

The varieties of impartiality, or, would an egalitarian endorse the veil?¹

Justin P. Bruner, University of Groningen

Matthew Lindauer, Brooklyn College, City University of New York

Forthcoming in *Philosophical Studies*

Abstract

Social contract theorists often take the ideal contract to be the agreement or bargain individuals would make in some privileged choice situation (i.e., an ‘original position’). Recently, experimental philosophers have explored this kind of decision-making in the lab. One rather robust finding is that the exact circumstances of choice significantly affect the kinds of social arrangements experimental subjects (almost) unanimously endorse. Yet prior work has largely ignored the question of which of the many competing descriptions of the original position subjects find most compelling. This paper aims to address this gap, exploring how attractive experimental subjects find various characterizations of these circumstances of choice. We find evidence suggesting that no one choice situation can fulfill the role that social contract theorists have hoped it would play. We also find that, contrary to what some prominent social contract theorists have expected, there is no robust relationship between an individual’s ranking of distributive principles and their ranking of various descriptions of the original position. In conclusion, we discuss the broader implications of these results for political philosophy.

1. Introduction

Political philosophers working in the social contract tradition have traditionally been interested in showing that their preferred theory is both uniquely choiceworthy and most suitable to justification to a democratic public. We will call these two criteria “uniqueness” and “publicity.” Previous experimental research has been taken by many to support the view that, within a given choice procedure or “impartiality frame,”² a particular contractarian conception of justice could meet these criteria. Yet no research thus far has given sustained attention to whether any given impartiality frame *itself* could satisfy uniqueness and publicity. Examining

¹ JB and ML contributed equally to the paper. Both authors designed the experiments and contributed towards interpreting and sussing out the philosophical significance of the results.

² We will use the term “impartiality frame” to cover the various choice procedures that we examine in this paper, but it is worth noting that, in addition to impartiality, these choice procedures can model other moral ideals. See Moehler (2018a,b) for an important discussion of the moral ideals of impartiality, impersonality, and equality.

these questions is very important because, as we shall see, if at the more general level of impartiality frame these criteria are not fulfilled, much of the motivation for examining uniqueness and publicity of the principles chosen within these frames is undermined. Our paper is the first, to our knowledge, to examine the extent to which prominent impartiality frames can meet these criteria. We provide evidence to support a pessimistic answer, and explore the implications of our results for further developments in social contract theory. Before presenting these results, we will say more about how uniqueness and publicity relate to one another and what is at stake in examining whether different components of conceptions of justice and, in particular, impartiality frames can meet these criteria.

In broad strokes, the social contract tradition considers the decisions of rational or reasonable³ individuals placed in various circumstances of choice designed to ensure some level of impartiality (we refer to a choice situation as an *original position* or an *impartiality frame*). Decisions are then taken to inform the features of a just social contract as well as justify various social arrangements. One assumption commonly made in the social contract tradition is *uniqueness*. This is the assumption that rational or reasonable individuals will converge on a uniquely choiceworthy way of organizing and structuring society. Uniqueness has wide support and is generally, though not universally, assumed by contractarians and contractualists alike. Rawls (1971), for instance, famously contended that decision-makers would converge in their judgments.

While critics of the social contract tradition have questioned whether the decisions of rational and reasonable individuals will coincide (see, for instance, Sugden 1990; Skyrms 1996, 2016; Thrasher 2013; Muldoon et al. 2014), recent experimental work on the social contract has been taken to provide strong evidence in favor of uniqueness. Norman Frohlich and Joe Oppenheimer (Frohlich and Oppenheimer 1993) have found that subjects placed in circumstances of choice approximating the Rawlsian original position and its ‘thick’ veil of ignorance by and large endorse the same social contract, average utilitarianism with a minimum floor or “restricted utilitarianism,” and these findings have now been widely replicated (Lissowski et al. 1991; Bruner forthcoming).⁴ Additional experiments investigating a ‘thin’

³ Reasonableness here is meant to convey the fact the individuals in question are not only advancing their own interests, but receptive to the claims and interests of others (cf. Rawls 1971). Some contractarians, such as Rawls, assume that the participants in the original position are both rational and reasonable. Nothing significant will turn on this point in the paper.

⁴ Miller (2001) takes the choice of the restricted utilitarian principle to support a pluralistic conception of justice. He argues that, whereas Rawlsian justice as fairness is a conception of justice that is overly focused on need, a pluralistic conception of justice must balance the sometimes competing desiderata of desert, need, and equality. Participants in Frohlich and Oppenheimer’s study appear to want to balance the needs of the badly off with the desert of those who worked harder in order to receive more in choosing the restricted utilitarian principle. It’s worth noting that this isn’t an argument against uniqueness in the selection of a uniquely choiceworthy principle or set of principles of justice, but rather against a singular focus on one

version of the veil based on Thomas Scanlon's contractualism (1998), as well as work exploring the judgments of informed but impartial third-parties, à la Adam Smith and David Hume, have also suggested that convergence is the norm.

As mentioned, experimental work of this kind is particularly significant since many theorists take what is unanimously selected by contractors to be the ideal social contract. Furthermore, prominent theories of political justification rely on a similar approach, privileging those principles overwhelmingly supported by rational and reasonable persons. Yet the same empirical evidence showing that unique principles are selected within prominent impartiality frames also demonstrates that there is significant divergence between the principles selected *across* frames. Principles nearly unanimously selected under one impartiality frame often gain little to no support from subjects deliberating from an alternative impartiality frame. For instance, while those behind a Rawlsian veil of ignorance opt for utilitarian principles, those tasked to perform as third-party spectators support egalitarian arrangements such as the difference principle, which holds that social and economic inequalities must be to the greatest benefit of the worst-off. Thus, this egalitarian principle gains substantial support, just not from those deliberating from behind the Rawlsian veil.⁵ This divergence suggests that while there may be a determinate social contract when all contractors employ the same impartiality frame, choice of impartiality frame significantly affects the properties of the contract that is agreed upon.

Divergence of this kind threatens to undermine the strategy traditionally employed by those in the social contract tradition. Since the circumstances of choice partly determine which contract agents assent to, there are either a plurality of ideal social contracts, an option that social contract theorists have traditionally hoped to avoid, or a case must be made as to which of the available frames should be privileged. We explore the latter option in this paper, which is consistent with the traditional contractarian commitment to a determinate and unique social contract, although we also provide a few thoughts regarding the former option in section 6. Recently, some social contract theorists have abandoned the commitment to uniqueness at the level of impartiality frame (Gaus 2010; Moehler 2018a). In virtue of militating against the view that some prominent impartiality frame should be privileged over others, we take the results of our experiments (section 3-5) to support this relatively new strand of contractarian thought.

Of course, impartial choice situations are motivated in various ways by the theorists who advance them, and the justification of an impartiality frame partly depends on how well it coheres with considered judgments regarding principles of justice associated with the frame. Yet it is clear that there is no agreement in the philosophical literature as to how these circumstances

value in the development of these principles. We are grateful to an anonymous reviewer for pushing us to clarify this point.

⁵ While the difference principle is sometimes characterized as a prioritarian principle, given that it is largely motivated by egalitarian reasoning and forms part of an egalitarian conception of justice, we will simply refer to it as an egalitarian principle in this paper. Throughout the paper, mentions of 'egalitarian principles' all refer to the difference principle.

of choice *should* be best characterized. We explore this issue further and present a series of experiments designed to investigate whether individuals are overwhelmingly drawn to a particular impartiality frame. A comparison of frames is particularly apt given that, as noted above, previous empirical research has only focused on behavior in a given impartiality frame, rather than the choice of frame itself.

It is worth noting that the fact that no prominent impartiality frame is regarded as uniquely choiceworthy may be more problematic for the traditional social contract theorist than the variation of principles chosen across frames. Upon seeing that experimental subjects converge on different principles of justice under different impartiality frames, the traditional social contract theorist may retreat to the position that the choice of principles should not be viewed as a statistical generalization based on people's choices. Rather, it is a mathematical result in the case of bargaining theories of justice, and a deduction in the case of traditional contract theory, one might think. We are doubtful that all social contract theorists can make this claim. For instance, while Rawls initially sought to provide a "strictly deductive" (1971, pp. 121, 104-5) argument for his favored principles of justice, his mature position qualifies this claim, holding that the selection of principles from the original position as matching the "ideal of rigorous deductive reasoning cannot...be fully attained" and must instead be based on "judgment informed and guided by reasoning" (2001, pp. 133-4). Rawls acknowledges that, rather than following deductively from the original position, principles of justice will have to be selected on the basis of considerations that can be balanced in different ways, although he is confident that a uniquely choiceworthy set of principles will still emerge from this process. But even if one did embrace this view for principles of justice, the choice of which impartiality frame to use when deriving principles has not been held to be a strictly deductive exercise, either from the concept of 'justice' or any basic assumptions. Hence, it is especially important to explore how ordinary people tend to view these impartiality frames and which, if any, is most compelling if we want to achieve agreement on principles of justice that are selected under them.⁶

Furthermore, the concept of uniqueness plays multiple roles in the social contract tradition, some of which depend on what ordinary people would agree to. Our experiments pertain not only to debates concerning the ideal social contract, but also the related issue of the *publicity* of principles of justice. That is, principles uniquely selected by rational or reasonable persons have been thought to be capable of being justified to a democratic public. Principles that are widely agreed to—and not merely imposed—form the grounds of a public conception of justice that will be stable over time *for the right kinds of reasons* (Rawls 1993). Their stability will rely on an ongoing commitment to these principles, rather than the merely self-interested reasons that ground a *modus vivendi*. As a result, these principles of justice can be justifiably implemented in designing and reforming the institutions of a democratic society. While not necessary to meet the requirement of publicity, since multiple sets of principles of justice could potentially be widely accepted, finding that principles are uniquely supported in a society is

⁶ We are grateful to an anonymous reviewer for pushing us to clarify this point and the following one.

sufficient to show that they meet this requirement. This is a further reason why many social contract theorists have been interested in uniqueness, and one that cannot be seriously addressed without attention to what principles of justice ordinary people find compelling. As we will see, our experiments address these issues and shed light on which principles are most likely to serve as components in a widely-accepted public conception of justice.

The paper proceeds as follows. First, we will discuss in more depth the role that uniqueness plays in the social contract tradition as well as the extant experimental literature on justice. Section 3, 4 and 5 then outline the design and results of three experiments we have conducted. Section 6 discusses their philosophical significance and concludes.

2. Uniqueness and the lab

Contemporary social contract theorists consider the contract rational and reasonable individuals would agree to. Importantly, contractors are placed in a choice situation that allows for some level of impartiality, and it is typically assumed that individuals will come to endorse the same contract. This latter assumption of “uniqueness,” as mentioned above, is clearly seen in the writings of a variety of recent thinkers. We discuss a few representative cases in what follows. Famously, Rawls (Rawls 1971) assumes that in the original position individuals select principles of justice from behind a ‘veil of ignorance.’ Not knowing their class, race or other differences, contractors will agree unanimously on a conception of justice. Relatedly, John Harsanyi (Harsanyi 1953) contends Bayesian rational decision-makers will all assent to the same (utilitarian) social contract from behind the veil.

Similarly, contractarians such as David Gauthier (Gauthier 1986) presuppose a unique contract will be selected by rational individuals. Gauthier, however, models justice not as agreement under conditions of uncertainty, but instead as a bargain among rational agents.⁷ For Gauthier, the uniquely ‘rational solution’ to the bargaining problem is for contractors to allocate resources in a way which minimizes the maximum concession made by those involved. In a similar vein, James Buchanan and Gordon Tullock (Buchanan and Tullock 1962) consider the case in which individuals are tasked with selecting institutional rules (such as a constitution) to govern collective decision-making. Due to epistemic limitations, individuals are unable to accurately predict the impact various rules will have on their interests, and are thus unable to promote their group’s interest when negotiating institutional rules. As a result, Buchanan and Tullock contend contractors will instead unanimously settle on rules benefiting the “average or representative player” (p. 80).

⁷ It is worth noting that many political philosophers that model justice as the outcome of a rational bargain are drawing on work from a tradition in the social sciences that has been largely focused on identifying ‘unique’ solutions to strategic encounters. Work in game theory and bargaining theory, for instance, has for decades attempted to pin down unique solutions to various games. This goal of identifying a determinate ‘solution’ to the bargaining problem has fed into and reinforced the desire of political philosophers to identify a unique social contract (see in addition to Gauthier 1986, Binmore 2005).

As mentioned, social scientists and philosophers have recently explored how experimental subjects behave when placed in scenarios approximating various impartial choice situations discussed in the literature. Contributions in this vein can be traced back to the pathbreaking work of Frohlich and Oppenheimer. Frohlich and Oppenheimer (Frohlich and Oppenheimer 1993; Frohlich, Oppenheimer and Eavey 1987) consider small deliberative groups tasked with discussing and then voting on various principles of justice. The principles that are chosen determine how much group-members earn for their participation in the experiment. Additionally, individuals are placed ‘behind the veil’ and are thus unable to determine the precise impact principles have on their earnings. Frohlich and Oppenheimer find that an overwhelming majority of groups gravitate toward the same principle (“maximize the average with a floor constraint” or “restricted utilitarianism”), and in some treatments, *all* groups select this principle. It is worth noting that this finding has been replicated multiple times. Frohlich and Oppenheimer themselves have considered variations of their original experimental design, and others have registered that ‘maximize the average with a floor constraint’ attracts wide support in other subject populations (Frohlich and Oppenheimer 1990; Lissowski et al. 1991; Bruner forthcoming).

Experimental evidence in favor of uniqueness extends beyond work on Rawls. James Konow’s work on ‘quasi-spectators’, for instance, suggests that consensus is the norm when individuals must stand as an informed third-party (Konow 2009). When given detailed information about a hypothetical scenario involving imagined individuals, experimental subjects converge on the same fairness judgments. Similar experiments have also registered that spectators tend to converge in their judgments when provided context and information about the individuals of interest (Faravelli 2007; Mitchell et al. 1993; Herne and Mard 2008; Konow 2008).⁸ Experimental work on Thomas Scanlon’s original position also indicates widespread agreement holds among subjects. This is particularly noteworthy since Scanlon himself does not claim a unique social contract would be agreed to in his original position. Nonetheless, Herne and Suojanen (2004) and Herne and Mard (2008) find that egalitarian principles are routinely selected when individuals know how various principles affect their earnings and all have the ability to reject each principle.⁹

Yet although the experimental work surveyed above bolsters the claim that individuals will converge in their judgments, a significant divergence is nonetheless observed. While utilitarian arrangements are near-unanimously agreed to behind the veil, egalitarian principles are selected when agents are either placed in Scanlon’s original position or asked to stand as an impartial, but nonetheless informed, third-party. These findings are somewhat robust, and hold

⁸ See, however, Michelbach et al. (2003) for evidence of dissensus among spectators as well as Schotter and Sopher (2003) for evidence of dissensus in a bargaining context.

⁹ It is worth noting that the majority of the experiments cited in this section provided monetary incentives to subjects, ensuring more accurate elicitation of preferences. Faravelli (2007) and Konow (2008) are exceptions in this respect.

across different subject pools and experimental setups.¹⁰ Taken together, these experiments suggest that the choice of principle crucially hinges on how the circumstances of choice are characterized. While this is perhaps not too surprising—one would expect details regarding the situation of choice to in part determine *what is in fact chosen*—the level of divergence we observe *across* impartiality frames is truly striking. Overwhelming support for utilitarian or egalitarian principles can arise depending on how one characterizes the original position.

Thus, while prior experimental work has given us some reason to think uniqueness holds for principles selected from a given impartiality frame, we are still not in a favorable position to describe the features of the ideal social contract. As mentioned in the introduction, one response at this point is to surrender the very goal of identifying an ideal social contract, and instead embrace a variety of different social arrangements as legitimate (see, for instance, Gaus 2010 and Moehler 2018a). We discuss this option more in section 6, but note for now that this avenue is likely to be viewed as unattractive by most social contract theorists, who have rather deep and long-held commitments to there being a unique, ideal social contract. We thus explore an alternative response they are likely to give in what follows. Namely, we attempt to determine whether uniqueness holds at the level of impartiality frame, i.e., whether certain descriptions of the circumstances of choice are seen as especially attractive. If widespread agreement exists with respect to how the original position should be constructed, along with widespread agreement on principles that should be selected from within that impartiality frame, divergence across frames becomes significantly less troubling. While different frames might not lead to the same principles of justice, the existence of a privileged frame garnering widespread support would allow us to recover a unique social contract.

Before we get to our experiments, however, a few words on the particular impartiality frames we consider. There are of course a variety of different frames that have been suggested and considered in the literature. Our aim here is not to provide a taxonomy nor an exhaustive experimental comparison of frames. Instead, we draw on a few canonical impartiality frames most familiar to political philosophers. In particular, we consider Rawls' veil of ignorance, the perspective of a disinterested and sympathetic third-party and Scanlon's account of the original position. This first frame involves Rawls' 'thick' veil, where individuals are ignorant of their abilities and social circumstances and must select principles of justice in this epistemically impoverished state (we refer to this frame as *Thick*). The second impartiality frame involves the judgments of a sympathetic and informed, but nonetheless detached, spectator who adjudicates the affairs of others, which is most often associated with the writings of Adam Smith (1759), but also traceable to the work of the other Scottish Enlightenment thinkers David Hume (1740) and Francis Hutcheson (1728). This impartiality frame, which we will refer to as *Spectator*, has been discussed at length in the experimental ethics literature (see, for instance, Konow 2012).

Finally, the third frame involves Scanlon's original position. In this case, individuals are placed behind a 'thin' veil and thus have knowledge of their social circumstances, abilities, and

¹⁰ For an experimental comparison of these frames, see Herne and Suojanen 2004, Herne and Mard 2008 and Aguiar, Becker and Miller 2013.

talents (we refer to this frame as *Thin*). Impartiality is achieved because individuals only consider those principles of justice no reasonable person could reject. In the next sections we discuss in more detail a series of experiments involving all three of these impartiality frames.

3. Experiment 1

In Experiment 1, we sought to determine whether uniqueness holds at the level of *impartiality frame*. There are, once again, two primary motivations for examining uniqueness at this level of political theory. First, while there is empirical evidence for consensus at the level of principles selected under a given frame, there is as yet no research examining whether particular impartiality frames receive more support than others. This is problematic insofar as agreement not just under some impartiality frame, but under the *correct* or *best* impartiality frame, is supposed to be evidence that a given set of principles are the true principles of justice. Second, many political philosophers want to offer not only a true theory of distributive justice, but a *public conception of justice* – one that could be agreed upon by real citizens and used to do real work in an actual democratic society. There is a tradition of regarding the ability of a set of principles to secure the agreement of citizens as a theoretical virtue of those principles, i.e. part of what should go into our evaluation of whether or not they are the true principles of justice (Rousseau 1762, Kant 1795, Rawls 1971). Since uniqueness is sufficient, though not necessary, for public justification, social contract theorists have been interested in publicity as an additional and important upshot of establishing the uniqueness of their preferred conception of justice. We are skeptical, however, that securing consensus on any one conception of justice is a reasonable goal for political philosophy, and one important reason for our skepticism is the plurality of appealing impartiality frames that are offered in the social contract tradition. We anticipate that none of the most promising impartiality frames will achieve anything like the level of consensus needed to meet the uniqueness condition.

To test this “non-uniqueness” hypothesis¹¹ regarding impartiality frames, we designed an experiment that presented the three most prominent types of frames found in the philosophical and empirical literature on social contract theory. They are the Rawlsian veil of ignorance (*Thick*), the Smithian impartial spectator (*Spectator*), and the Scanlonian reasonable acceptance (*Thin*) frames. In Experiment 1, participants¹² were given brief instructions that explained what

¹¹ We are grateful to an anonymous reviewer for suggesting this label and pushing us to clarify that our results shouldn’t be taken to imply that uniqueness has no role to play in contractarian thought. For instance, Moehler (2018a) holds that while there may be no unique social contract, in the absence of such a contract, agents may come to agree on what he calls a “pure instrumental morality” in order to ensure peaceful cooperation on an ongoing basis.

¹² Our participants were 104 MTurk users located in the United States. We limited the study to users with a 98% or better approval rating for their previous work. Participants were required to follow instructions and enter a randomly-generated code that was given to them in order for their results to be counted.

impartiality frames and principles of distributive justice are, along with some background information about distributive justice and political philosophy. Subjects were then asked to imagine a society called Alpha that was trying to determine what distributive principles were required by justice. Subjects were told they were not in fact members of Alpha. The members of Alpha were entertaining the idea of using one of these impartiality frames to determine which distributive principles were just. Participants read a description and then evaluated each of these impartiality frames by registering their agreement with the statement “This is a good way of determining whether a principle is just or not” on a 7-point Likert scale (endpoints and midpoint were labeled in the prompt as follows: “1 = Not at All, 4 = Somewhat, 7 = Strongly Agree”). The presentation of the impartiality frames to each participant was randomized to control for potential order effects.¹³

It is worth noting the criteria that we used to test uniqueness. Clearly, it would be too exacting a standard to expect that every single participant would favor one impartiality frame uniquely over the others. Below this standard, uniqueness can be positively spelled out in a number of ways, and it is beyond the bounds of this paper to adjudicate among them. Instead of giving a positive account of what kind of support an impartiality frame would have to receive for it to be regarded as uniquely choiceworthy by participants, we adopt the more theoretically modest criterion of testing impartiality frames against a plausible and minimal necessary condition for uniqueness. That is, an impartiality frame must at least receive greater support than its competitors, to a statistically significant extent, if it is to be regarded as a uniquely choiceworthy impartiality frame. Any plausible characterization of what being regarded as uniquely choiceworthy by participants consists in will have to accept this as a necessary condition. If there is no statistically significant difference between the ratings given to one frame in comparison to its competitors, we can’t establish whether any apparent difference, if one is observed at all, is a genuine difference or merely due to chance. Of course, this is only a necessary condition, and not a sufficient condition for uniqueness. For instance, an impartiality frame could receive significantly greater support than its competitors while all frames receive a very low level of support. In such a case, it would be implausible to think that such a frame is regarded as uniquely choiceworthy by participants – the frame simply fares better than other undesirable options. But as we shall see, using the minimal necessary condition of significantly greater support than the alternatives allows us to demonstrate that the prominent frames that we are examining are not regarded as uniquely choiceworthy without having to provide a full analysis of uniqueness. We are thus able to establish our conclusions while remaining neutral among competing conceptions of uniqueness, since any plausible criterion of uniqueness will include this necessary condition.

¹³ Furthermore, we surveyed an additional 130 subjects to determine the extent to which subjects comprehended the different descriptions of the three impartiality frames and found no statistically significant difference in reported level of comprehension across frames.

Figure 1 displays histograms for the three impartiality frames (*Thick*, *Spectator* and *Thin*):

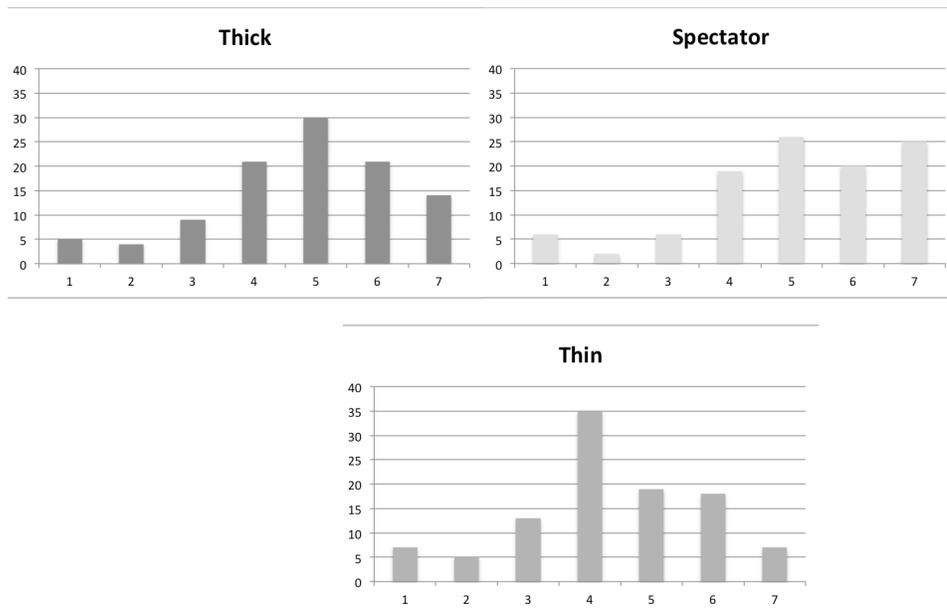
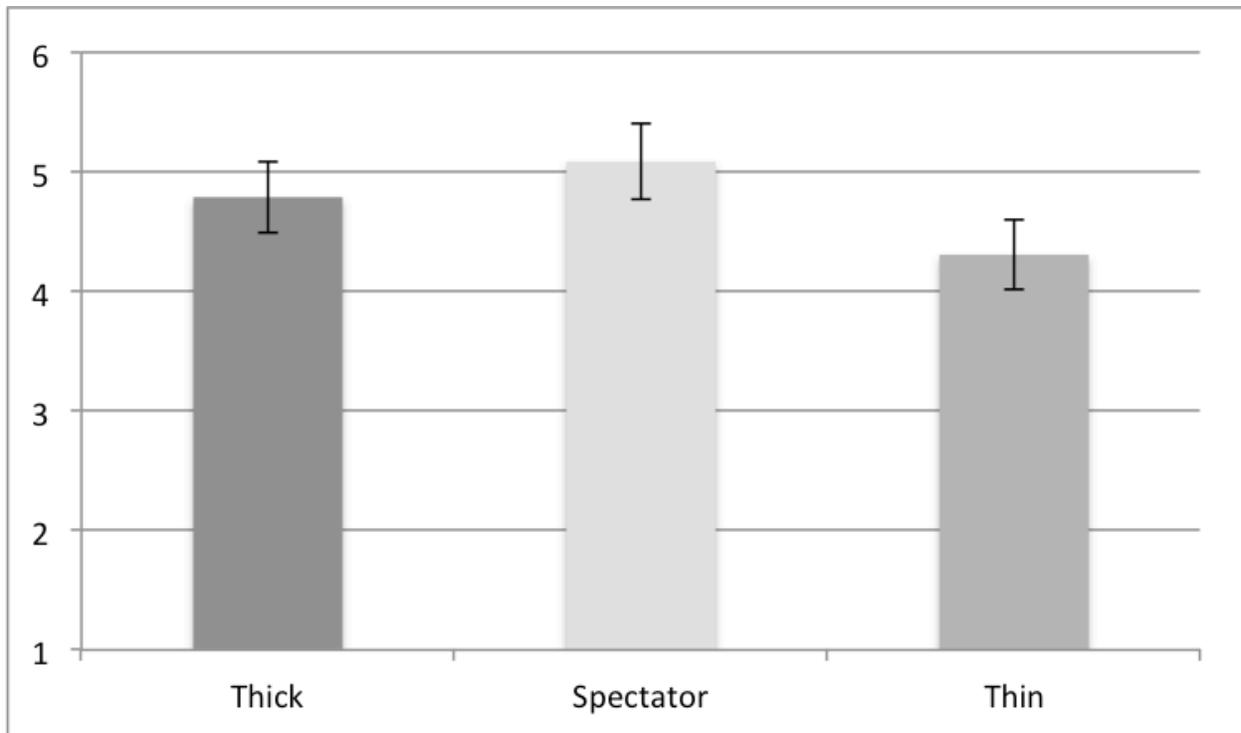


Figure 2 shows the mean ratings that participants gave to each impartiality frame:



Error Bars: 95% CI

There was a significant difference between the mean ratings that participants gave to each impartiality frame.¹⁴ In particular, the mean ratings that participants gave to *Thick* ($M = 4.79$, $SD = 1.54$) and *Spectator* ($M = 5.09$, $SD = 1.64$) were significantly higher than those given to *Thin* ($M = 4.31$, $SD = 1.53$).¹⁵ However, the difference between the ratings given to *Thick* and *Spectator* was not significant.¹⁶

The results of Experiment 1 suggest that none of these impartiality frames are regarded as uniquely choiceworthy. Again, while we do not attempt to give a positive criterion for uniqueness in this paper, it is plausible that in order to be regarded as uniquely choiceworthy, one of these frames would have to receive significantly greater support than its competitors. However, while the *Thick* and *Spectator* frames were viewed significantly more favorably than the *Thin* frame, neither the *Thick* nor the *Spectator* frame could claim to have received significantly greater support than the other. Hence, while this is not a sufficient condition for uniqueness of an impartiality frame, significantly greater support than the alternatives as a necessary condition for uniqueness is not met by any of the competing impartiality frames. Further, despite the significant differences that we did observe, none of the frames scored particularly highly among participants. Thus, our results give us reason to doubt that any of these frames could serve as a uniquely chosen impartiality frames in a society.

4. Experiment 2

The results of Experiment 1 support the non-uniqueness hypothesis about impartiality frames. While the *Thick* and *Spectator* frames were rated significantly higher than the *Thin* frame, there was no significant difference in judgments between them. The frames in Experiment 1 were, as mentioned above, set out as third-personal frames for participants – they were to imagine a society called Alpha that they wouldn't be a member of. However, one might think that some frames will perform better than others when participants imagine using them for a society that they will be a member of. In particular, the *Thick* Rawlsian frame might be thought to perform best for people imagining the kind of society that they would want to live in, and hence how their own society should be reformed to match their considered judgments about justice. One might expect this because thinking about the veil of ignorance naturally invites one to consider what rules one would agree to in the absence of knowledge of one's own socioeconomic position and social group membership. In Experiment 2, we tested whether presenting the frames in a first-personal way would make a difference ($N = 99$). We hypothesized that the non-uniqueness result would remain robust despite this change.

¹⁴ Repeated-measures ANOVA, $F(2, 206) = 7.268$, $p = .001$, $\eta^2 = .066$.

¹⁵ Paired-samples, *Thick* vs. *Thin*: $t(103) = 2.587$, $p = .011$; *Spectator* vs. *Thin*: $t(103) = 3.542$, $p = .001$. Because we will be making three comparisons, the Bonferroni correction sets the alpha level for significance at .01667.

¹⁶ Paired-samples, $t(103) = -1.411$, $p = 1.61$.

Figure 3 displays histograms for three impartiality frames, first-person (*Thick*, *Spectator* and *Thin*):

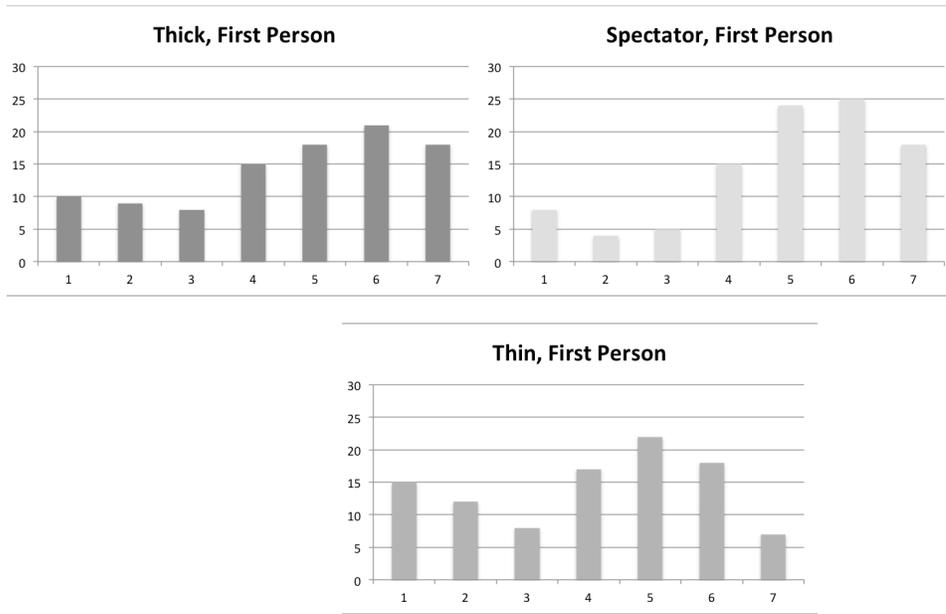
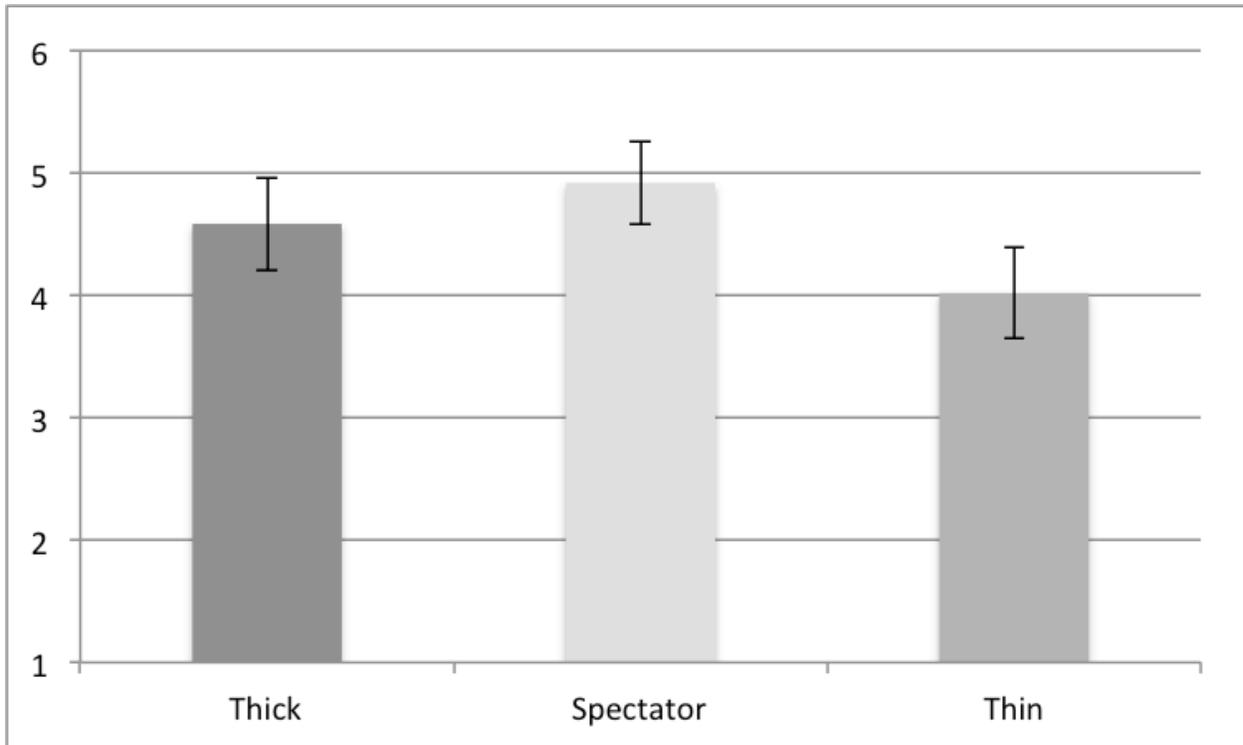


Figure 4 shows the mean ratings given to each first-person impartiality frame:



Error Bars: 95% CI

There was a significant difference between the mean ratings given to each impartiality frame.¹⁷ Also following the pattern of results found in Experiment 1, *Thick* ($M = 4.59$, $SD = 1.93$) and *Spectator* ($M = 4.92$, $SD = 1.74$) performed significantly better than *Thin* ($M = 4.02$, $SD = 1.88$),¹⁸ but there was no significant difference between *Thick* and *Spectator*.¹⁹

The presentation of impartiality frames as first personal rather than third personal did not change the fact that no frame was regarded as uniquely choiceworthy. The result that the *Thick* and *Spectator* frames were rated higher than the *Thin* frame was also robust between Experiments 1 and 2. Most importantly for our purposes, there was still no significant difference in the ratings given to the *Thick* and *Spectator* frames when presented first-personally as opposed to third-personally. This suggests that the non-uniqueness result found in Experiment 1 is robust and cannot be attributed to the prior third-personal framing.

5. Experiment 3

In the previous two experiments, we examined support for prominent impartiality frames, and found that across third-personal (Experiment 1) and first-personal (Experiment 2) variations, none of these frames was regarded as uniquely choiceworthy. In a third experiment ($N = 319$), we investigate whether a connection holds between preferences for impartiality frames and preferences for distributive principles. In other words, this experiment allows us to determine whether those inclined to support more utilitarian or egalitarian distributions tend to favor a particular impartiality frame. A clearer understanding of this connection is significant for a number of reasons. For one, it allows us to better understand what factors determine level of support for various frames. Additionally, it allows us to register whether people tend to endorse coherent principle-frame ‘packages.’ For instance, do egalitarians tend to favor frames that lead to the selection of egalitarian principles (such as *Spectator*)? Or, are egalitarians drawn to those frames more traditionally associated with egalitarian principles (such as *Thick*)? In other words, *would an egalitarian endorse the veil?* Notice that if the answer to these latter two questions is ‘no,’ then the situation appears rather dire for Rawls. Not only do subjects fail to select his favored distributive principle from behind the veil of ignorance, but those who favor egalitarian principles are not drawn to Rawls’ characterization of the original position. But more broadly, if coherent principle-frame packages do not tend to be endorsed, much of the motivation for contractarianism and its focus on securing principles through contractual procedures is called into question.

For this experiment we asked participants two questions, the order of which was randomized. One question began with a description of two distributive principles: Rawls’ difference principle and utilitarianism with a floor or the “principle of restricted utility” (see

¹⁷ Repeated-measures ANOVA, $F(2, 196) = 8.28$, $p < .001$, $\eta^2 = .078$.

¹⁸ Paired-samples, *Thick* vs. *Thin*: $t(98) = -1.387$, $p = .012$; *Spectator* vs. *Thin*: $t(98) = 4.336$, $p < .001$.

¹⁹ Paired-samples, $t(98) = 2.56$, $p = .017$.

appendix for exact wording). We then asked individuals to rank these principles. Similarly, our question pertaining to impartiality frames presented subjects with *Thick* and *Spectator*, the two frames that received the highest levels of support in Experiments 1 and 2. Subjects were then asked to select the frame they found most attractive.

We found that subjects were once again somewhat evenly split between *Thick* (42.6%) and *Spectator* (57.4%). While these proportions are significantly different,²⁰ clearly if 42.6% of the population prefers *Thick*, *Spectator* is not uniquely preferred. With respect to distributive principles, however, a little under three-fourths (74%) of subjects preferred the principle of restricted utility over the difference principle.²¹ This is quite striking. One possible explanation for this latter significant finding is that individuals took themselves in this experiment to be selecting principles from ‘behind the veil.’ That is, subjects may have imagined themselves to be stakeholders selecting principles that would ultimately impact their well-being in some difficult to determine fashion. Alternatively, it could be the case that individuals just by and large are inclined to endorse more efficient distributions but veer from this tendency only when placed in particular circumstances (such as the *Spectator* frame). Either way, it is clear some of our experimental subjects are *more* utilitarian than others, and we exploit this fact to then register whether a tendency toward utilitarian principles predicts one’s preferences over impartiality frames. In particular, we examine whether those inclined to rank utilitarian principles over egalitarian principles are more likely to endorse *Thick* over *Spectator*.

What we uncover is rather surprising. We find that choice of principles is *not significantly associated with choice of impartiality frame*. In fact, it appears that there is little connection *at all* between an individual’s choice of principle and choice of frame. For instance, of the 236 subjects who preferred the restricted utilitarian principle, 58.5% of them preferred *Spectator* to *Thick*. This is (nearly) the same proportion of the total population that ranked *Spectator* above *Thick* (57.4%). Similarly, 55.0% of those who preferred the difference principle also ranked *Spectator* over *Thick*. Preference for distributive principle does not appear to predict preference for impartiality frame.²²

Table 1 drives this point home. We compare the proportion of subjects in our experiment who endorse a particular principle-frame combination to the proportion of subjects one would expect to endorse this combination on the assumption that choice of principle and choice of frame are completely independent of one another. As Table 1 illustrates, the actual proportion of individuals endorsing a principle-frame combination is nearly identical to the proportion one would expect on the assumption of independence.

²⁰ We used a Chi-Square Goodness-of-Fit Test to compare these proportions, $\chi^2(1) = 4.63, p = .03$.

²¹ $\chi^2(1) = 46.06, p < .001$.

²² Additionally, 25% of those who preferred *Spectator* also preferred the egalitarian principle and 28% of those who preferred *Veil* preferred the egalitarian principle. Thus preference for frame does not appear to predict a subject’s preference for distributive principle. A Chi-Square Test for Independence confirmed that there is no association between the distributive principles and impartiality frames that participants chose, $\chi^2(1) = .004, p = .951$.

A few points are in order. It is worth mentioning that almost a majority of individuals selected the utilitarian principle in combination with the *Spectator* frame. This is noteworthy for two reasons. First, it undercuts a conclusion one may be tempted to draw from the experiments of Frohlich and Oppenheimer, namely, that the combination of utilitarianism and the *Thick* veil is the preferred principle-frame package. Our experiment indicates that while this package is attractive to experimental subjects, it is not the most popular principle-frame combination on offer. That said, the most popular principle-frame combination, utilitarianism and *Spectator*, was nonetheless striking as utilitarianism is not typically selected by experimental subjects asked to play the role of a disinterested but informed third-party.²³ In other words, experimental subjects tended to endorse a frame which, prior research suggests, is not judged to support the very principle they find most compelling. This mild inconsistency is somewhat puzzling and could be due to the fact that subjects either (i) do not take into account what principle would be chosen from a given frame when deciding among frames, or (ii) incorrectly predict that utilitarianism would be selected from the *Spectator* frame. Future experimental work is required to determine which of these explanations best captures these surprising results concerning the lack of a relation between preferences for impartiality frames and preferences for distributive principles.

	Spectator 57.4%	Thick 42.6%
Utilitarian 74.0%	43.3% (74.0% x 57.4% = 42.5%)	30.7% (31.5%)
Egalitarian 26.0%	14.1% (14.9%)	11.9% (11.1%)

Table 1: Percentage of individuals who independently endorse a distributive principle and impartiality frame. Numbers in parentheses refer to proportions of subjects who endorse a principle and impartiality frame combination on the assumption that endorsement of frame and endorsement of distributive principle are statistically independent.

²³ Recall that egalitarian principles were typically selected from the *Spectator* frame (Herne and Mard 2008).

6. General Discussion

In this paper we presented the results of three experiments suggesting that the search for a uniquely choiceworthy impartiality frame is misguided. When comparing among three prominent impartiality frames, subjects rated two of them (*Thick* and *Spectator*) significantly higher than the third (*Thin*) across both first and third personal variations, but neither of these two performed significantly better than the other in this comparison (Experiment 1 and 2). When we performed a runoff between these two impartiality frames, one performed significantly better than the other (*Spectator*), but not enough to justify the thought that it was the *uniquely* preferred choice (Experiment 3). We contend these results undermine much of the force of traditional approaches to contractarianism in political philosophy. If no impartiality frame is uniquely choiceworthy, neither are the principles of justice which follow from them.

Proponents of contractarianism might respond that the judgments of the folk are irrelevant. Such proponents might appeal to the supposed difference in the judgments of philosophers and non-philosophers, contending that while the former are reliable, the latter are not. Such appeals to the so-called “expertise defense,” it should be noted, have a shoddy empirical track record (Weinberg 2009; Weinberg et al. 2010; Schwitzgebel and Cushman 2012; Tobia et al. 2013), and we are skeptical of them. However, even if one is willing to accept the expertise defense for some contexts and topics, the matter of finding a *publicly justifiable* conception of justice relies on the premise that such a conception will be stably agreed upon by ordinary democratic citizens. As a result, there are no good grounds for excluding the judgments of reasonable and rational ‘non-experts’ in this particular debate.

As noted above, we take our results to be troubling for traditional contractarian approaches. The absence of a privileged impartiality frame leads to indeterminacy, thereby posing a problem for those committed to uniqueness at this level of theory. Yet while our findings undercut orthodox approaches to the social contract, this is not to say that social contract theorizing is for naught, or that the contractarian tradition should be abandoned. If anything, our experimental results provide additional motivation for a very recent turn in social contract theory exemplified by the work of Gerald Gaus (2010) and Michael Moehler (2018a). Gaus explicitly rejects uniqueness, contending a variety of distinct social contracts may be publicly justified and further argues no procedure for selecting a social contract is *uniquely* justified. Instead of invoking a collective decision procedure or impartiality frame, Gaus takes a pseudo-evolutionary approach and argues social evolution will ‘break the symmetry,’ resulting in eventual coordination on a publicly justified contract. Moehler takes a similar approach and advocates for a Humean account of the social contract involving an evolved system of moral rules that constitute a contract members of society are assumed to enter.²⁴ Our experiments provide new

²⁴ As mentioned in note 11 above, Moehler rejects the view that there is a unique social contract, but retains a role for uniqueness in his theory. He endorses a hybrid view, according to which agents may come to agree on a unique pure instrumental morality when there is widespread divergence regarding moral commitments. However, in better societal conditions, where values

empirical support for these recent developments in social contract theory, reinforce Gaus' contention and suggest where theorists working in the tradition should place their efforts.²⁵

Appendix

1. Background information (for Experiment 1 and 2)

Thank you for participating in this experiment. We are interested in how people think about justice. Your responses will help us learn more about this.

Distributive justice is concerned with how resources and goods should be fairly and justly distributed in society.

Not surprisingly, there are many different principles of distributive justice that people argue for – principles that say how and when a distribution of goods and resources can be just.

In addition to principles of distributive justice, there are different methods or procedures for choosing those principles. It is often argued that a principle is required by justice because it would be selected by a method or procedure that we have reason to think is just.

Now imagine that you are a member of a new society called Alpha that has been established. The people of Alpha must determine how to justly and fairly distribute goods and resources in their new society. In other words, they must determine **which distributive principle is required by justice**.

Imagine that you and the other members of Alpha are going to use a particular method or procedure to decide which distributive principle is required by justice. We're going to ask you **which distributive principle you would choose, given this way of deciding on a principle that we will describe**.

are more widely shared, a plurality of evolved norms for different societies will emerge on his account.

²⁵ We would like to thank Geoff Brennan, Matthew Kopec, John Matthewson, Ryan Muldoon, Jennifer Nado, Wlodek Rabinowicz, Nic Southwood, Katie Steele, Jeremy Strasser, Justin Sytsma, Joe Ulatowski, and Jonathan Weinberg for helpful discussion on the points raised in the paper. For written comments, we are especially grateful to Bob Goodin, Kaisa Herne, Serene Khader, Joshua Knobe, RJ Leland, Michael Moehler, Brian Skyrms, Jiewuh Song, and two anonymous reviewers for the journal. We would also like to thank the audiences at the Australasian Experimental Philosophy Workshop, the Economics and Philosophy Workshop at ANU, UC Irvine, and the 2nd Meeting of the PPE Society.

2. Impartiality Frames

Thick:

Members of Alpha consider what distributive principles they would select if they did not know anything about themselves. In other words, the distributive principle required by justice is the distributive principle they would select if they didn't know their race, gender, socioeconomic status, their talents, or how motivated they were.

Spectator:

Members of Alpha consider what distributive principle they would select if they were fully informed impartial spectators. In other words, the distributive principle required by justice is the distributive principle someone would choose if they knew all the relevant facts about the members of Alpha and didn't have any reason to favor any participant over others.

Thin:

Members of Alpha consider only distributive principles no member of society could reject for good reasons. In other words, if the members of Alpha don't have good reasons to reject a principle, that distributive principle can govern Alpha.

3. Distributive Principles

Difference Principle:

"The most just distribution of income is that which maximizes the floor or minimum income in society."

Utilitarianism with a Floor (Principle of Restricted Utility):

"The most just distribution of income is that which insures that all members of society can meet their basic needs and, once this condition is secured, maximizes the average income in that society."

References

Aguiar, F., Becker, A. & Miller, L. (2013). Whose impartiality? An experimental study of veiled stakeholders, involved spectators and detached observers. *Economics & Philosophy*, 29(2), 155-174.

Binmore, K. (2005). *Natural Justice*. Oxford University Press.

- Bruner, J. (forthcoming). Decisions behind the veil: an experimental approach. In T. Lombrozo, J. Knobe and S. Nichols (Eds.), *Oxford Studies in Experimental Philosophy, Vol 2*. Oxford University Press.
- Buchanan, J. & Tullock, G. (1962). *The Calculus of Consent: Logical Foundations of Constitutional Democracy*. University of Michigan Press.
- Faravelli, M. (2007). How context matters: A survey based experiment on distributive justice. *Journal of Public Economics*, 91, 1399-1422.
- Frohlich, N., Oppenheimer, J. & Eavey, C. (1987). Choices of Principles of Distributive Justice in Experimental Groups. *American Journal of Political Science*, 31(3), 606-636.
- Frohlich, N. & Oppenheimer, J. (1990). Choosing Justice in Experimental Democracies with Production. *American Political Science Review*, 84(2), 461-477.
- Frohlich, N. & Oppenheimer, J. (1993). Choosing Justice: An experimental approach to ethical theory. University of California Press.
- Gaus, G. (2010). *The Order of Public Reason: A theory of freedom and morality in a diverse and bounded world*. Cambridge University Press.
- Gauthier, D. (1986). *Morals by Agreement*. Oxford University Press.
- Harsanyi, J. (1953). Cardinal utility in welfare economics and in the theory of risk-taking. *Journal of Political Economy*, 61(5), 434-435.
- Herne, K. & Mard, T. (2008). Three versions of impartiality: an experimental investigation. *Homo Oeconomicus*, 25, 27-53.
- Herne, K. & Suojanen, M. (2004). The role of information in choices over income distributions. *Journal of Conflict Resolution*, 48(2), 173-193.
- Hume, D. (1740). *A Treatise of Human Nature*. D.F. Norton and M.J. Norton, eds., Oxford University Press, 2000.
- Hutcheson, F. (1728). *An Essay on the Nature and Conduct of the Passions and Affections, with Illustrations on the Moral Sense*. J. Smith and W. Bruce.
- Kant, I. (1795). Perpetual Peace. In H. Reiss (ed.), H. Nisbett (trans.), *Kant: Political Writings*, Cambridge University Press, 1991: 93–130.
- Konow, J. (2008). The Moral High-Ground: An experimental Study of Spectator Impartiality. MPRA Paper No. 18558.
- Konow, J. (2009). Is fairness in the eye of the beholder? An impartial spectator analysis of justice. *Social Choice and Welfare*, 33(1), 101-127.

- Konow, J. (2012). Adam Smith and the modern science of ethics. *Economics & Philosophy*, 28(3), 333-362.
- Lissowski, G., Tyszka, T., & Okrasa, W. (1991). Principle of Distributive Justice: Experiments in Poland and America. *Journal of Conflict Resolution*, 35(1), 98-119.
- Michelbach, P., Scott, J., Matland, R. & B. Bornstein (2003). Doing Rawls Justice: An Experimental Study of Income Distribution Norms. *American Journal of Political Science*, 47(3), 523-539.
- Miller, D. (2001). *Principles of Social Justice*. Harvard University Press.
- Mitchell, G., Tetlock, P., Mellers, B., & Ordonez, L. (1993). Judgments of social justice: Compromises between equality and efficiency. *Journal of Personality and Social Psychology*, 65(4), 629-639.
- Moehler, M. (2018a). *Minimal Morality: A Multilevel Social Contract Theory*. Oxford University Press.
- Moehler, M. (2018b). The Rawls-Harsanyi Dispute: A Moral Point of View. *Pacific Philosophical Quarterly*, 99(1), 82-99.
- Muldoon, R., Lisciandra, C., Colyvan, M., Martini, C., Sillari, G., & Sprenger, J. (2014). Disagreement behind the veil. *Philosophical Studies*, 170(3), 377-394.
- Rawls, J. (1971). *A Theory of Justice*. Harvard University Press.
- Rawls, J. (1993). *Political Liberalism*. Columbia University Press.
- Rawls, J. (2001). *Justice as Fairness: A Restatement*. Harvard University Press.
- Rousseau, J.J. (1762). *The Social Contract, Or Principles of Political Right*, transl. G.D.H. Cole, public domain.
- Scanlon, T. (1998). *What We Owe to Each Other*. Harvard University Press.
- Schotter, A. & Sopher, B. (2003) Social learning and coordination conventions in intergenerational games: an experimental study. *Journal of Political Economy*, 111(3), 498-529.
- Schwitzgebel, E. & Cushman, F. (2012). Expertise in moral reasoning? Order effects on moral judgment in professional philosophers and non-philosophers. *Mind and Language*, 27(2), 135-153.
- Smith, A. (1759). *The Theory of Moral Sentiments*. D.D. Raphael and A.L. Mackie, eds., Oxford: Oxford University Press, 1976.
- Skyrms, B. (1996). *Evolution of the Social Contract*. Cambridge University Press.

Skyrms, B. (2016). Evolution, Norms and the Social Contract. *Arizona State Law Journal*, 48, 1087-1100.

Sugden, R. (1990). Contractarianism and Norms. *Ethics*, 100(4), 768-786.

Thrasher, J. (2012). Uniqueness and symmetry in bargaining theories of justice. *Philosophical Studies*, 167(3), 683-699.

Tobia, K., Buckwalter, W. & Stich, S. (2013). Moral intuitions: Are philosophers experts? *Philosophical Psychology*, 26(5), 629-638.

Weinberg, J. (2009). On doing better, experimental-style. *Philosophical Studies*, 145(3), 455-464.

Weinberg, J., Gonnerman, C., Buckner, C. & Alexander, J. (2010). Are philosophers expert intuiters? *Philosophical Psychology*, 23(3), 331-355.