What experiments can teach us about justice and impartiality: vindicating experimental political philosophy

Aurélien Allard

Institute for Ethics, History and Humanities, University of Geneva

Florian Cova

Philosophy Department & Swiss Center for Affective Sciences, University of Geneva

Forthcoming in H. Viciana, F. Aguiar, & A. Gaitan (Eds.), Issues in Experimental Moral Philosophy. Routledge.

In the past twenty years, philosophers, psychologists and neuroscientists have been working together to improve our understanding of how we think about morality. This field, which we can either call "moral psychology" or "experimental moral philosophy" has been flourishing and it has become extremely hard to keep track of all the work that is being done in this area. This success provides a stark contrast with another potential field: "experimental political philosophy". In 2016, political philosopher Nicole Hassoun published a chapter entitled "Experimental or empirical moral philosophy", in which she called for a better integration of political philosophy with psychological and behavioral sciences, akin to the one that can be found for the study of morality (Hassoun, 2016). However, so far, her call does not seem to have fostered a lot of work in this perspective[1].

As we will see in our first section, this neglect is all the more surprising since political philosophers have long acknowledged the relevance of empirical data for normative political theory. Given the explosion of research in moral and political psychology, coupled with recent developments

[1] For notable exceptions, see Bruner, 2018; Bruner & Lindauer, 2020; Buckland et al., 2021; Davis & Preece, 2021; Lindauer & Barry, 2017.

in experimental philosophy, political philosophers now have all the tools they need to incorporate empirical data in their work. In the second section, we provide two concrete examples of how debates in political philosophy can be informed by the results of empirical studies. Finally, in the last section, we focus on the lessons that can be drawn from the existing empirical literature for philosophical debates on theories of justice[2].

1. The relevance of empirical data to political philosophy

Political philosophy has long acknowledged the philosophical relevance of empirical data about folk conceptions of what makes a good and fair society. This is particularly striking in John Rawls' seminal article "The Independence of Moral Theory", where Rawls first introduces the distinction between *narrow* and *wide* reflective equilibrium (Rawls, 1974). In Rawls' opinion, moral theory is primarily about "the substantive moral conceptions that people hold, or would hold, under suitably defined conditions". Thus, the first task of political philosophers is the rational reconstruction of popular ideas concerning justice and normative ideas. For Rawls, we should think of the moral theorist "as an observer, so to speak, who seeks to set out the structure of other people's moral conceptions and attitudes" (p.7). As such, the construction of moral theory is "a kind of psychology and does not presuppose the existence of objective moral truths" (p.9).

Such a task is what the moral theorist aims to achieve by attaining *narrow* and *wide* reflective equilibria. *Narrow* reflective equilibrium aims at identifying the structure of our current moral conceptions. Rawls explicitly claims that philosophers have to conduct this inquiry as sociologists or psychologists, recording popular theories and common ideas, and cataloguing widespread intuitions about normative issues. To be clear, Rawls does not claim that philosophers' aim is simply to record popular theories about fairness and justice. Such explicit theories should be considered as manifestations of an underlying structure people are not necessarily conscious of. Narrow equilibrium is thus already a philosophical reconstruction, through which philosophers strive to bring to light the

[2] Another domain which has recently benefited from fruitful collaborations between philosophers and empirical scientists is the theory of democracy, particularly regarding the epistemic value of deliberation and majority voting. For some examples, see Bor et al., 2021; Mercier et al., 2021; Mercier & Landemore, 2012.

underlying coherent moral principles that manifest themselves through various, sometimes inconsistent, popular opinions.

However, Rawls' method of reflective equilibrium does not presuppose that everyone shares the same moral conceptions, as Rawls explicitly acknowledges the possibility of moral pluralism. In this case, the outcome of narrow equilibrium should be the reconstruction of all rival moral theories that are already present, at some implicit or explicit levels, within a given society. The results of this first phase of philosophical inquiry will then provide the starting point of the second phase, which Rawls calls *wide* reflective equilibrium. In this second phase, philosophers should try to compare the different moral systems reconstructed during the first phase in order to identify where they agree and where they disagree, but also to "investigate what principles people would acknowledge and accept the consequences of when they have had an opportunity to consider other plausible conceptions and to assess their supporting grounds" (p.8).[3]

From this perspective, there is no doubt that empirical data are relevant to philosophical theorizing, as they can help identify and reconstruct the moral conceptions of people. Moreover, the fact that Rawls acknowledges the possibility of moral pluralism means that it would be foolish for the political theorist to only rely on their *own* intuitions.

Of course, not all philosophers accept Rawls' conception of political theory (see for example Singer, 1974). Rawls' own view is guided by the need to find a justification for political theory that does not presuppose the existence of objective moral truths. Depending on one's meta-ethical commitments, this lack of reference to objective standards can be seen as a virtue or a vice. However, even if one rejects Rawls' approach, there are still many reasons to accept the idea that empirical data from behavioral sciences can be relevant to political theory.

[3] There has not been a lot of empirical studies investigating how people manage to reach reflective equilibrium and how they handle contradictions between their considered judgments. For a rare example in the context of population ethics, see Schoenegger & Grodeck (this volume).

A first traditional argument starts from the observation that many philosophical arguments are based on fundamental intuitions that serve as basic and undemonstrated premises. To the extent that philosophical theorizing is grounded in such intuitions, it seems relevant to ask whether these intuitions are shared, and what factors can influence them (Knobe, 2007).

One widespread objection to this kind of argument is that non-expert intuitions are irrelevant to philosophical theorizing, because they are not as reliable as expert intuitions (Kauppinen, 2007; Sosa, 2007). However, the weight of this objection might not be as heavy for political philosophy as it is for metaphysics or epistemology. Indeed, while most of us are not accustomed to muse about complex epistemological and metaphysical cases (such as, let's say, Gettier cases), we often have the occasion to reflect about moral and political questions. This is why some authors (Burke, 2014; Hayek, 2012) have suggested that common sense intuitions about moral and political issues are not completely naive and clueless, as they have gradually been honed by a process of cultural evolution.

Another argument comes from the purpose of political philosophy. According to this argument, political philosophy should be seen as a public endeavor that aims to help people who disagree, but have to live together, converge towards some kind of consensus. Under this view, one role of political philosophers is to offer public advice in favor of conserving, eliminating or revising certain common ideas (Fischer et al., 2007). However, if this is the goal of political philosophers, then they cannot completely ignore common-sense principles and intuitions about fairness and other political issues: even an "objectively true" political theory would have no chance to succeed (and be useful) if it was not grounded in arguments the premises of which are intuitively accessible to non-expert citizens. To quote Miller (2001), a theory of justice should be "publicly justifiable". An accurate picture of the intuitiveness of various moral and political principles might thus delineate which political theories are out-of-question or, on the contrary, "on the table". Some have even argued that the main criterion of a good moral or political theory is its ability to bring all citizens together by appealing to principles that are intuitive to all. It's on this basis that Joshua Greene (despite his moral anti-realism) champions utilitarianism: because it is grounded on the idea, shared by all, that happiness is good (Greene, 2014).

Pushing this line of argument further, one might even argue that empirical research on popular intuitions should be even more crucial and relevant to political philosophy than it is to other philosophical domains. Indeed, the very practice of political philosophy is based on the legitimacy of public reason: political rules and principles should not be sought in some afterworld, but should be the product of public discussion and deliberation. In a word, philosophers need to know how much legitimacy and support a particular theory is likely to garner.

2. Two examples of empirically-informed political philosophy: natural injustice and behavior behind the veil of ignorance

So far, we have advanced general considerations about the relevance and usefulness of empirical data to political philosophy. However, precise examples might prove more convincing than such abstract arguments. In this section, we present two concrete cases in which empirical data can be shown to be relevant to debates in political philosophy.

2.1. Appeal to intuitions and the existence of natural injustice

As mentioned in the previous section, one reason why empirical research can be relevant to political philosophy is that political philosophers sometimes base their arguments on premises that are supposed to be intuitive. One can then investigate to which extent said premise is indeed intuitive, and thus whether it is suitable to present this argument in the context of a public discussion. Doing so implies investigating people's *beliefs* and *judgments* about certain situations. To illustrate this point, let's take an example: the debate about the existence of natural injustice.

Most authors agree that justice and its opposite, injustice, are central concepts to political theorizing. However, there is also some disagreement about the *scope* of these concepts, i.e. the kind of entities they can be aptly applied to. To some, only human actions and the consequences of human actions can count as "unfair" (Anderson, 1999; Hayek, 2012). However, to others, certain states of

affairs can be considered as "unfair" even if human beings have nothing to do in their production (Cohen, 2008; Temkin, 1986).

As Nagel (1997) put it, this is far from a purely conceptual question: saying that something is "unjust" or "unfair" carries a normative force that is not shared by other concepts such as "misfortune". Saying that something is "unfair" often implies that we *should* do something about it, as long as it is possible and not too costly for us to do so. This is why questions about the delineation of the scope of our concept of "injustice" can have important normative implications. For example, among those who consider that only human actions and their consequences can be considered unfair, some consider that only the consequences of actions of *individuals* can be considered unfair, and that the consequences of collective, uncoordinated behaviors do not fall into the scope of the concepts of "justice" and "injustice". This position is mainly defended by classical liberal thinkers, such as Hayek (2012), as one of its consequences is that bad consequences of the current social organization should not be considered as unfair, as long as they are not the predictable product of individual actions.

Claims about the *scope* of injustice are conceptual claims. One way to probe a concept is by studying experts' intuitions about how this concept is properly applied. However, in opposition to concepts like "neutrino", mastering the concept of "injustice" is not reserved to a small set of people. Rather, as "injustice" is part of ordinary language, it seems that most of us are competent experts in how the concept should be applied. Claims about the concept of "injustice" can thus be considered to entail certain predictions about the intuitions of competent speakers. More precisely, if the *scope* of the concept of injustice is restricted to human actions and their consequences, then most competent speakers should feel that it is inappropriate to say that states of affairs that are not caused by human actions are "unjust" or "unfair".

Is it the case, then? To investigate competent speakers' intuitions about the scope of "injustice", we conducted a survey in which participants were presented with the following vignette (Allard & Cova, 2018):

> In the country of Maraland, a strange illness has appeared. When they reach the age of 20 or so, one out of a hundred young adults die suddenly. These young people first start to cough weakly, but after only a few days, they become so weak that they die. These young people hadn't done anything wrong, and the illness is not caused in any way by their behavior.

Depending on the condition, participants were then presented with a different source for the illness:

- *Natural source:* In reality, the illness is caused by genetic factors, which are not predictable at the current state of science.
- *Social source:* In reality, the illness is caused by the pollution resulting from industrialization and technological development, even though nobody knows that this is causing the illness.
- *Individual source:* In reality, the illness is caused by the pollution resulting from the activities of an important firm, even though the chairman and the share-holders of this firm are unaware that they are causing the illness.

In each condition, half of the participants were then told that "scientists had found a way to cure these patients" but that "treating all the ill young people would be so expensive that the society had decided not to do anything". The other half was told there was simply no way to cure these patients.

After reading the vignettes participants were asked to answer the following question: "How unfair do you think this situation is?" (on a scale from 1 = "Not unfair at all" to 7 = "Totally unfair"). They were also asked to choose which of the following statements best described the situation:

- It is unfortunate (but not unfair) that one out of a hundred young people die when they get 20.
- It is unfair (but not unfortunate) that one out of a hundred young people die when they get 20.
- It is both unfair and unfortunate that one out of a hundred young people die when they get 20.
- It is neither unfair nor unfortunate that one out of a hundred young people die when they get 20.

	Treatment	*No treatment*
Nature	5.84 (1.17) 95%	4.91 (2.04) 66%
Society	5.75 (1.43) 93%	5.40 (1.43) 95%
Individual	6.33 (0.98) 100%	6.02 (1.31) 93%

Table 1. Participants' answers depending on the source of the illness (nature, society, individual) and the availability of treatment. First line in each row indicates the mean (and *SD*) for participants' answers to the 7-points scale question. Second line indicates the % of participants who selected a description describing the situation as "unfair". (After Allard & Cova, 2018.)

Participants' answers to both questions for all six conditions are presented in Table 1. As can be seen, a majority of participants considers that the situation is unfair, even when the source of the illness is natural and there is no available treatment. Admittedly, the proportion of participants judging the situation to be unfair is even higher when the illness is produced by society or by an individual firm. But our results clearly speak against the idea that any competent user of the "injustice" concept would refuse to apply it to purely natural consequences.

Thus, arguments that start from the observation that the concept of "injustice" cannot be properly applied to natural events in order to reach the conclusion that the State's duties towards people in need are very limited rest on implausible assumptions about our ordinary concept of "injustice".

2.2. Participants' behavior behind a veil of ignorance

The former section gave the example of a thought experiment that probed directly into people's intuitions on the concept of justice. However, one might fear that naïve intuitions are unreliable, as they may be biased by political or social prejudices that philosophy ought to eliminate. This is why philosophers have developed what Bruner and Lindauer (2020) call "impartiality devices": thought experiments aimed at producing impartial judgments. In this section, we review recent debates regarding the possibility of improving the reliability of one's moral and political intuitions via impartiality devices.

The two most famous of impartiality devices are Rawls' original position, and the impartial spectator, promoted by Adam Smith. In his masterwork, *A Theory of Justice*, Rawls famously introduced the notion of "veil of ignorance". The veil of ignorance is part of what Rawls calls the "original position". In Rawls's original position, we have to imagine that individuals have to choose the principles of justice and social organization that would govern their society, while at the same time ignoring their own social standing, or the kinds of skills and qualities that will be rewarded in this society. On the other hand, according to the idea of the impartial spectator, we are to judge moral situations as a spectator that has both perfect knowledge of every minutia of a moral situation and feels equally benevolent towards all individuals involved. In the context of experimental philosophy, one could argue that having participants make judgments and decisions with the assistance of such "impartiality devices" is an improvement over traditional experimental designs, as it allows experimenters to tap into participants' "pure" and "objective" moral intuitions, without considerations about self-interest intruding.

But is it really the case? In recent years, psychologists and philosophers have evaluated whether putting participants under a veil of ignorance or asking them to adopt the perspective of the impartial spectator would shift their moral judgments. For example, in the context of the COVID-19 pandemic, Huang and colleagues (2021) investigated participants' judgments about resource allocations among patients. They focused on the question of whether age should be taken into account, and whether young

patients should be prioritized over older ones. Unsurprisingly, they found that older participants were less likely than younger ones to accept this principle. However, this difference was much lower when participants were first asked to envision the problem from a veil of ignorance perspective, notably because it led older participants to be more likely to endorse the principle. Huang and colleagues conclude that the veil of ignorance procedure led participants to be less "self-serving" and more impartial in their judgment. This suggests that such a procedure can indeed help us better capture participants' moral principles, by reducing the intrusion of self-serving considerations in their judgments.

However, claiming that impartiality devices improve people's judgment does not mean that they are a perfect means of discovering moral principles. Indeed, if impartiality devices were perfect, we would expect that they would all converge on a unique set of correct ideas; but a study by Aguiar and colleagues (2013) suggest that not all impartiality devices favor the same decisions. In a comparative study, they observed that participants in an "impartial spectator" position made on average more egalitarian decisions than participants "under the veil". Thus, different "impartiality devices" might yield different decisions, raising the question of which "impartiality device" is the most relevant one. Bruner and Lindauer (2020) asked their participants which impartiality they would advise a society to adopt so as to choose their principles of justice. They found that participants showed no marked preference for one of these two impartiality devices over the other.

Despite these limits, impartiality devices do seem to improve the impartiality of our moral intuitions. So do they give the intuitions expected by philosophers? Would participants put under a veil of ignorance choose Rawlsian principles? In a seminal study, Norman Frohlich and his colleagues (1987) tried to reproduce the conditions of impartiality corresponding to Rawls' veil of ignorance within the setting of a lab experiment. They constituted groups of four participants and told them that, at the end of the study, the group would earn some amount of money. However, participants had to decide how the money should be distributed across members of the groups. To do so, they were asked to choose between different principles of justice. The trick was that each participant would be randomly assigned to one of five "income classes" (or "social positions") but did not know in advance which one. Thus,

they could not know in advance whether they would belong to the "top earners" or the "bottom earners" and had to decide on how to distribute the money between the five income classes without having any information on their "social position". This design was supposed to mirror Rawls' "veil of ignorance".

Participants had the choice between four principles of justice:

1) maximizing the average income (a principle akin to utilitarianism),
2) maximizing the average income with a floor constraint (guaranteeing a certain minimum to every member of the group),
3) maximizing the average income with a range constraint (thereby limiting the disparities between members of the group),
4) maximizing the gains of the worst-off (a 'maximin' principle similar to Rawls' principle of difference).

Most groups (25 out of 29) ended up choosing the second principle: maximizing the average income with a floor constraint. None chose the fourth principle.

What conclusions can be drawn from these studies? For their authors, they constitute an objection to Rawls' philosophical argument. Indeed, Rawls predicted that people in the original position, and thus behind a veil of ignorance, would converge on the fourth principle. This is not the case here, as participants clearly thought poorly of this principle. However, participants in Rawls' original position differ from average human beings in significant ways: they are strongly risk-averse and have no idea of their own principles of justice. They can be considered as perfectly egoistical, rational agents. The behavior of such agents in the original position cannot thus be inferred from the behavior of 'real' human beings - rather, Rawls uses decision theory to predict the principle they would converge on.

One might reject Rawls' construction of the original position as strange and arbitrary, but one should understand that, for Rawls, his construction of the original position is precisely constrained by our intuitions (or 'considered judgments'). For Rawls, the original position must be construed in this

particular way because these are the conditions under which the agents in the original position will converge on the principles of justice that best reflect our intuitions about fairness. Thus, it is doubtful to think that one has 'refuted' Rawls simply because real agents under veil of ignorance do not behave as Rawls expected agents in the original position to behave.

Does this mean that such studies and their results are worthless? Not at all. We saw that, for Rawls, the 'right' way to construct the original position is determined by our intuitions about fairness and justice. But such studies provide us with precious information about the kind of principles participants consider to be intuitive. In this case, the results suggest that participants favor a blend of meritocracy and guaranteed minimum. Overall, this suggests that studies investigating how participants behave and make decisions under such impartiality devices constitute a source of relevant information about people's moral convictions.

We now close this chapter by providing a survey of what empirical results can bring to a debate that is central to political philosophy: the debate about theories of justice.

3. Empirically-informed political philosophy and pluralism in theories of justice

Debates on 'folk' or 'naive' theories of justice have a long history. In 19th century philosophy, John Stuart Mill (2002) and Henry Sidgwick (1907) tried to show the inherent contradictions within folk notions of justice, in order to vindicate the need for a radical utilitarian overhaul. In the 20th century, claims about the lack of intuitiveness of utilitarianism were in turn used by deontologists to argue in favour of the rejection of utilitarianism (McCloskey, 1965; Ross, 2003).

In 2001, David Miller published *Principles of Justice*, in which he argues for the importance of studying folk conceptions of social justice for philosophical theorizing. On the basis of a review of the empirical literature, he argues that naive conceptions of justice support a *pluralist* view of justice that integrate three main principles: needs-satisfaction, desert, and equality. According to Miller, which principle should trump the others will depend on the type of human relationship that is relevant to a

given situation. For people in a relation of *solidarity*, the corresponding principle will be the principle of needs-satisfaction. For people in a relation of *instrumental association*, the relevant principle will be the desert principle. Finally, for people in a relation of *citizenship*, equality should be the central principle. To ground his claim, Miller points to studies showing that people tend to choose more egalitarian principles of social justice when making decisions for groups of friends, rather than strangers.

More generally, we can roughly characterize theories of justice based on how much weight they give to the following five factors: (i) protecting individual property rights, (ii) minimizing inequalities, (iii) giving people what they deserve, (iv) satisfying people's basic needs, (v) maximizing efficiency or global welfare. For example, libertarian theories accord maximal weight to protecting individual rights, but give no importance to the other principles. Utilitarian theories only give intrinsic value to welfare maximization, and give mere instrumental value to other principles, such as equality or property rights. Each of these principles has found support in the literature, both in terms of philosophers willing to defend their importance and in terms of empirical data showing that they matter for people's understanding of justice. However, in the context of the present chapter, we will focus on three main theories of justice, due to their prominence in recent debates[4].

First, *sufficiency theories* attribute a high degree of importance to basic needs, and to reaching a minimal level of functioning (Frankfurt, 1987; Huseby, 2010; Waldron, 1986). Sufficiency theories are characterized by the idea that justice is fulfilled when people reach an adequate level of functioning, and argue that egalitarian principles have little value above the level where everyone has reached a sufficiency threshold. Sufficiency theories tend to be silent regarding the place of other constraints, such as desert, in the just society.

[4] We choose to sideline the issue of the importance of property rights, because it has received comparatively little attention in the literature; the most important studies have focused on children's intuition, in order to see how early the concept of ownership arises (Noles et al., 2012; Rossano et al., 2015). In terms of philosophical debates, however, we think that this lack of concern has little import; indeed, almost all viable philosophical theories support the value of property rights. What is debated is whether other principles of justice should be reduced to property rights. In this regard, the importance that people seem to give to desertist or need-based principles refutes the intuitiveness of libertarian theories, which are solely based on individual rights.

Second, *desert-based theories* consider that individuals should be rewarded according to their contribution, or, alternatively, that inequalities are fair if and only if they are deserved. Theories that are only based on desert are relatively rare in contemporary political philosophy. Most desertist theories are actually pluralist theories of justice; for instance, both David Miller (2001) and David Schmidtz (2006) attribute value to the satisfaction of basic needs, desert, and equality.

Third, egalitarian theories of justice consider that the just distribution is the one that either minimizes inequalities (egalitarianism proper; Temkin, 2003) or that maximizes the situation of the worse-off (prioritarianism; Weber, 2014). These two broad categories are united in the recognition that, while basic needs are important, they are not all that matters; and in the claim that inequalities cannot be justified by appeal to desert.

How does each of these theories fare (as descriptions of folk conceptions of justice) when compared to available empirical evidence? In our opinion, the empirical literature supports Miller's claim that people's conception of justice balances several principles, including needs-satisfaction, desert and equality. In this regard, Miller's contention that the folk conception of justice is pluralist is confirmed. However, not all principles receive the same amount of support from empirical evidence. Desert-satisfaction finds maximal support. Needs-satisfaction also seems crucial for folk understanding of justice: at least among minimally deserving citizens, it seems crucial that everyone has their basic needs satisfied. However, folk support for equality is less obvious. Rather, equality seems to function mostly as a baseline in contexts of equally-deserving recipients. This quasi-lexicographic order, with desert-satisfaction being the most important principle, and equality finding the least empirical support, seems to stand in contrast with major egalitarian theories of justice. In the rest of this section, we survey the evidence supporting this claim.

3.1. In terms of avowed ideals, people support pluralistic principles of justice

To study people's intuition about justice, we should make a broad distinction between the avowed principles people defend, and the theory that can be reconstructed from their implicit judgments. The two theories, the implicit one and the explicit one, can be quite different. For instance, in the domain of

retributive justice, people often claim to give strong importance to the consequences of punishment in terms of deterring future crime. However, their implicit judgments tend to be based mostly on retributive motives, that is, based on giving criminals what they deserve, without any regard for future consequences (Carlsmith et al., 2002; Darley et al., 2000).

In terms of explicit theories, people in western democracies seem to adhere to pluralist theories of justice; that is, when asked explicitly, people generally claim that needs-satisfaction, desert, and equality are all important distributive principles. One major piece of evidence for this claim comes from the *Social Value Survey*, a large-scale survey conducted in 1999 among 30,000 citizens of 24 countries in the European Union. The survey asked participants to determine which of the following criteria a society had to fulfill to be considered as just: "Recognizing people on their merits", "Guaranteeing that basic needs are met for all, in terms of food, housing, clothes, education, health", and "Eliminating big inequalities in income between citizens". On average across the 24 countries, 90% of Europeans agreed with the importance of the need criterion, 81% with the desert criterion, and 65% with the equality criterion (Forsé & Parodi, 2006). Thus, all three values seem to receive strong support, with the need criterion being the most popular.

Another international survey providing strong support for pluralism is *Social Inequality IV*, from the *International Social Survey Programme*, in which scientists recruited participants from 41 (mostly European) countries, with a sample of at least 1,000 participants per country. The survey asked participants which criteria they found important for determining wages. The available criteria included both desert-based criteria, such as performance, and need-based criteria, such as the amount of resources necessary to take care of one's family. In this case, desert-based criteria found the strongest degree of support: in each country, they were supported by at least 93% of participants. However, need-based criteria, while less popular, were still deemed important for a large proportion of participants, since at least 56% of participants found them important in any country, the global average being 82% (Allard, 2019).

The discrepancies between the two surveys (need being the most popular criterion in the *Social Value Survey*, desert being the most popular in the *International Social Survey*) is probably due to the difference in framing between the two surveys. First, the desert criterion in the *Social Value Survey* is kept relatively vague ("recognizing merit"), while it is made much more precise in *Social Inequality IV* ("how well he or she does the job"). Second, the first survey is asking about a fair society in general, while the second is asking about a particular case (the fairness of income, thus making salient work-based contributions). This suggests that the respective relevance of the different principles might depend on context.

While these data represent strong evidence for pluralism, they can be subject to criticism, in that they do not necessarily reveal what choices participants would make if judging specific societies and political arrangements: there could be a discrepancy between the criteria for their implicit and explicit judgments. Second, and most importantly, the description of the principles of justice themselves were relatively vague and thus open to diverging interpretations. We now turn to less explicit measures of people's conceptions of justice.

3.2. Disentangling the relationship between desert and needs-satisfaction

The previous section has made evident the importance of needs-satisfaction for social justice. However, solely considering principles formulated in abstract terms obscures implicit assumptions made by participants. Consider participants' extremely high level of agreement on the importance of the following criterion: the State should be "guaranteeing that basic needs are met for all, in terms of food, housing, clothes, education, health". This seems to show support for the unconditional importance of basic needs satisfaction for a just society. However, social science research has consistently shown that support for welfare and for helping needy recipients is actually predicated on the deservingness of the recipients themselves. Desert and needs-satisfaction are not independent criteria, but actually interact. In other words, most people do not consider satisfying citizens' basic needs as an unconditional requirement of justice, but as a right that is conditional on minimal degrees of desert.

For example, in an influential review published in 2000, Wim van Oorschot suggested that support for welfare was predicted by five major dimensions. One category of support for welfare was simply neediness: people should receive support only if they need it. Three other criteria, however, incorporated cues for deservingness. Oorschot argued that people support welfare for recipients who are not responsible for their plight (low control), who are grateful for the support (attitude), and who have worked and contributed in the past (reciprocity). All three criteria can lead to inferences on moral character, or on the fact that welfare recipients would be contributing to global welfare if only they had the chance. In this context, support for the needy seems to be conditional on their deservingness. (Oorschot's fifth criteria, identity, reflects ingroup preference. While there may be some dimension to it that reflects reciprocity considerations, it may also simply reflect pure ingroup bias.)

In the same article, Oorschot also presented participants with different descriptions of welfare recipients. He observed that participants were strongly supportive of a recipient who was not *able* to work, but strongly opposed welfare for a recipient who was not *willing* to work (for converging evidence, see Aarøe & Petersen, 2014). This conditionality of needs-satisfaction on desert is also nicely illustrated by an experiment conducted by Lisa Farwell et Bernard Weiner (2000). One of the most interesting features of their article relies on the fact that they investigated differences in United States participants' support for welfare across the political spectrum. In their experiment, they asked participants to imagine that they were members of a charity that had to allocate resources between two unemployed people. They varied the reason for their unemployment: one of the unemployed persons had been fired due to lack of efforts, while the other one had become unemployed because their company had gone bankrupt. Participants had to rate how much money they would give to each, how much money a typical conservative would give, and how much money a typical liberal would give. Farwell & Weiner found large levels of agreement between conservatives and liberals on this issue: people on both sides of the political spectrum gave less money to the person who had been fired for lack of effort, even though liberals were indeed less reluctant than conservatives to give them money. In the case of the unemployed person who was not responsible for their plight, the authors found no difference between liberals and conservatives: people from both sides considered that the person should

be helped. Moreover, the differences between the two political camps were much smaller than participants expected. This experiment fits within a broader literature that suggests that differences between people from different political orientations tend to be much smaller than expected, once we switch from abstract declarations about general rules to intuitions about particular cases (Aarøe & Petersen, 2014; Graham et al., 2012; Parker et al., 2021).

While needs-satisfaction is dependent on deservingness cues, it still remains a cornerstone of people's intuitions about justice. In a recent series of experiments (Allard & Cova, 2020), we investigated whether concerns for basic needs survived in societies which can be described as broadly meritocratic. We asked participants to imagine one society (the Beta society). The Beta society was described as a perfectly meritocratic society: positions tied to higher wages were described as requiring more effort and talent, and as making a more important contribution to society as a whole. However, the Beta society lacked a welfare system covering citizens' basic needs. Thus, covering one's basic needs required only $1400 a month in Beta society, while the existing minimum wage was set at $1000 a month. After giving participants the choice to raise the minimum wage, we found that 58 to 66% chose to do so. Moreover, a large proportion of participants chose to increase the minimum wage to $1450 a month, that is, just enough to satisfy one's basic needs. Thus, even in a perfectly meritocratic society, most participants still sought to guarantee need-satisfaction, suggesting that the principle enjoys some level of independence from the desert principle.

Together, these results suggest that the need-satisfaction and desert are two considerations that play a major role in people’s conception of justice, and that they interact to some extent.

3.3. Desert and equality

We have seen that needs-satisfaction and desert are considerations that play an important role in people’s conceptions of justice, even though needs-satisfaction might be conditional on desert. But what

about equality? Do people have a concern for equality *per se*, or can this concern be reduced to their concern for needs-satisfaction and deservingness?

The relevance of equality as a principle of social justice has been debated both among philosophers (e.g. Frankfurt, 1987) and psychologists. Indeed, some psychologists have consistently argued that equality is not a principle that people take into account when judging the fairness of a distribution (Konow, 2003; Starmans et al., 2017). We consider here three ways in which equality plays a role in folk conceptions of social justice: (i) as a baseline when there is no way to allocate resources according to merit, (ii) as a constraint on deserved inequalities, and (iii) as a means for a better, though not necessarily fairer, society.

Two studies from the 1990s by Ordóñez et Mellers (1993) and Mitchell et al. (1993) showed the primacy of desert over equality in people's conception of justice. Ordóñez et Mellers showed that, if people had to make trade-offs between satisfying desert and equality, they were largely more likely to prefer unequal meritocratic societies to egalitarian non-meritocratic ones (where a meritocratic society was operationalized as a society with a correlation of 90%, vs 10%, between talents, efforts, and income). In another study, Mitchell and colleagues assigned participants to three conditions, each condition presenting a different level of meritocracy. The participants had to choose between different income distributions with varying degrees of inequality and average income, leading participants to make trade-offs between equality and utility. The main finding of their study is that participants made choices in conformity with Rawls's difference principle under conditions of low or moderate meritocracy, but shifted to maximizing efficiency under conditions of high meritocracy.

These experimental results suggest that desert is a basic principle of justice. Interpreting these results in terms of people's native preference for equality and inequality is trickier. It seems clear that, when it is impossible to allocate resources based on desert, then people consider an egalitarian distribution of resources preferable to other possible distributions. This can be interpreted as participants having a baseline preference for equality, that can be overcome and trumped by considerations about desert.

But equality does not function only as a baseline but also as a constraint on the range of inequalities. Let's return to the experiment described in section 3.2 (Allard & Cova, 2020), where participants had to choose whether to increase the minimum wage in Beta society. In addition to the case we already presented, we introduced several variations. First, we varied the difference between the highest and lowest wages in the society: in some versions, inequalities were *low* while, in others, inequalities were *high*. Moreover, we compared the Beta society to another society, the Alpha society. Alpha society differs from Beta society because it has a robust welfare system, so that $200 a month is enough to cover one's basic needs. Given that the minimum wage in Alpha is also $1000 per month, all workers in Alpha society already have enough money to cover their basic needs. Still, 36 to 50% of participants were willing to increase the minimum wage in Alpha society. Their behavior can be seen as manifesting a preference for equality, as this raise cannot be explained by a concern for need-satisfaction. Indeed, even when we raised the minimum wage to $3000 in a follow-up study, around 40% of participants were still willing to raise the minimum wage.

However, our results also suggest that participants' concern was *limited*. Indeed, most participants who chose to raise the minimum wage did not choose the maximum raise available. Thus, it seems that, even if some participants are trying to limit inequalities, they are also disposed to tolerate high inequalities when societies are described as meritocratic.

Finally, the fact that participants are tolerant of inequalities and make decisions that often sacrifice equality to desert does not mean that participants do not *value* equality. In two studies, Scott et al. (2001) and Michelbach et al. (2003) showed that people evaluate egalitarian societies positively, even in the case of highly meritocratic societies. Still, they observed that only a minority of participants behaved in a way consistent with Rawls's principle of difference or with a strict egalitarian principle when trade-offs were involved: rather, people were often ready to sacrifice equality in favor of efficiency. People thus seem to prefer egalitarian societies, even though they are unwilling to take strong actions (such as raising the minimum wage) in cases of deserved inequalities. These results can be interpreted as indicating that people do value equality, but do not necessarily consider it to be a requirement of justice in cases of meritocratic societies. Indeed, political philosophers have stressed

multiple reasons to value equality independently of justice considerations: more equal distributions can contribute to maximizing total welfare, promote feelings of solidarity or community, or enhance the stability of political systems (Miller, 2001; Scanlon, 2020). Further research should investigate the relative weight of these different considerations when people indicate preferences for egalitarian societies.

4. Conclusion

While psychologists and political scientists have long investigated issues of interest to philosophers, the development of political experimental philosophy has remained limited. This slow progress is surprising, given that political philosophers commonly acknowledge the relevance of empirical data for normative theorizing. In this chapter, we illustrated the importance of empirical data by outlining recent developments in three domains related to theories of justice, where empirical results reinforce or endanger popular philosophical theories. Our first showcase concerned the boundaries of the concept of fairness. While both libertarian thinkers and social egalitarians (Nagel, 1997; Anderson, 1999) have united in rejecting the possible existence of natural inequalities as contrary to our common understanding of the concept, we showed that people actually do consider that unpreventable natural inequalities can be described as unfair. Second, we took the example of impartiality devices, such as the veil of ignorance or the idea of an impartial spectator, to show that such devices are actually efficient at shifting people's intuitions in a more impartial manner. However, it remains unclear whether different impartiality devices lead people to come to different conclusions on principles of justice. Third, we reviewed empirical data on fairness intuitions to show that people conform to a pluralist theory of justice, valuing desert, needs-satisfaction, and equality at the same time. However, distribution according to desert enjoys a privileged position: people mostly care for the satisfaction of basic needs of deserving recipients, and people are willing to tolerate large inequalities in meritocratic contexts. The centrality of desert leads to dilemmas for mainstream egalitarian theorists, as theories that do not

recognize the value of desert could fail to find support or legitimacy within the population (Scheffler, 1992).

References

Aarøe, L., & Petersen, M. B. (2014). Crowding Out Culture: Scandinavians and Americans Agree on Social Welfare in the Face of Deservingness Cues. The Journal of Politics, 76(3), 684–697. https://doi.org/10.1017/S002238161400019X

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. https://doi.org/10.1017/S0266267113000175

Allard, A. (2019). Le mérite: Signification, possibilité et valeur [These de doctorat, Paris 8]. https://www.theses.fr/2019PA080012

Allard, A., & Cova, F. (2018). Les limites de l'injustice. Klesis, 41.

Allard, A., & Cova, F. (2020). Equality Beyond Needs-Satisfaction: An Empirical Investigation. Journal of Applied Philosophy, 37(2), 273–298. https://doi.org/10.1111/japp.12388

Anderson, E. S. (1999). What Is the Point of Equality? Ethics, 109(2), 287–337. https://doi.org/10.1086/233897

Bor, A., Mazepus, H., Bokemper, S. E., & DeScioli, P. (2021). When Should the Majority Rule? Experimental Evidence for Madisonian Judgments in Five Cultures. Journal of Experimental Political Science, 8(1), 41–50. https://doi.org/10.1017/XPS.2020.8

Bruner, J. P. (2018). Decisions Behind the Veil: An Experimental Approach. In Oxford Studies in Experimental Philosophy, Volume 2. Oxford University Press. https://doi.org/10.1093/oso/9780198815259.003.0008

Bruner, J. P., & Lindauer, M. (2020). The varieties of impartiality, or, would an egalitarian endorse the veil? Philosophical Studies, 177(2), 459–477. https://doi.org/10.1007/s11098-018-1190-8

Buckland, L., Lindauer, M., Rodríguez-Arias, D., & Véliz, C. (2021). Testing the Motivational Strength of Positive and Negative Duty Arguments Regarding Global Poverty. Review of Philosophy and Psychology. https://doi.org/10.1007/s13164-021-00555-4

Burke, E. (2014). Revolutionary Writings: Reflections on the Revolution in France and the First Letter on a Regicide Peace (I. Hampsher-Monk, Ed.; Annotated edition). Cambridge University Press.

Carlsmith, K. M., Darley, J. M., & Robinson, P. H. (2002). Why do we punish? Deterrence and just deserts as motives for punishment. Journal of Personality and Social Psychology, 83(2), 284–299. https://doi.org/10.1037/0022-3514.83.2.284

Cohen, G. A. (2008). Rescuing Justice and Equality (First Edition). Harvard University Press.

Darley, J. M., Carlsmith, K. M., & Robinson, P. H. (2000). Incapacitation and Just Deserts as Motives for Punishment. Law and Human Behavior, 24(6), 659–683. https://doi.org/10.1023/A:1005552203727

Davis, R. W., & Preece, J. (2021). Individual Valuing of Social Equality in Political and Personal Relationships. Review of Philosophy and Psychology. https://doi.org/10.1007/s13164-021-00527-8

Farwell, L., & Weiner, B. (2000). Bleeding Hearts and the Heartless: Popular Perceptions of Liberal and Conservative Ideologies. Personality and Social Psychology Bulletin, 26(7), 845–852. https://doi.org/10.1177/0146167200269009

Fischer, J. M., Kane, R., Pereboom, D., & Vargas, M. (2007). Four Views on Free Will. Wiley-Blackwell.

Forsé, M., & Parodi, M. (2006). Justice distributive. Revue de l'OFCE, no 98(3), 213–244.

Frankfurt, H. (1987). Equality as a Moral Ideal. Ethics, 98(1), 21–43. https://doi.org/10.1086/292913

Frohlich, N., Oppenheimer, J. A., & Eavey, C. L. (1987). Choices of Principles of Distributive Justice in Experimental Groups. American Journal of Political Science, 31(3), 606–636. https://doi.org/10.2307/2111285

Graham, J., Nosek, B. A., & Haidt, J. (2012). The Moral Stereotypes of Liberals and Conservatives: Exaggeration of Differences across the Political Spectrum. PLOS ONE, 7(12), e50092. https://doi.org/10.1371/journal.pone.0050092

Greene, J. (2014). Moral Tribes: Emotion, Reason, and the Gap Between Us and Them (Reprint edition). Penguin Books.

Hassoun, N. (2016). Experimental or Empirical Political Philosophy. In A Companion to Experimental Philosophy (pp. 234–246). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118661666.ch16

Hayek, F. A. (2012). Law, Legislation and Liberty. Routledge.

Huang, K., Bernhard, R. M., Barak-Corren, N., Bazerman, M. H., & Greene, J. D. (2021). Veil-of-ignorance reasoning mitigates self-serving bias in resource allocation during the COVID-19 crisis. Judgment & Decision Making, 16(1).

Huseby, R. (2010). Sufficiency: Restated and Defended. Journal of Political Philosophy, 18(2), 178–197. https://doi.org/10.1111/j.1467-9760.2009.00338.x

Kauppinen, A. (2007). The Rise and Fall of Experimental Philosophy. Philosophical Explorations, 10(2), 95–118. https://doi.org/10.1080/13869790701305871

Knobe, J. (2007). Experimental Philosophy. Philosophy Compass, 2(1), 81–92. https://doi.org/10.1111/j.1747-9991.2006.00050.x

Konow, J. (2003). Which Is the Fairest One of All? A Positive Analysis of Justice Theories. Journal of Economic Literature, 41(4), 1188–1239.

Lindauer, M. (2020). Experimental philosophy and the fruitfulness of normative concepts. Philosophical Studies, 177(8), 2129–2152. https://doi.org/10.1007/s11098-019-01302-3

Lindauer, M., & Barry, C. (2017). Moral Judgment and the Duties of Innocent Beneficiaries of Injustice. Review of Philosophy and Psychology, 8(3), 671–686. https://doi.org/10.1007/s13164-016-0329-9

McCloskey, H. J. (1965). A non-utilitarian approach to punishment. Inquiry, 8(1–4), 249–263. https://doi.org/10.1080/00201746508601433

Mercier, H., Dockendorff, M., Majima, Y., Hacquin, A.-S., & Schwartzberg, M. (2021). Intuitions about the epistemic virtues of majority voting. Thinking & Reasoning, 27(3), 445–463. https://doi.org/10.1080/13546783.2020.1857306

Mercier, H., & Landemore, H. (2012). Reasoning Is for Arguing: Understanding the Successes and Failures of Deliberation. Political Psychology, 33(2), 243–258. https://doi.org/10.1111/j.1467-9221.2012.00873.x

Michelbach, P. A., Scott, J. T., Matland, R. E., & Bornstein, B. H. (2003). Doing Rawls Justice: An Experimental Study of Income Distribution Norms. American Journal of Political Science, 47(3), 523–539. https://doi.org/10.2307/3186113

Mill, J. S. (2002). Utilitarianism (G. Sher, Ed.; Second Edition). Hackett Publishing Company, Inc.

Miller, D. (2001). Principles of Social Justice (Third Printing edition). Harvard University Press.

Mitchell, G., Tetlock, P. E., Mellers, B. A., & Ordóñez, L. D. (1993). Judgments of Social Justice: Compromises Between Equality and Efficiency. Journal of Personality and Social Psychology, 65(4), 629–639.

Nagel, T. (1997). Justice and Nature. Oxford Journal of Legal Studies, 17(2), 303–322.

Noles, N. S., Keil, F. C., Bloom, P., & Gelman, S. A. (2012). Children's and Adults' Intuitions about Who Can Own Things. Journal of Cognition and Culture, 12(3–4), 265–286. https://doi.org/10.1163/15685373-12342076

Oorschot, W. van. (2000). Who should get what, and why? On deservingness criteria and the conditionality of solidarity among the public. Policy & Politics, 28(1), 33–48. https://doi.org/10.1332/0305573002500811

Ordóñez, L. D., & Mellers, B. A. (1993). Trade-offs in fairness and preference judgments. In B. A. Mellers & J. Baron (Eds.), Psychological perspectives on justice, Theory and application (Cambridge University Press). Barbara Mellers and Jonathan Baron.

Parker, V. A., Feinberg, M., Tullett, A., & Wilson, A. E. (2021). The Ties that Blind: Misperceptions of the Opponent Fringe and the Miscalibration of Political Contempt. PsyArXiv. https://doi.org/10.31234/osf.io/cr23g

Rawls, J. (1974). The Independence of Moral Theory. Proceedings and Addresses of the American Philosophical Association, 48, 5–22. https://doi.org/10.2307/3129858

Ross, D. (2003). The Right and the Good (P. Stratton-Lake, Ed.; 2nd edition). Clarendon Press.

Rossano, F., Fiedler, L., & Tomasello, M. (2015). Preschoolers' understanding of the role of communication and cooperation in establishing property rights. Developmental Psychology, 51(2), 176–184. https://doi.org/10.1037/a0038493

Scanlon, T. (2018). Why does inequality matter? Oxford University Press.

Scheffler, S. (1992). Responsibility, Reactive Attitudes, and Liberalism in Philosophy and Politics. Philosophy & Public Affairs, 21(4), 299–323.

Schmidtz, D. (2006). The Elements of Justice. Cambridge University Press.

Schoenegger, P. & Grodeck, B. (forthcoming). Concrete Over Abstract: Experimental Evidence of Reflective Equilibrium in Population Ethics. In H. Viciana, F. Aguiar, & A. Gaitan (Eds.), *Issues in Experimental Moral Philosophy*. Routledge.

Scott, J. T., Matland, R. E., Michelbach, P. A., & Bornstein, B. H. (2001). Just Deserts: An Experimental Study of Distributive Justice Norms. American Journal of Political Science, 45(4), 749–767. https://doi.org/10.2307/2669322

Sidgwick, H. (1981). The Methods of Ethics. Hackett Publishing Company.

Singer, P. (1974). Sidgwick and Reflective Equilibrium. The Monist, 58(3), 490–517. https://doi.org/10.5840/monist197458330

Sosa, E. (2007). Experimental philosophy and philosophical intuition. Philosophical Studies, 132(1), 99–107. https://doi.org/10.1007/s11098-006-9050-3

Starmans, C., Sheskin, M., & Bloom, P. (2017). Why people prefer unequal societies. Nature Human Behaviour, 1(4), 1–7. https://doi.org/10.1038/s41562-017-0082

Temkin, L. S. (1986). Inequality. Philosophy & Public Affairs, 15(2), 99–121.

Temkin, L. S. (2003). Egalitarianism Defended. Ethics, 113(4), 764–782. https://doi.org/10.1086/373955

Waldron, J. (1986). John Rawls and the Social Minimum. Journal of Applied Philosophy, 3(1), 21–33.

Weber, M. (2014). Prioritarianism. Philosophy Compass, 9(11), 756–768. https://doi.org/10.1111/phc3.12175