Felix Gerlsbeck
fhg2106@columbia.edu

The Epistemological Analysis of Normative Principles

Abstract: This paper contributes a methodological position on the question of how to evaluate and compare different prescriptive principles in normative political philosophy. In particular, I argue that there is a role for an analysis of the *epistemic assumptions* of these principles, and suggest a framework that allows us to exactly categorize the kind of epistemic assumptions prescriptive principles imply. This makes it possible to compare theories according to their epistemic characteristics. Finally, I argue that for the comparative assessment of normative principles, this method is superior to the distinction between *ideal* and *non-ideal* theory that is sometimes used to categorize how "utopian" a given principle is.

1. Introduction

A recent debate in normative political theory circles around the question of how "utopian," "realistic" or "fact-sensitive" theories of justice can or should be.[1] A common critique of the elaborate construction of normative theories from Plato to Rawls has been that the proposed theoretical social construct is "infeasible," i.e. that it would be impossible to implement, that it would be instantly corrupted by the imperfect people living under it, or that it places unrealistic demands either on individuals or on collective institutions, and that it is therefore pointless to theorize about it. The time of the theorist, in the eyes of these critics, would be better spent on devising realistic practical innovations that improve existing conditions, or on solving questions of political ethics that agitate many people in practice.

[1] Consider, for example, G. A. Cohen, Rescuing Justice and Equality (Cambridge, Mass. ; London: Harvard University Press, 2008). ch. 6 as an example of ideal theory, and the replies in David Miller, "Political Philosophy for Earthlings," Political Theory: Methods and Approaches, eds. David Leopold and Marc Stears (Oxford: Oxford University Press, 2008). or Thomas W. Pogge, "Cohen to the Rescue!," Ratio 21.4 (2008).

Many normative theorists, however, seem undisturbed by criticism of this sort, as they claim a clear distinction between the task of the political philosopher, who deals with the question of what makes a good (or just, or free) society unconcerned with practical concerns, and the applied political theorist, who figures out what aspects of the ideal theory can be applied in practice, and/or whether they can be expected to work.

This debate is also often framed in the terms of *ideal* vs. *non-ideal* theory. The former is normative theory that is free to assume ideal circumstances, i.e. a situation in which there are few or no practical problems with implementation. The most famous example of self-identified ideal theory is perhaps Rawls' *Theory of Justice*, which creates a normative theory of a social order for citizens who are motivated (to some extent) by their own sense of justice. The non-ideal kind, however, is theory that takes into account for instance the possibility that citizens may not behave in spirit of the theory. Here, people are taken as the imperfect beings we can observe everyday, and the theory is designed to fit their characteristics. A famous non-ideal theory of justice may be Hobbes's account, who based his theory on premise that humans are in the last instance purely self-interested and will resort to lethal violence if they perceive themselves to be threatened.

Consequently, it is sometimes claimed that there is a fundamental difference between *ideal* and *non-ideal* theory; and that they are both different *kinds* of intellectual pursuits, with different standards and aims. This goes to the extent that theories proclaimed within either paradigm are said to be incommensurable with each other. Instead of mutually criticizing the respective other approach for being too conservative (non-ideal theory) or too unrealistic (ideal theory), the solution therefore is typically to assert some division of labour between them. Ideal theory is formulating the "free-floating" idea of the good, and non-ideal theory is trying to see how this ideal can be applied in practice, and, if not, what a second-best solution could be.

In this paper I argue that the distinction between *ideal* and *non-ideal* theory and the thesis of incommensurability of approaches are not helpful. All normative political theories are "ideal" in some sense. Therefore, it is useful to examine the *particular* assumptions made by proposed political theories that make them more or less "ideal." Revealing this can lead to some surprising conclusions, where in some sense some purportedly "non-ideal" theories are actually rather utopian, and some alleged "ideal" theories much less so, as I shall hope to show in the following sections. Although the fundamental argument of this paper is meant to be universally applicable to all kinds of ideal assumptions political theories might include[2], this paper concentrates exclusively on *epistemic* assumptions that normative theories of justice make, and how they affect the plausibility of the theory itself.[3] If we have a clear picture of these, we can use this a criterion to assess theories of justice.

The idea that epistemic problems, or informational constraints,[4] have to play an important role in assessing a normative political theory is hardly new. However, a classification of different *kinds* of epistemic assumptions, and their *particular* effects or consequences may lead towards a more rigorous way to compare different theories. It contributes, that is, a view on the question how the *rational* resolution of differences between normative theories might work.[5]

The paper is therefore partly an exercise in clarification of existing normative positions, and partly defends a methodological standpoint about how to do normative political philosophy, especially how to choose among rival theories. I do not claim that for their evaluation the *epistemological* analysis of normative principles is in any way more important than any other analysis of these principles, but I do claim that it may have *a* role, among others, in our comparative assessment of them.

---

[2] e.g. about human behaviour, about factual knowledge, about technological development
[3] This paper does not deal with other assumptions made by normative theories, such as assumptions about whether people will comply with the law, or whether they are motivated by justice.
[4] I shall use these terms interchangeably
[5] it does not, though, take a position on the metaethical issue whether all normative differences are in the end rationally resolvable or not.

2. Epistemic Assumptions and Political Theory

a. Sen's Analysis

Most theories of justice require individuals to know some fact in order to act justly in the way the theory prescribes, or alternatively, they require an institution to know some fact in order to ensure a just order. For example, assume that justice at a minimum requires that every citizen have enough resources to avoid starvation. Assume furthermore that justice requires that a political agency (the state) distribute resources accordingly. This basic theory of justice (sometimes called "sufficientarianism") requires extensive knowledge of (1) how many citizens there are within the state, (2) where they live and how they can be reached, (3) how high their income is, (4) how to raise the money necessary to distribute funds accordingly.

The institutions or individuals tasked with this distribution therefore need to know all of these things, which is a requirement that significantly exceeded the capacities of most governments in history, and still exceeds the capacities of some governments in the world.

All moral principles have some epistemological assumptions. This is also not a new idea. As Sen suggests, moral principles can actually be categorized according to their *informational* characteristics.[6] In other words, normative principles characteristically *require* some types of information to be taken into account, while *forbidding* certain other types of information.

For example: act-utilitarianism *requires* us to take into account information about the happiness of individuals, in particular the expected happiness of individuals.[7] At the same time, it *forbids* the taking into account of other information, such as, for instance, how the present state of the world came about.

---

[6] Amartya K. Sen, "Informational Analysis of Moral Principles," Rational Action: Studies in Philosophy and Social Science, ed. R. Harrison (Cambridge: Cambridge University Press, 1979).
[7] Whether this refers to information about the happiness of *all* people, of *all relevant* people, or some other subset of all people is an open question, but not one that need immediately concern us here. The key point is that utilitarianism requires some knowledge about some people's happiness.

To take another example: a principle that enjoins equality of resources *forbids* the taking into account of particular information about individuals (e.g. their merits or capacities), and *requires* information about both the amount of available resources, and the number of people in the world (or within some other morally relevant boundary).

And a final example: what has come to be known as "luck egalitarianism" *requires* the taking into account of how every individual's own (unforced) actions contributed to which state of the world came about, as well as the available amount of the good to be equitably distributed and the number of recipients in the world. At the same time we are asked to *ignore* information about every individual's characteristics and those actions for which he is not responsible.

b. Why does it matter?

The question is now, why does this matter? Why should we worry about the types of information "necessary" for our normative principles? After all, justice is justice, and if we hold one principle to be valid, should we not have to bite the bullet and accept its requirements, be they moral or practical. The problem is, however, that epistemic assumptions can make a given principle *indeterminate*. This means that under some conditions the principle may not yield a definite answer to a moral question. If a principle we know to be correct requires us to take into account whether fact A or fact B obtains (for instance, say historic facts about how an agent ended up in some morally problematic situation), and we don't know whether A or B obtains, our principle, however unambiguously we hold it, cannot tell us what to do. This is indeterminacy with respect to the *action-guiding* quality of normative principles.

But even if we reduce our requirements and accept that principles do not need to be action-guiding in every instance (hardly any of them are), we have the problem of *evaluative*

indeterminacy. If our evaluation of a state of the world depends on whether fact A or B obtains, and we don't know which of them does obtain, we cannot say whether the state of the world (distribution, set of institutions, etc.) is in fact just, or how closely it approximates justice.

Now it seems that this is not a problem for all kinds of normative theories of justice. With respect to many theories, the *rightness* of a principle does not seem to depend on any features of the outside world. Accordingly, it seems, we need not worry about indeterminacy – even if a principle is indeterminate, why should this matter for our evaluation of that principle? We can of course accept indeterminacy of normative principles, but it should be clear that they cannot in all cases be used for the guidance of moral actions or normative evaluation of different states of the world. This may include hypothetical scenarios – if a theory is completely indeterminate, we might not even be able to fully specify what a perfectly just state of the world would look like. Again, I am not here advocating the position that indeterminacy of normative principles is necessarily a bad thing – in fact, there might be good reasons to hold indeterminate principles. However, the fact that they are indeterminate should not be ignored when we ask whether a principle is right. I now want to illustrate this with respect to three kinds of "methodological approaches" to normative political philosophy that initially may seem to be relatively immune from practical concerns: I shall call them intuitionism, constructivism, and foundationalism.[8]

For intuitionist[9] thinkers, the rightness of a theory, much like in the "exact sciences" derives from its fit with our moral "data," where "data" refers to the moral judgements we

[8] This list is obviously not complete, and Given the space constraints, I am perhaps over-generalizing in the characterization of these theories and I acknowledge that I am probably to some extent representing caricatures here, but the goal is merely to get the methodological point across, not to give a criticism of these political theory approaches.

[9] I am not here referring to the meta-ethical position called "intuitionism" that links the truth of normative principles to human intuitions, but to the position on the methodology of political theory, which is well represented by Cohen, Rescuing Justice and Equality. ch. 6, or Daniel McDermott, "Analytical Political Philosophy," Political Theory: Methods and Approaches, eds. David Leopold and Marc Stears (Oxford: Oxford University Press, 2008).. This is a position about how to find correct principles, not on the question what makes them true.

(intuitively) make with respect to particular cases.[10] For instance, we might think that letting someone die we could easily save is as morally wrong as murder. So we might develop the theory that having any positive or negative causal role in someone's death is morally wrong. But at the same time, we observe that people in the West let people in developing countries starve all the time, and we might think that while this is regrettable, we don't think it is equivalent to murder. Given this new piece of "data" we have to amend our theory – perhaps we qualify the initial statement to the effect that letting someone die who is spatially proximate is wrong, while the badness diminishes with distance. Then we can test this theory against other data, and so on.

We can immediately see how epistemic problems matter with respect to this method: which of our individual moral judgement "data" become relevant is determined by the epistemic implications of the theory. For instance, we might think that there is no moral distinction between people living now and people living in the future, *and* that justice requires that everyone has an equal number of preferences fulfilled. So we build our distributive theory around these data points. But then we observe the epistemic problem that we do not know what the preferences of future generations are or how many people live in the future, so the theory is completely indeterminate: we cannot assess different paths of action or different states of the world today. So we might revise our "data" – maybe we think that since it is too difficult to find out what people's preferences are, this cannot possibly be required by justice, and we choose to distribute some other good. Or we might think that we should keep the preference-fulfilment egalitarianism, but decide that this only applies to present generations, where it is at least not logically impossible to find out what people's preferences are. In other words, even if we are intuitionists, if we become clear about the epistemic requirements of a theory, we might change our moral judgements ("data") on which the theory is based.

[10] Subject to sufficient clarity of mind, internal consistency, or any number of other conditions of "rationality" we might want to impose on the set of all moral intuitions

Constructivist thinkers, among which I here I would count Kant, Rawls and Scanlon[11] and proponents of some forms of virtue ethics, among others, tend to *construct* theories by assuming some moralized conception of an individual or a situation, and then judge theories by whether they would be chosen or followed by this moralized individual, or in the ideal choice situation. For instance, constructivists like Scanlon may think that it is important that a normative theory be reasonable, and so start with the conception of a *reasonable* person – and assess theories by whether they are what a reasonable person would not reject. On some versions of virtue ethics, a perfectly virtuous person takes the place of the reasonable person. Or they would, like Rawls, think that it matters that a theory is *fair*, and therefore judge any theory by whether it would be chosen in a *fair* choice situation, like the Original Position.

Epistemic constraints enter at the characterization of the reasonable person or the ideal choice situation. We might think that for instance, that it is part of what it means to be reasonable to reject a theory that requires us to know something that is practically impossible to know. For instance, consider that we come to the conclusion that a reasonable person would judge a theory optimal which calls for an economic system that maximizes utility compared to all alternatives. Then we might observe that for any human chooser it is practically impossible to know all alternatives to the present system, and much less what their consequences for utility would be – so we might change our opinion about what the reasonable person would choose. Perhaps we then think a reasonable person would choose as normatively best only such an economic system which does not involve strong epistemic requirements, given the state of the world as it is.[12] Again, becoming clear about epistemic requirements may change our opinion about the relative merits of different "bases" of our constructivism.

[11] See for instance, John Rawls, "Kantian Constructivism in Moral Theory," Collected Papers, ed. Samuel Richard Freeman (Cambridge, Mass.; London: Harvard University Press, 1999)., or Thomas Scanlon, What We Owe to Each Other (Cambridge, Mass. ; London: Belknap Press of Harvard University Press, 1998).
[12] Rawls himself would rule out this possibility, because knowledge of the present state of the world is not allowed in the Original Position, and cannot be taken into account by any reasonable chooser.

Finally, what I have summed up here under *foundationalism* are viewpoints that base their normative assessment of theories on their consistency with some ideal that is valuable independently of facts. All Platonic theories, divine- or natural-law theories, as well as theories based on a highly moralized philosophical anthropology belong in this category. If "the Good," having value independently of what anyone thinks about it, requires us to do X, then, it seems, we have to do X, regardless of whether we actually *can* do it, or can even *know* in any given situation what exactly we should do.

Informational constraints in this view cannot, or should not, count as reasons for holding one principle rather than the other. However, they do enter at the level of deriving more specific principles of action or evaluation from the basic principles.

Let's assume we find out that it is wrong to cause the death of a human being – this could be based on revelation, the categorical imperative, or the observation that it is part of the essence of humans to want to maintain human population on earth. Now, consider that we want to morally assess state S, which has the practice of capital punishment. State S justifies this by reference to the alleged fact that this practice keeps the murder rate down, which, so it claims, means the state actually is causally responsible for *fewer* deaths than if it did not have the capital punishment. Thus we face an epistemic uncertainty in the application of the principle (since we have to know what would happen under the different alternatives), *even if* we will not consider that a reason against that more fundamental principle. In this case, the required information could perhaps be obtained,[13] which would remove the uncertainty, and thus enabling a moral judgement – but unless we do know which policy would lead to more deaths, our principle exhibits indeterminacy. In other words, even if we believe firmly that "the state ought not to cause the death of its citizens," we cannot decide whether the principle "the state ought not to actively kill its citizens" and "the state ought to take all measures to

[13] Although for obvious reasons we might want to forgo any *experimental* methods in determining which policy will lead to fewer deaths overall.

prevent killing from occurring," is the correct specific interpretation of the basic principle, when applied to state S.

In this section I hope to have shown how epistemic problems might be worth considering for three different methods of normative political philosophy,[14] in the sense that they do, or ought to, influence our decision between different possible moral principles. In the following, I shall illustrate more clearly how uncovering epistemic assumptions can influence our assessment of the theories.

c. Who has to have the information?

In the examples above of course the indeterminacy results from the fact that we don't know whether A or B obtains, so the obvious course of action, if at all possible, would of course be to obtain that information. In some cases and for some theories this information may be in principle unobtainable, but in many cases the information whether A or B is the case *can* be obtained, even though it is difficult, or costly.[15] This immediately gives rise to the next question: *who* needs to obtain that information?

The answer to this depends on what has come to be known as the question of the "site" of normative political theory.[16] It fundamentally affects the epistemological analysis of normative principles. In some views, political theories give moral obligations to all individuals to act in accordance with the principle. They are, therefore, effectively also theories of individual morality and they morally constrain individual actions.[17] In this case, therefore, it

[14] The Rawlsian concept of "reflective equilibrium" I see as a combination of what I have termed here *intuitionism* and *constructivism*.
[15] More on this in section 3.
[16] On this debate, consider G. A. Cohen, "Where the Action Is: On the Site of Distributive Justice," Philosophy and Public Affairs 26.1 (1997). and Thomas W. Pogge, "On the Site of Distributive Justice: Reflections on Cohen and Murphy," Philosophy and Public Affairs 29.2 (2000).
[17] they are not always complete theories of individual morality, since they only regulate actions that affect the political sphere in some way.

might be the case that every individual needs to personally obtain the relevant information is necessary to act justly.

Other theories again place not place any obligations on determining morally valuable courses of action on all *individuals*. Here, political theory applies to political decisionmakers, institutions or administrative agencies – they need to obtain the relevant information, legislate accordingly, and the duty of the citizen is to comply with the laws.[18] Of course, collecting information is a very different task for a state or its organs than it is for an individual.

Finally, some theories only require the initial setup of the system to be just, and both legislators and citizens are not morally bound by the theory itself. Here, the institutional setup itself constrains actions and guarantees a just outcome. Rousseau's state of the Social Contract may be a case in point. Here the relevant knowledge has to be known only by the lawgiver *qua* designer of the structure of constraints and incentives that pushes citizens in the right direction. We could also imagine an idealized constitutional convention that has to collect the relevant information, and then set up an constitutional order accordingly.

This categorization, by the way, is not meant to imply that epistemic assumptions become less "ideal" the fewer people have to obtain the relevant knowledge, as I will show in the following sections. In fact, some types of knowledge cannot easily be obtained by centralized agencies, or much less by individuals in founding situation.

[18] Rawls' theory could be an example of this. However, he places a duty on individuals to obey and support only a *just* order, so perhaps citizens do have to collect information in order to determine whether they order under which they live is in fact just or not.

3. A classification of epistemic assumptions.

We can classify epistemic constraints on normative theories along two dimensions: according to the *depth* of the problem, which I define as the difficulty to entirely remove the uncertainty, and according to the *conditionality* of a theory, that is, the character and strength of the logical *link* between the theory and the epistemic assumptions. These two dimensions will be explicated below. We can take this information into account when assessing the plausibility of a given normative theory. If we have more exact information about the epistemic assumptions made by theories, we can avoid the compartmentalization between ideal and non-ideal theory, and perhaps rationally resolve conflicts between principles.

There is no linear relationship between these two dimensions: so we can have a theory that requires us to know things that are really difficult to find out, but does not require that information under all circumstances (that is, the epistemic problem will occur only under some specific *conditions*). Conversely, we can have a theory with epistemic requirements that are relatively easy to fulfil, but which are logically linked to the theory (that is, they occur under *all conditions*).

a. Depth

As mentioned above, the "depth" of the epistemological assumption is defined as the difficulty to entirely remove the uncertainty. The concept is not intended to imply that that differences in depth can be represented in a continuous function, nor that there are no qualitative differences in depth. Accordingly, the depth of an epistemic uncertainty cannot be used to accurately quantitatively "measure" the plausibility of political theories. The question of depth is much rather the question what *exactly* is required to remove the uncertainty

associated with the theory, and whether that exact requirement makes it more or less plausible.

The relevant knowledge required by a theory can range from "radically or logically unknowable" via "practically unknowable" to "cost-intensive but conceptually easy to find out," to "self-evident" or "self-revelatory." To illustrate the concept I will now discuss these categories as if they were categorical variables, or, if you will, Weberian ideal types, while admitting that they will probably rarely come up in isolation.

Logically unknowable things

If a political theory demands the logically unknowable from us, it is self-defeating or incoherent. A fact is logically unknowable if there is no possible world in which it could be known. For instance, say a normative theory requires us to know the identities of all possible future individuals (say, in order to assess different paths of action that would affect the future differently). This is logically impossible, because the amount of knowledge required here is infinite. The same goes for theories that, conversely, require us to know the *full* causal history of the state of the world we are in. This again requires going back indefinitely and amassing an infinite amount of knowledge. Given the assumption that gathering knowledge requires a positive amount of time, theories with these epistemic assumptions are completely indeterminate, since one cannot assess different states of the world (or paths of action) unless one spends an infinite amount of time gathering information.

These examples seem rather outlandish, and so it might be doubted whether anyone ever holds them. Consider, however, a theory of intergenerational justice that requires us to distribute utility fairly among all generations, starting with our own. Initially, this sounds like a plausible theory. In order to evaluate a given state of the world, however, we need to know

not only how many people there will be in the future,[19] but we need to know their identity (meaning their "utility function"). Now, even if it were possible to find that out somehow[20] it would be an infinite process to amass that information, and it has to be arbitrarily truncated at some point in order to make evaluation possible.

A different kind of a logically unknowable fact is what technological innovations will take place in the future. This is based on a paradox as far as I know developed by Popper[21]: either we know which things will be discovered in the future, or we do not. If we don't, we just face epistemic uncertainty. If we *do* know what will be discovered in the future, it is not in the future – we basically have to *have* already discovered it. Otherwise, how would we know what it is? Notice that when I say that "I know that cold fusion will be discovered in the future" I am not really stating that I know that fact. Because if I was in fact certain about it, I would have to know *that* and *how* it works (to justify my claim of knowledge), and then I might as well do it right now. The notion of knowing with certainty what will be discovered in the future is incoherent.[22]

Finally, and following a similar logic, a theory faces a logically unknowable fact when it demands of us to know what "a perfectly rational being," or an "omniscient" one, for instance a God, would do. Of course we cannot know what a perfectly rational or an omniscient being would do because knowing that would require us to be perfectly rational or omniscient ourselves. Now one could argue whether omniscience or perfect rationality are logically possible notions, and whether we could imagine a world in which humans built a machine that makes perfectly rational decisions, so this might belong into the category of practically

[19] We could perhaps imagine technological or demographical advances that enable us to specify that number for any given year.
[20] E.g. if biology were to advance so much that accurate predictions about future generations' mindsets could be made.
[21] In Karl R. Popper, The Poverty of Historicism, 2nd ed. (London: Routledge, 2002).
[22] This, by the way, applies not only to technological, but to all "discoveries." For instance, it is logically impossible to know which new planets will be discovered in 50 years' time, or how much oil will be discovered in a 100 years.

unknowable things. In any case, removing the relevant uncertainty when trying to answer this question seems quite difficult, if not impossible.

Logically unknowable facts pose particular problems for "ideal" normative theory, since there is no possible state of the world "ideal" enough so that we could know them. Accordingly, theories for which finding out these "facts" is fundamental (i.e. theories which are necessarily indeterminate without this knowledge) may be incoherent.

Practically unknowable things

The category of practically unknowable things is larger and more complex. Here are things that can presently not be known, but one could imagine a way of obtaining them, even if it does not exist (yet). For instance, until about a century ago, the question how many people actually live in one's jurisdiction was practically unanswerable with any degree of certainty. Only with the advent of the census and similar methods of registration and social control did it become possible to obtain this fact, at which point it moved categories, from "practically unknowable" to "costly to find out."

Important elements in this class of things are firstly, knowledge about mental states, and secondly, things outside the present limits of natural- and social-scientific knowledge – with respect to cause-effect relations or the "laws" of the natural or the social world.

Some theories of social justice or individual morality require us to find out about mental states of individuals. In many cases (e.g. classical utilitarianism), this requires us to know precisely how *happy* a given person is at one given point. Other theories that require knowledge about other peoples' mental states are those that posit the ideals of "recognition" or "flourishing," to the extent that these values represent neither an objective (observable) condition, nor merely a cause of happiness. The most striking problem obtains here when the mental states are somehow to be used cardinally.

For instance, say my normative position is that people desire and therefore deserve to feel that they are recognized by their peers, and that the state should be organized as to maximize the incidence of people enjoying this feeling. Now, unless I define recognition a behavioural (i.e. observable) state, it will be practically impossible to find out when this state of affairs obtains. In response to this, the recognition theorist may say that we can just go ask everyone whether they feel recognized, and how much so. With modern methods of representative sampling and surveying it should not be a problem to obtain this information. One could perhaps make a case that people can in general tell in a dichotomous manner whether they feel recognized or not, but I think any information about the level of recognition in different possible states of affairs would not be particularly reliable. People are mistaken about their feelings, especially with respect to the past or to hypothetical situations. Note that I am not here rejecting the *ideal* of recognition per se, I am pointing out that under some conditions, the principle of *maximizing* recognition is indeterminate.[23]

Now, the uncertainty associated with this category of theories can in theory be removed, for instance through advances in science. We can imagine that we can reliably find out through analysis of electro-chemical processes in the brain whenever people are happy, feel recognized, or fulfilled. Accordingly, it seems entirely possible that in the future we may be able to maximize the overall happiness in principle, if we wanted to do so.

The second class of "facts" in this category are things outside our present knowledge of the natural or social sciences. Causal connections with respect to the natural sciences are much better established than in the social sciences, nevertheless there remains some uncertainty. While we can reliably say when the laws of gravity will apply, highly complex systems like the climate are still incompletely understood. Here, given the present limitations

[23] Note also that this epistemic problem does not arise with all variants of utilitarianism. Consider preference-fulfilment utilitarianism: since preferences are (typically) considered a behavioural category, whatever the person in our grand survey would reply we can take to actually be the case. Here I need not go into the mental states of the individuals.

of our knowledge, we cannot completely remove the uncertainty as to the consequences any given action may have on the climate.

Causal connections with respect to social scientific facts are even less well established, probably since the systems social scientists deal with are even more complex than those of meteorology. For instance, consider we chose the Difference Principle as our theory of distributive justice, and we try to assess different basic structures we observe in the world. Now we can easily find out in which among the ones we see the least-well-off group has the most resources. What we cannot find out with certainty is whether they are as well off *as they could possibly be*. Perhaps some additional redistribution would make them even better off – or the Laffer curve is correct, and that would reduce overall production to such levels that they would be less well off. We cannot, in other words, *know* whether the inequalities we find are justified or not. In addition, if we allow fundamental differences between societies we cannot even judge the different societies against each other. Perhaps in society A *Atlas Shrugged* is widely read, such that in A greater redistribution would lead to the most productive laying down their tools and, much like John Galt, refusing to contribute to the social product. And perhaps in society B, G.A. Cohen's *If You're an Egalitarian, How Come You're so Rich?* is widely read, such that the most productive individuals produce at their maximum not because of their remuneration, but because they believe in justice. When purely looking at the fate of the worst-off groups in A and B, we cannot conclude purely from their levels of resources which one is the more just society. Nevertheless, we could imagine that social science advances, and that at some point, we could obtain this knowledge about societies, again perhaps through advances in neuro-economics or similar disciplines.

Theories that rely at their core on this type of practically unknowable knowledge are "ideal" in the sense that they rely on facts that we know are there, but we cannot presently obtain. The solutions in practice for these problems is either to approximate the knowledge by formulating either probabilistic law statements (someone who has read *Atlas Shrugged* is *more*

likely to withdraw her productive power if there is redistribution than someone who has read *If You're an Egalitarian, How Come You're so Rich?*), or by measuring something else as a proxy for the required facts – for instance we do not measure actual happiness, but people's responses on a four-point scale that asks "how happy are you now compared to last year?" etc. Nevertheless, the discrepancy between the approximation and the true facts means the theories are, *for any one given case*, indeterminate. After all, we could be facing one of those societies in which everyone has read *Atlas Shrugged* but still nobody will be following John Galt's example. Given that the discrepancy also cannot in principle be quantified (because for that I would have to know the true facts, and could dispense with the approximation), we do not know the range of our uncertainty. Therefore claims about knowledge based on either proxy measures or probabilistic laws ought to be treated with extreme caution.

Things that are costly to find out

The third category of knowledge I consider here contains facts that can be obtained, but for which there is a cost attached to finding out about them. Now, this covers almost all kinds of knowledge that *can* actually be obtained, but there are differences both in quality and quantity of this cost. I want to consider three different kinds of problems here: when the obtaining of information is just straight-up costly, where it is self-defeating, and where it violates some other non-material values.

Some normative theories require information the collection of which reduces the overall amount of resources available to a considerable extent. They exhibit "friction costs." For instance, say we believe, as many European welfare states seem to assume, that resources ought to be distributed such that everyone has *enough* to live a certain approved minimally valuable lifestyle. This requires any potential redistributor, agency or individual, to find out many details about the said individual, how many children she has, how large her apartment

is, how much she might have saved under the mattress etc. If she already has an amount of resources that can be considered "sufficient," no transfer payments need to be made. If we want to go deeper into this, we would also have to look at whether she spends her money prudently, such that she leads the approved lifestyle with as few resources as possible. In any case, we can see that the required information can be easily obtained in principle (in that there are no problems in principle how to measure these facts), but that it requires quite a lot of resources to do so. We have to send people round her house periodically to check up, we have to check her bank accounts, and potential other places where money is tucked away (mattress, Switzerland, etc.), check her associations – perhaps she can obtain money from friends or family, and so on.

We can immediately see that the cost of this supervision for every potential recipient of payments is much larger than the resources available to most single individuals. It seems therefore that it would be problematic to hold the above mentioned normative theory as a theory of individual morality – we cannot really expect anyone to spend more than the total of his resources to accurately determine how much he should redistribute. However, as a collective distribution mechanisms (say, by the state) where the costs are shared, the theory seems more feasible. However, we should still not lose sight of the fact that in the process of obtaining the relevant information, there is a deadweight loss of "friction costs." This means that in the end there are fewer resources to distribute. This may not be significant when compared to the total of the resources distributed, but it may influence our decision about which theory to choose.

The second element in this category are theories in which the process of gathering the relevant information undermines the normative justification of the theory itself. A great example of this in Elster's analysis of child custody lawsuits.[24] These have been governed in

[24]Jon Elster, Solomonic Judgements : Studies in the Limitations of Rationality (Cambridge: Cambridge University Press, 1989).

many places for a long time by the principle that the decision ought to be made only in the best interests of the child. Unfortunately the lengthy and adversarial process of determining which solution is in the best interests of the child may itself not be in the interests of the child at all. We need to consider whether the benefit for the child of finding out the "correct" solution outweighs the "cost" of any psychological damages that are obtained by the process of determining that information (i.e. by the trial).

Finally, there are processes of obtaining information that just violate other values we might hold. Say we have become "luck egalitarians," and condemn all inequalities which are not the result of actions for which individuals can be held responsible.[25] In other words, we have to check whether the inequalities are "deserved" or not. When assessing a pattern of distributions therefore we then have to look at how it came about. Distributive patterns are created by aggregate individual actions, and so this involves finding out for given inequality-creating action, whether the individual can be held responsible for it. This may require us to invade their privacy (which we might hold as a value in itself), or curtail their freedom of action (because they may be thinking constantly about being monitored as to the justice of their actions, and act differently than they would have done).

Most luck egalitarians therefore include in their theories a "prerogative" within which individuals can act freely without their actions being judged. It seems therefore that we do not need to find out about what happens within the personal prerogative. This, however, neglects the fact that as long as there is an area of human action that is unobserved, there is uncertainty as to what happens within it. Imagine we grant the prerogative that there is no information to be gathered about whatever happens in a person's home. This of course will not do, since actions within a home can greatly contradict "luck egalitarianism" – whether it is undeserved inequality within the family, or stock trading on the internet. Assume then instead

[25] see, for instance, Larry S. Temkin, "Egalitarianism Defended," Ethics 113 (2003).; for a thorough critical discussion of luck egalitarianism, see Elizabeth Anderson, "What Is the Point of Equality?," Ethics 109 (1999).

that we define the prerogative with respect to categories of action – say individuals are free to choose their career, but not free to choose their salary. In this case, we *do* have to observe all actions of individuals again – if the observed action falls inside the prerogative (choosing occupation), we ignore it for the purposes of justice. However, although we go on to ignore it in assessing the justice of the overall situation, we have to know what the action was.

Self-evident facts

Lastly, there is the fourth qualitative category of knowledge – knowledge the obtaining of which does not have a cost attached.[26] Included among this are firstly simple facts about oneself, like one's own preferences or a non-cardinal assessment of one's own emotions, as well as one's abilities and general life situation.

Secondly, this includes behavioural facts that are easy to obtain, since it is automatically collected and/or published with some other process. For instance, we can at no extra cost obtain the information how may people voted for Proposition 8 in California last November, or how many people have a broadband connection at home, or what Congress decided last week. These facts are self-revealing, since it is an essential part of the process of voting, of selling broadband internet access, or of legislating, to obtain and record this information. As social choice theory tells us, however, we cannot infer from this information what the underlying preferences are, since the choice process itself may distort the link between preferences and behaviour.

Thirdly, there are facts that are revealed through the distribution process itself. A good case in point are auctions. The goal of an auction from the point of view of the seller is to sell a good to the person with the highest reservation price, and the process of an auction reveals just that information. We could imagine someone going around asking everyone how much

[26] actually, there is probably a minimal cost attached to most of the elements in this category.

they are willing to pay for a good that can not be easily valued (say, a rare baseball card), and then put it on the market at the highest price that was mentioned, but it is obvious why an auction is a better way to go about this. Theoretically, these relative valuations are revealed in all perfect market processes, but as we know, they do not operate in this way in actuality. For instance, theoretically, the supply side of the labour market reveals individuals preferences over different jobs – with the least valued jobs theoretically commanding the highest salary. However, as we know, the demand side, as well as power differentials between employers and employees distort this theoretical revelation quite considerably.

These four categories of depth of epistemic assumptions, defined as the difficulty to remove all uncertainty associated with the assumption, are probably exhaustive, although the elements in each of the categories presented here are not. I have merely selected some examples to illustrate the point of each of the categories, according to the role they play in a number of normative political theories. However, I hope to have shown how the epistemological analysis of any given principle of justice with respect to the *depth* of the epistemic assumptions, might work. We can, and should, take these assumptions into account when we reflect on different principles, because, as argued in section 2.b above, this might influence our judgment of what justice actually requires.

b. Conditionality

Now, in response to the categorization just now, an ideal-theorist could try to argue that he does not have to concern himself with these epistemic problems – since ideal theory assumes circumstances under which the epistemic problems will not arise. I propose that we can exactly assess how "ideal" or "utopian" this assumption is in a given case, and, in connection with the "depth" of the problem, use it to assess theories.

In particular, we can see how their normative theories are in fact connected to the respective epistemic requirement. This I have termed the *conditionality*, since it can be defined as the set of *conditions* under which the epistemic problem arises for the theory. Firstly, the gathering of information could be just logically entailed, meaning that under *all* circumstances we need to obtain this knowledge to use the theory. If this is the case, there is no amount of utopian assumptions that will get around the epistemic problem. *Secondly,* the knowledge might only be required under some certain factual circumstances. In this case, the theory is "ideal" to the extent that it can get around the epistemic problems only if these circumstances do not obtain – and we can clearly assess these assumptions, not only whether they are realistic, but also whether they embody plausible normative assumptions. *Finally,* there might be theories that can be modified such that they do not face the epistemic problem at all.

For some theories some knowledge is absolutely essential for the theory to have meaning. Consider the example of Nozick's theory of justice.[27] Roughly speaking, the theory states that a just distribution only obtains if it has come about by only just (read: rightful) transactions from an initial distribution that conformed to the theory of just initial acquisitions. Nozick is rather unclear about what exactly the latter would entail, but this does not matter for the point here. The point is the knowledge of how goods were initially acquired and transferred is absolutely key to our assessment of any possible distribution. There are, as it were, no possible circumstances under which we could dispense with that knowledge.

Imagine then that we are trying to adjudicate whether Nozick's theory is better than some alternative. Let's say we want to find out whether his theory is adequate as an ethical principle governing individual behaviour. In order to act justly towards others, therefore we have to know whether the goods we possess, and the goods the relevant others possess, have been acquired justly, and whether they have throughout been transferred justly. Now, we might think that the depth of this epistemic problem actually is "logically unobtainable" (since

[27] Robert Nozick, Anarchy, State, and Utopia (Oxford,: Basil Blackwell, 1974).

it possibly involves going back in time indefinitely). Accordingly, we might judge that this cannot possibly be a principle of individual morality, since if it were, it would demand the impossible of people. Now, I am sure Nozickians can get around this problem somehow, and I do not here want to give a critique of historical-entitlement-theories of justice, but I hope the process of how to use these categories for assessment should be a little clearer now.

Another class of "linkages" are those that hold only under certain circumstances that can be clearly specified. For instance, as discussed above, the application of the Difference Principle by, say, a constitutional convention, requires extensive knowledge of social scientific facts. I have classed this under practically unobtainable knowledge, so it seems that this *prima facie* constitutes a strong objection against the principle. However, the epistemic problem only arises if we somehow have to actively distribute goods (either by individual or by collective action). Imagine in contrast a "very ideal" society, say one where everyone has read both *A Theory of Justice* and *If You're An Egalitarian, How Come You're So Rich?*. Here everyone would presumably work as hard as she can without accepting more wages than are absolutely (i.e. objectively) necessary to motivate her to do so. In such a society, therefore, the epistemic problem would not arise, since we *know* that by definition the outcome conforms to the Difference Principle. If nobody in society accepts any unjustified inequality-producing goods for herself, the outcome will be in accordance with the principle.

Now we can assess how "ideal" this assumption is: Firstly, we can of course point out that it seems somewhat unrealistic to assume that everyone will in fact behave morally. However, this is not a decisive objection against "ideal" theory: even if murder occurs frequently, we can still justifiably argue that it is wrong. So while the "very ideal society" may not be realistic, it could still be an ideal to which to aspire.

Secondly, however, we can look at the very ideal society more closely and assess what it really means. We can point out that it would, for instance, entail that individuals can no

longer choose their occupation according to their preferences, but must choose it according to their moral requirements – namely that occupation in which they can be most productive given their own skills. This introduces an epistemic problem at the individual level – can we really expect individuals to collect full knowledge about their expected productivity before making a choice of occupation? Or is this knowledge even possible to obtain? If it is not, are people violating their moral requirements if they base their occupational choice on incomplete information? These moral questions, which are an upshot of the epistemic issues, *have to be* addressed by the proposed very ideal theory.

Furthermore, we might think that the assumption of ideal behaviour violates the essential value of freedom – that a world in which people can choose to work in whatever occupation they want is *better* than one in which everyone works in that occupation where she is most productive. So, in response to that, let's say that we allow a personal prerogative that everyone can choose their own occupation, and limit the requirements of justice to something like the *basic structure*. This is of course Rawls's own solution. But then we are back with the initial epistemic problem – the lawgivers who assess different basic structures require extensive social scientific knowledge, which I have classified as *practically impossible* to obtain.

None of these considerations are supposed to defeat the Difference Principle as a principle of distributive justice. They are intended to clarify what the Difference Principle actually entails. This might, of course, influence our judgement of this principle, but it need not do so, if we are prepared to bite the bullet of either indeterminacy due to epistemic problems (when limiting justice to the basic structure) or really rather strong moral constraints on individual behaviour (when assuming the "very ideal" society).

Consider another example to clarify. In his *Critique of the Gotha Programme*, Marx proposed that under Communism distribution will be (or, as some commentators interpret him, ought to be) governed by the principle "from each according to his ability, to each

according to his needs!"[28] Now, on the face of it, this requires every individual to know (1) her ability, which includes how that ability might be most usefully employed, and (2) her needs. One's own abilities and needs I have categorized as costlessly obtainable. This may be disputed since people might, for instance, systematically overestimate what they actually *need*. However, since *need* is not per se a mental state, I believe with sufficient clarity of mind this could be found out quite reliably. So it seems that this principle has no problematic epistemic requirements at all.

A problem arises, however, when we realize that individuals have to figure out how their abilities might be most gainfully employed. This is similar to the problem identified in the "very ideal" society described above, and the instinctive response would perhaps be to then burden the state, or the basic structure with the task of assigning jobs.[29] However, the principles are not the same: in Marx's principle, the total social product necessary for justice is not the *maximum*, but merely the *sufficient amount* to cover everyone's needs. Now, this is still an epistemic problem, but we could imagine a situation in which productivity is so high that everyone's needs are covered if everyone works in their preferred occupation, provided that they give everything they can according to their ability. This obviously does not maximize the social product (this could be achieved by shuffling individuals around into the most productive occupations), but it may cover the needs of all. Note also that this is of course Marx's own view, since the fundamental way Communism will be brought about is by expansion of the productive forces, that is, potential productive capacity to levels at which Communism become the most adequate principle of economic organization (the optimal "relations of production").

We can see that the theory is "ideal" to the extent that it works only if there is quite a high level of productivity, but under this condition it has no problematic epistemic

[28] Karl Marx, "Critique of the Gotha Programme," Karl Marx: Selected Writings, ed. David McLellan (Oxford: Oxford University Press, 2000). p. 615
[29] many nominally "Communist" states did and still do exactly that, with not infrequent bad consequences

requirements. Again, we can use this analysis to aid in our assessment of the principle relative to other possible distributions.

Finally, there are principles that have no, or just minimal, epistemic requirements at all. This is true for some interpretations of Conservative or Minimalist principles. An extreme conservative principle, very roughly, would be one that states that unless there are secure and immediate gains to be had from change, we should do today what we did yesterday. This has no epistemic requirements (beyond remembering what has been done yesterday[30]), precisely because it takes epistemic concerns so seriously. The extreme version I have stated here considers any epistemic uncertainty as a sufficient reason against doing something. There are more moderate (and sensible) interpretations of Conservatism that allow for more change under uncertainty, and they therefore introduce various epistemic concerns, but in principle this could be one way to completely circumvent epistemic problems.

A minimalist principle is one that states that it does not matter what we do, as long as the outcome is stable, or prevents catastrophe, or fulfils some other minimal requirement. Hobbes and Hume both articulated versions of this principle, as for them essentially the search for political arrangements that fulfil some pre-specified ideal is probably self-defeating since it will be unstable. Instead, any arrangement that embodies, in contemporary parlance, a coordination equilibrium will do.[31] These types of principle yield a prescription *only if* there are no epistemic problems. This again makes clear that the epistemic requirements cannot be our only criterion of assessment of a moral theory – the fact that they are relatively immune to uncertainty certainly does not mean that they necessarily are the best principles we can have.

[30] subject to Bayesian updating, if new facts have arisen.
[31] Hobbes adds the claim that the only principle that would makes a coordination equilibrium possible is that of absolute authority concentrated in the hands of a single sovereign with absolute power (individual or assembly).

4. Conclusion - Evaluating evaluative theories

In the examples above, it has already become clear how the epistemological analysis of normative political principles might proceed. It has also been indicated how it could be used in evaluating these principles. This section treats this question is a little more detail. Let us assume that we are faced with some distributive theory *D*, say it is proposed in a university seminar, and we would like to decide whether it is a good principle or not. The first step could be to give what is called an "immanent" critique – do the conclusions follow from the premises, do the different elements of the theory contradict or supplement each other, do the normative recommendations fit the problem identified, etc. etc.?

But assuming *D* clears the hurdles of the immanent critique, we might still want to evaluate it further: is the ideal proposed by *D* a desirable one, are the premises on which it is based realistic or normatively valuable, and, importantly, which implications does it have, and are these implications acceptable? The epistemological analysis contributes to this last question.[32] To return to the brief discussion of methodologies in section 2.b above we can see how the additional information could influence our judgement.

If we are intuitionists in the sense indicated above, we can check whether *D* fits with our intuitions about particular moral questions. Assume then that *D* prima facie contradicts none of our intuitive judgments. Even if this is the case, the epistemic implications of *D* might do so. Perhaps the process of gathering the required information violates some other ideals we hold, like freedom or privacy. Or perhaps *D* requires us to find out something practically unknowable, and we intuitively think "ought implies can." In that case there would be another clash between principle of intuition. Whether in response to that we amend the theory, or we follow some process of "reflective equilibrium," does not matter here. The point

[32] And to some extent it may aid in identifying any internal contradictions *D* might have.

is, that the epistemological analysis of *D* may uncover new contradictions that have to be resolved, and cannot be just ignored by the reference to "ideal theory" (of course, they may be resolved *in favour of* ideal theory, for instance if we drop the "ought implies can" requirement).

If *D* is a constructivist theory, for instance, one based on the idea of a rational moral agent, beyond the immanent question whether the agent would *really* choose *D* there is the question whether the characterization of the rational moral agent is adequate. Let's assume that *D* requires a redistributive scheme that has so much deadweight loss (i.e. "friction") from the gathering of the required information that most people under it will be worse off than in a rough-and-ready (but less just) rule-of-thumb distribution with little informational requirements. Could being a rational (and presumably reasonable) moral agent mean taking this fact into account? The epistemic problem provides a *pro tanto reason* against *D*, and therefore there has to be an argument made why it should be ignored, or it should be considered when setting up the basis of constructivism.

Finally, assume *D* is a foundationalist principle, based, say, on a detailed ideal of human flourishing, which is based in a philosophical anthropology. Then assume that *D* has an informational requirement that is self-defeating – for instance, the collection of the relevant information to ensure the distributive ideal of *D* could take so much time that people do not have time left over to flourish, compared to alternative principles. This does *not* damage the underlying ideal of human flourishing, but it does deny that *D* is the way to realize this ideal. That is, *D* is neither an adequate *action-guiding,* nor an adequate *evaluative* principle derived from the foundational value. Perhaps we should choose principle *D**, which has fewer informational requirements, and so leaves more time to individuals to flourish. The ideal proposed by *D* might still be more attractive than *D**, but since there is no way to realize it without undermining itself, *D** might be the better principle, even *given the foundational ideal.*

These are some illustrations of how the framework of assessing normative principles might work in practice. Instead of the somewhat blunt *ideal/non-ideal* instrument, we can identify the exact assumptions principles involve. The degree to which principles are utopian or unrealistic with respect to epistemic assumptions is determined by the depth of the epistemic uncertainty of the assumption, and the conditions under which it will occur. The analysis also has some interesting outcomes, a few of which I want to point out here, by way of concluding the paper.

Firstly, not all minimal principles are the same. Desert- or rights-based ones like Nozick's have extremely strong epistemic requirements, while coordination-equilibrium theories like Hume's or Hayek's have very low epistemic requirements. While both kinds of theories arrive at similar policy prescriptions, they are based on fundamentally different premises and one kind is clearly more "utopian" than the other with respect to epistemic assumptions.

Secondly, as we have seen, Marx's principle of distribution, in political discourse often cited as a quasi-paradigm of being unrealistic, has almost no epistemic assumptions under the condition of high productivity, especially compared to alternatives that are commonly considered more "reasonable," like utilitarianism or the difference principle. The "from each, to each" principle is idealistic to the extent that it will work only if certain external conditions are fulfilled, namely a sufficient rise in productivity.[33]

And finally, the analysis of conditionality above implies that the circumstances may matter decisively for how problematic epistemic assumptions are. We can even under some circumstances get rid of epistemic problems altogether – for instance, if there was a way to create a perfect labour market determining the adequate wage for each person and each job would no longer be a problem. This fact would move from practically unknowable to self-

[33] The "from each, to each" principle also shares with many other normative theories the assumption that people will by and large abide by the principle.

evident. However, this also indicates conversely that in the absence of a perfect labour market we are faced with the epistemic problem. In other words, claims that the outcome of the *actually existing* labour market is necessarily just should be treated with caution – instead under present conditions this question is practically unanswerable.

WORKS CITED

Anderson, Elizabeth. "What Is the Point of Equality?" Ethics 109 (1999): 287-337.

Cohen, G. A. Rescuing Justice and Equality. Cambridge, Mass. ; London: Harvard University Press, 2008.

---. "Where the Action Is: On the Site of Distributive Justice." Philosophy and Public Affairs 26.1 (1997): 3-30.

Elster, Jon. Solomonic Judgements : Studies in the Limitations of Rationality. Cambridge: Cambridge University Press, 1989.

Marx, Karl. "Critique of the Gotha Programme." Karl Marx: Selected Writings. Ed. David McLellan. vols. Oxford: Oxford University Press, 2000. 610-17.

McDermott, Daniel. "Analytical Political Philosophy." Political Theory: Methods and Approaches. Eds. David Leopold and Marc Stears. vols. Oxford: Oxford University Press, 2008.

Miller, David. "Political Philosophy for Earthlings." Political Theory: Methods and Approaches. Eds. David Leopold and Marc Stears. vols. Oxford: Oxford University Press, 2008.

Nozick, Robert. Anarchy, State, and Utopia. Oxford,: Basil Blackwell, 1974.

Pogge, Thomas W. "Cohen to the Rescue!" Ratio 21.4 (2008): 454-75.

---. "On the Site of Distributive Justice: Reflections on Cohen and Murphy." Philosophy and Public Affairs 29.2 (2000): 137-69.

Popper, Karl R. The Poverty of Historicism. 2nd ed. London: Routledge, 2002.

Rawls, John. "Kantian Constructivism in Moral Theory." Collected Papers. Ed. Samuel Richard Freeman. vols. Cambridge, Mass.; London: Harvard University Press, 1999.

Scanlon, Thomas. What We Owe to Each Other. Cambridge, Mass. ; London: Belknap Press of Harvard University Press, 1998.

Sen, Amartya K. "Informational Analysis of Moral Principles." Rational Action: Studies in Philosophy and Social Science. Ed. R. Harrison. vols. Cambridge: Cambridge University Press, 1979. 115-32.

Temkin, Larry S. "Egalitarianism Defended." Ethics 113 (2003): 764-82.