proprietors of news media corporations from exercising partisan influence on what the company reports and who it hires. The general approach is to build institutions that are impervious to moneyed interests exercising undue political influence on their agendas. While there may be hurdles to crafting enforceable legislation with the just-right principle, there is abundant democratic promise in its achievement. Domination in the background culture would be outlawed when such domination would have political consequences—these reforms would be analogous to existing policies that attempt to thwart domination in political institutions. Free speech and domination can then be seen as two sides of the same coin: sometimes allowing one person to speak as loudly as she is able to afford will lead to the domination of those who cannot spend as much on their speech. Crucially, our democratic values limit this principle to only contexts where direct or indirect political influence is at stake. The theory underpinning the promise of the just-right principle can be found in Anderson's conception of democratic equality. Negatively, Anderson argues against any type of oppression that comes in the form of social relationships. Positively, Anderson argues for "a social order in which persons stand in relations of equality."[6] When it comes to matters of political or political-public influence, standing in a relation of equality to one another is particularly important.

IV

In this paper, I offered an improvement to the egalitarian approach to political influence. After considering Josh Cohen's argument for a narrower principle of political equality, I entertained Ryan Pevnick's objections to this principle. Taking seriously too his objections to a wider principle, I came up with a compromise principle that still faces objections from Cohen, Pevnick, and First Amendment advocates. I hope to have shown how we can ward off these concerns, rendering the just-right principle an attractive alternative for egalitarians concerned about political equality.

---

[6]Anderson, Elizabeth S, "What is the Point of Equality?" *Ethics* 109.2 (1999), 313.

References

Anderson, Elizabeth S. "What is the Point of Equality?." *Ethics* 109.2 (1999): 287-337.

Cohen, Joshua. "Money, Politics, Political Equality." *In Fact and value: Essays on ethics and metaphysics for Judith Jarvis Thomson*, edited by Robert Stalnaker, Alex Byrne, and Ralph Wedgwood, 47-80. Cambridge, Mass: MIT Press, 2001.

Christiano, Thomas. "Money in politics." In *The Oxford Handbook of Political Philosophy*, edited by David Estlund, 241-258. Oxford: Oxford University Press, 2012.

Mill, John S. *The Basic Writings of John Stuart Mill*. New York: Modern Library, 2002.

Pevnick, Ryan. "Does the Egalitarian Rationale for Campaign Finance Reform Succeed?" *Philosophy Public Affairs* 44.1 (2016a): 46-76.

Pevnick, Ryan. "The Anatomy of Debate about Campaign Finance." *The Journal of Politics* 78.4 (2016b): 1184-1195.

Rawls, John. *A Theory of Justice*. Cambridge, Mass: Harvard University Press, 1971.

On Relativizing Simultaneity

Rashad Rehman
Univeristy of Western Ontario

Abstract

Within contemporary parlance, one is often confronted with the notion that relativity theory has shown that absolute simultaneity is not to be found within our space-time reality. It is only with the advent of modern science, in particular relativity theory, that we can now know there is no such thing as absolute simultaneity. While this is no doubt a common belief among defenders of 'relativizing simultaneity', there is a hastiness in this judgement. This paper concerns itself with laying out contemporary quantum physical data which implies that there exists absolute simultaneity. I do this by consulting Michael Tooley's paper "A Defense of Absolute Simultaneity", in which he argues that since we have empirical grounds to reject the special theory of relativity (formulated by Einstein), we can accept the quantum physical data which implies absolute simultaneity. Whether the argument Tooley presents is accepted, it is a contribution to the debate regarding the special theory of relativity and quantum physics, and has consequences extending well beyond the confines of philosophical analysis.

In this paper, I intend to exegetically lay out an interesting argument—relying on contemporary quantum physical data—in favor of preserving absolute simultaneity and why, contrary to a *prima facie* acceptance of relativity theory, absolute simultaneity is plausibly a feature of our space-time. In modern parlance, it is alleged that relativity theory shows how there cannot be absolute simultaneity; however, I argue that contemporary quantum physical data provides an argument that the application of relativity theory as an epistemic defeater of absolute simultaneity is mistaken, since it is plausible that relativity theory itself will have to be ultimately given up. It is an interesting fact about ourselves that within our temporal, phenomenological experience, events seem to, on occasion, happen at the same time—put more formally, *simultaneously*. However, these events do not only appear simultaneous relative to us, but occurs *de facto* absolute simultaneously. While relativity theory provides the best known defeater for this wide-spread and very ingrained belief of ours, it is my contention that upon closer inspection it is plausible that relativity theory does not successfully provide a defeater for our belief in absolute simultaneity. The Special Theory of Relativity[1], formulated by Albert Einstein, has the implication that all events in time are relative to a reference (inertial) frame, and so there exists—ontologically—no privileged (standard, objective, absolute) reference frame which is the standard to which events conform. Michael Tooley's quantum mechanical argument, however, developed in his paper "A Defense of Absolute Simultane-

---

[1] Hereafter STR. Einstein first presented the STR in 1905.

ity"[2], is an argument which seeks to break through the horns of an inclusive disjunction regarding STR and Quantum Mechanics[3] themselves, namely, that either we accept quantum mechanics or the special theory of relativity as *incomplete*. As such, I will lay out the STR conceptually along with Tooley's argument for an alternative picture of QM and STR. I will thereafter present the philosophical problems that arise, and conclude that the objections to absolute simultaneity are plausibly unsuccessful.

The STR, formulated and developed by Einstein himself, has the implication that absolute simultaneity does not, and therefore cannot, occur. Here is the problem it poses: If we take absolute simultaneity to be a relation between events in the world, there must exist a privileged reference frame to which the "simultaneous" events conform; however, there is no non-arbitrary reference frame devisable. Any frame, therefore, will either be conventional or arbitrary (and probably conventional anyway)[4]. So, we must give up the notion of absolute simultaneity. While the details of the argument are (admittedly) quite daunting, the point is obvious enough—if there is no objective, privileged reference frame, there can be no relation of absolute simultaneity since events must be simultaneous *relative* to the same reference frame. (And this reference frame is lacking). Instead of casting doubt on STR in isolation, says Tooley, we should rather "appeal to quantum mechanics", especially concerning "the collapse of a wave packet" which "involves the notion of simultaneity"[5]:

> ... the argument can be put as follows. Consider an electron that has been fired towards the screen of a cathode ray tube. According to quantum mechanics, the initial conditions do not causally determine what point on the screen the electron will strike. All that follows from the laws of quantum mechanics is that there are various non-zero probabilities of the electron's striking different parts of the screen. But once the electron strikes the screen, the probabilities of it hitting other parts of the screen must change from having non-zero values to being equal to zero. But then must it not be the case that all of these changes are absolutely simultaneous with the event that consists of the electron's hitting the screen in a certain location?[6]

Tooley concludes that since this shows STR and QM to be inconsistent theories, one must be given up, though Tooley says that it "is [a] perfectly reasonable response"[7] to simply give up QM from the thought experiment. An objection here, though, might be posed as follows: Perhaps there is more to adjudicating the credibility of a scientific (or metaphysical) hypothesis than merely selecting, for instance, STR over

---

[2]Printed in William Lane Craig and Quentin Smith's *Einstein, Relativity and Absolute Simultaneity*. (New York, NY: Routledge, 2008), p. 229-243. For the remainder of the paper I will limit my discussion to the second argument Tooley gives (p. 239-243) under the subsection "The scientific argument: quantum mechanics and absolute simultaneity."

[3]Hereafter QM.

[4]See my last footnote for a caveat to this phrase.

[5]Craig and Smith, p. 239. This is very similar (structurally) to what has been called a "Cambridge Chain."

[6]Ibid., p. 239. It should be noted that in understanding Tooley's experiment, Stephen Hawking and Leonard Mlodinow's section relevant to Cathode Rays—and the historical background—was especially helpful. See *The Grand Design*. (New York, NY: Bantam Books, 2012), p. 47-48.

[7]Ibid., p. 239.

QM in virtue of STR implying absolute simultaneity. In fact, there are a plethora of 'epistemic virtues' of scientific theories which help adjudicate between equi-probable scientific theories i.e., explanatory power, explanatory scope, predictability, et cetera. In this sense, the philosophical implication is that there is more to a scientific theory than mere evidence; and hence while evidence is ground on which a theory is built, epistemic virtues to scientific theories can be both informative and decisive. However, the objection from giving up QM on the grounds that there are epistemic virtues to scientific theories which contribute to adjudicating a scientific theory's truth value is wholly consistent with Tooley's argument, and supports it. In fact, Tooley's appeal to quantum mechanical evidence i.e., Bell's theorem, is constitutive of the scientific theory virtue of *conformity to prior evidence*. Indeed, it is precisely this prior evidence which constitutes, says Tooley, the rejection of STR. With this in mind, experimentally speaking, the argument can be re-cast in terms of a theorem in quantum mechanics, the implication of which is that absolute simultaneity is kept—and so STR must be given up.

Tooley begins narratively, beginning with the Einstein-Podolsky-Rosen (EPR) argument in the form of a thought experiment. The alleged implication of the EPR is that QM entails an incomplete description of reality. Here 'incompleteness' does not mean 'false'; it simply means that in the end if QM is withheld in the fundamental theory of physics (if there will be such a thing), it will be "QM + x" and not *just* QM. As such, the EPR begins with the following quantum mechanical observation[8]: "...quantum physics does not attribute any properties to an object until a relevant measurement has been made."[9] In other words, there does not exist a pre-measured determinate property of any phenomena i.e., the spin of an electron prior to a measurement of it. Put formally (where 'e' is an electron, 'p' a determinate property, 'm' a measurement and 't' some arbitrary time):

Quantum Mechanical Principle 1 (QMP1): e has p iff m at t occurs.[10]

However, there exists a counter-example to this notion—requiring pre-measured determinate properties—which would, if it were true, contradict this quantum physical presupposition. Imagine, then, two electrons (let's call them e and e') which, having come into being, have the same total spin on the same axes such that the value of the spin of e is the same as e'. Now suppose that e and e' are separated but at t, m is performed to only e. It will be the case that e has a determinate value (m has occurred); but, a deep problem arises: "..the spin of [e], along the given axis, has a determinate value. But, what about the other electron? When does it have a determinate spin, along the axis in question?"[11] To clarify, the question derivative from QMP1 is how, if the spin of e determines the spin of e', e' could fail to have m and yet have a determinate property—a strict violation of QMP1. A worry here is as follows: Perhaps

---

[8]No *double entendre* intended.

[9]Ibid., p. 239.

[10]This is very contra-intuitions about the structure of reality. Thus A. N. Prior in his "Some Free Thinking About Time": "...and you can't have a thing existing from one point of view but not existing from another, although of course its existence may be *known* to one person or in one region, without being known to or in another." Reprinted in Peter van Inwagen and Dean Zimmerman's *Metaphysics: The Big Questions*. (Oxford, UK: Blackwell Publishers, 1998), p. 107. This is an expression of epistemic, not ontic skepticism. (I am indebted to Prof. J. Thorp's metaphysics class lectures for this distinction).

[11]Ibid., p. 240.

this thought experiment really posits just *e* and not *e'*. If *e* and *e'* have the same spin and come into being together, why posit their ontic independency? To clarify, why not concede that by making the assumption that *e* and *e'* are distinct, this is merely *nominal* and not *ontological*?[12] A reply would be to suggest that between *e* and *e'* there are further properties which are constitutive of their ontological independence from one another. Let me lay out what, in metaphysical terms, is 'constitutive of their ontological independence' amounts to. Consider Leibniz' Law: $(x)(y)[(x=y) \longrightarrow (P)(Px \longleftrightarrow Py)]$.[13] (In our case, let 'x' and 'y' represent '*e*' and *e*"). If *e* and *e'* are to be the same thing, then they must share *all* their properties; however, even granting their temporal properties to be identical, as well as their properties relevant to their spin, they have one decisive difference in properties, namely, their *spatial* properties. Since they exist at different space-time co-ordinates, they have a different property and so cannot be the same—hence, this objection does not work.[14] While one can just deny the legitimacy of Leibniz' Law, the problem of distinct spatio-temporal co-ordinates remains (unless one wanted to deny that this was a necessary condition for ontological independence).[15]

Thus, there are, says Tooley, two options available from the EPR. The first option posits that neither *e* nor *e'* ever acquired a determinate spin (Einstein, Podolsky and Rosen favored this interpretation), since grammatically the verb 'to acquire' presupposes it was not already there since, ontologically speaking, the spin exists in a pre-measurement way and thus, so the thought goes, "measurement enables us to know what those determinate values are; it does not bring those determinate values into existence."[16] Contrarily, the second view suggests that *e'* acquires its value when *e* does. Unfortunately, there is a paradoxical problem: When does this 'acquiring' happen? Again, we have two options: Either *e'* obtains its value later than *m* (of *e*) or simultaneously with *m*. On this view, either there exists a temporal gap between *e* obtaining its value and *e'* obtaining it's value, in which case "[this] would imply a violation of the principle of conservation of spin"[17], or else—and this is the second position—it could be the case that "[*e'*] has a determinate spin at every time throughout some immediately following, open temporal interval."[18] The problem with the latter approach is that if *e* and *e'* get their value simultaneously, it cannot be *absolutely*

---

[12]I would like to thank Professor John Thorp for raising this worry to me on the first draft of this paper.

[13]Moreland, J.P. and William Lane Craig. *Philosophical Foundations for a Christian Worldview*. (Downers Grove, Illinois: Intervarsity Press, 2003), p. 194.

[14]This criteria for ontological independency has both a limited domain over which it quantifies, as well as a limited functional role, however. For instance, in cases i.e., human beings, *uniqueness* cannot be derivative from mere ontological independency i.e., spatial location. This seems to be a mistake made evident from Samuel Delorme's characterization of Nietzsche's theory of the self in his "Nietzsche and The Meaning of Life": ". . . each bud ['bud' representing metaphorically a single individual person] diverges into its own path [on a tree, a metaphor for life]. . . it distinguishes itself from all others by its space. . ." *The Oracle* X.I: 13. While the addition of 'and direction' (of each 'bud') might be significant, it seems nonetheless that this is trivial with respect to uniqueness *fundamentally*.

[15]Of course, though, Leibniz' Law might not be universally valid (for all we know). In this case, though, it seems philosophically justified to use it for my argument.

[16]Craig and Smith, Ibid., p. 240.

[17]Ibid., p. 240.

[18]Ibid., p. 240-241. Tooley takes it as axiomatic that *e'* cannot acquire its determinate property before *e* does (since it would be a self-contradiction).

simultaneous (otherwise the STR has to be rejected). So, what can it mean? Tooley suggests that it means relative to some inertial frame '*F*.' But, he argues, if this were correct than if *e* and *e'* obtain their values relative to *F*, then there is another inertial frame *F** that, relative to *F*, requires that *e'* is obtained after *e*—and thus there is a temporal gap (and so faces the problem the first alternative entailed). So, relative simultaneity does not work here; absolute simultaneity is, logically, deduced necessarily.

Since suggesting that *e* and *e'* obtain their values relatively simultaneously leads to the violation of conservation laws, another alternative to interpreting the EPR thought experiment seems required. As such, this position posits that "while the second electron does not have a determinate spin at the time when the first electron acquires one," instead "it has a determinate spin at every time throughout some immediately following open temporal interval."[19] Tooley notes, though, that this faces the aforementioned violation problems since the open temporal interval itself requires that "one has to make use of the idea that the instant that immediately precedes that interval is simultaneous with the instant at which the first electron acquires a determinate spin" and thus "only absolute simultaneity will do."[20] Given that both these attempts to avoid absolute simultaneity do not work, Tooley wants to make a comment on the EPR thought experiment which shapes the rest of his discussion. He notes that what the EPR thought experiment really shows is that either particles have determinate values prior to measurement (requiring the incompleteness of QM) or that correlated particles acquire determinate properties absolutely simultaneously (which requires the incompleteness of STR).[21] Ultimately, then, it seems that we are back to deciding which (now admittedly deflated) theory we want to adopt—either QM or STR. At this point, though, Tooley adds that "there is more to the story."[22]

Tooley notes physicist John S. Bell's 1964 paper in which his results entailed that "quantum mechanics logically preclude there being properties that make it the case, for instance, that an electron possess determinate spins along various possible axes before any measurements are made."[23] The result of this is that the inclusive disjunction that "Either QM or STR is incomplete" is now that either STR is incomplete, or QM is *false*. With the empirical confirmation that QM metaphysically entails that QMP1 is true, we are stuck with not merely the notion that QM is incomplete, but that STR is incomplete too and—on the basis of empirical findings—inconsistent with QM i.e., since it does not "provide a complete description of the spatio-temporal relations between events,"[24] namely, absolute simultaneity. Hence, in order to hold STR, one must think QM not merely incomplete—but *false* (which is contrary to the evidence Tooley suggests). Tooley notes that the debate on absolute simultaneity is not merely theoretical, it is now empirically testable—to which he firmly suggests that QM is correct and thus STR given up.[25] The implication, to come back to *e* and

---

[19]Ibid., p. 241.

[20]Ibid., p. 241.

[21]Ibid., p. 241.

[22]Ibid., p. 241.

[23]Ibid., p. 242. Tooley's footnote here quotes references to the formulations of Bell's theorems aimed at a non-specialist level.

[24]Ibid., p. 242.

[25]The evidence presented remains without explicit display, but with explicit reference. Tooley references here Aspect, A., P. Grangier, and G. Roger (1981) 'Experimental Tests of Re-

e', is that if one says either that the value of *e'* is acquired at the same time as the *m* of *e*, or that the value of *e'* is acquired "at every moment after a time that is absolutely simultaneous with the time at which the first electron acquires a determinate spin,"[26] either way absolute simultaneity is unavoidable, and therefore "there are experimental grounds for holding that absolute simultaneity is a relation that obtains between events in our world."[27]

In this paper, I have outlined Tooley's quantum mechanical argument for absolute simultaneity. Contrary to STR, the quantum mechanical evidence—largely from Bell's theorem—in conjunction with other evidence, supports the notion that we do not have to give up that human intuition of, and phenomenological experience that, events can be (whether epistemically accessible or not) absolutely simultaneous. If Tooley's argument is sound, what is next is choosing a physical interpretation of the mathematical formalism of QM which involves absolute simultaneity.[28] Finally, what Tooley has not done is provide an absolute reference frame, but has instead shown that absolute simultaneity—rejected by STR—cannot be given up. This to my mind is progress, and perhaps we shouldn't be surprised with attempts to non-arbitrarily devise an objective, metaphysical reference frame given that absolute simultaneity is an indispensable fact about the world.[29] To put it more clearly, perhaps it is not surprising that Plato, in his late dialogue *Laws*, had it correct when he said "Let us, then, speak to them thus:— 'O men, that God who, as old tradition tells, holdeth the beginning, the end, and the center of all things that exist. . .'"[30], to which Boethius rightly continued, his De Trinitate, that "*divinum vero "nunc" permanens neque movens sese atque consistens aeternitatem facit. . .*"[31]

---

alistic Local Theories via Bell's Theorem', *Physical Review Letters* 47, 460-67; Aspect, A., P. Grangier, and G. Roger (1982) 'Experimental Realization of Einstein-Podolsky-Rosen-Bohm *Gedankenexperiment*: A New Violation of Bell's Inequalities', *Physical Review Letters* 48, 91-94.

[26]Craig and Smith, Ibid., p. 242.

[27]Ibid., p. 242.

[28]This is a sort of 'meta-analytical' note for those doing the work in the philosophy of quantum physics.

[29]I have in mind here the interesting argument from W. L. Craig's work, especially his "The Metaphysics of Special Relativity" in Ibid., p. 11-49; see also his *Time and the Metaphysics of Relativity*. (The Netherlands, Dordrecht: Kluwer Academic Publishers, 2001), Chapter 10.

[30]*Platonis Opera* 715e, ed. John Burnet. Oxford University Press. 1903.

[31]"...but God's 'now,' abiding, unmoved and immovable, connotes eternity." "De Trinitate" in *Boethius: The Theological Tractates and The Consolation of Philosophy*. Trans. Stewart, H. F. E. K. Rand and S. J. Tester. (Cambridge, Massachusetts: Harvard University Press, Loeb Classical Library), p. 23.

References

Aspect, A., P. Grangier, and G. Roger (1981) 'Experimental Realization of Einstein-Podolsky Rosen-Bohm *Gedankenexperiment*: A New Violation of Bell's Inequalities', *Physical Review Letters* 48, 91-94.

——— (1982) 'Experimental Tests of Realistic Local Theories via Bell's Theorem', *Physical Review Letters* 47, 460-67.

Boethius. "De Trinitate" in *Boethius: The Theological Tractates and The Consolation of Philosophy*. Trans. Stewart, H. F. E. K. Rand and S. J. Tester. Cambridge, Massachusetts: Harvard University Press, Loeb Classical Library. Print.

Burnet, John. *Platonis Opera*. Oxford University Press. 1903. Greek and English text Retrieved from: http://www.perseus.tufts.edu/hopper/text?doc=Perseus

Conee, Earl and Theodore Sider. *Riddles of Existence: A Guided Tour of Metaphysics*. New York, NY: Oxford University Press, 2004. Print.

Craig, William Lane and Quentin Smith. *Einstein, Relativity and Absolute Simultaneity*. New York, NY: Routledge, 2008. Print.

——— *Time and The Metaphysics of Relativity*. The Netherlands, Dordrecht: Kluwer Academic Publishers, 2001. Print.

Hawking, Stephen and Leonard Mlodinow. *The Grand Design*. New York, NY: Bantam Books, 2012. Print.

Moreland, J.P. and William Lane Craig. *Philosophical Foundations for a Christian Worldview*. Downers Grove, Illinois: Intervarsity Press, 2003. Print.

van Inwagen, Peter and Dean Zimmerman. *Metaphysics: The Big Questions*. Oxford, UK: Blackwell Publishers, 1998. Print.