Utilitarianism for the Error Theorist

François Jaquet

Utilitarianism for the Error Theorist

François Jaquet

Abstract:

The moral error theory has become increasingly popular in recent decades. So much so indeed that a new issue emerged, the so-called "now-what problem": if all our moral beliefs are false, then what should we do with them? So far, philosophers who are interested in this problem have focused their attention on the *mode* of the attitudes we should have with respect to moral propositions. Some have argued that we should keep holding proper moral beliefs; others that we should replace our moral beliefs with fictional attitudes, beliefs in natural facts, or conative attitudes. But all these philosophers have set aside an important question about the *content* of these attitudes: which moral propositions, and more generally which moral theory, should we accept? The present paper addresses this neglected issue, arguing that moral error theorists should adopt a *utilitarian* moral fiction. In other words, they should accept the set of moral principles whose general acceptance would maximize overall well-being.

According to a prominent version of the moral error theory, all atomic moral propositions are false because they entail the existence of categorical reasons while all the reasons we have are hypothetical. Consider the proposition *Torture is wrong* by way of illustration. If the error theory is correct, then this proposition entails that we have a reason not to torture people regardless of our desires. But this implication is false, for all the reasons we have depend on our desires—they are reasons to act in a way that would satisfy our desires. Hence, the proposition *Torture is wrong* is false. And the same can be said about every atomic moral proposition.

The error theory raises the so-called "now-what problem" (Lutz 2014): as error theorists, what should we do with the moral practice—with moral thought and discourse? Obviously, this is not a moral question. Since the error theory entails that there is nothing we morally ought to do, it entails a fortiori that there is nothing we morally ought to do with our moral thought and discourse. The question is prudential and therefore presupposes only that there are things we prudentially ought to do, which may be the case provided that prudential reasons are hypothetical.

This has a crucial consequence: since desires vary across people, prudential reasons fluctuate accordingly, and no solution to the now-what problem will apply to *all* error theorists. The best we can hope for, then, is a solution that works given some

common desires. That there are such desires is of course an empirical hypothesis, but one that is to some extent confirmed by the literature on the now-what problem. For instance, philosophers who take part in this debate want to live in a society whose members cooperate and resolve their conflicts pacifically. The now-what problem concerns only those error theorists who have such standard desires. Unless specified otherwise, the pronoun "we" will hereafter refer to these error theorists.

Another clarification is in order. There are, in fact, two versions of the now-what problem. One question is what we should do with our moral beliefs, as individual error theorists in the present society, broadly made of success theorists. Another is what we should do with our moral beliefs, as individual error theorists in a possible future society mainly made of error theorists. Although less urgent than the former, the latter question has more of an "existential" flavor. For it is at the heart of the widespread worry that wide acceptance of the error theory might lead to the end of civilization. This is the question I will focus on for the purposes of this investigation.

Five main solutions have been proposed to the now-what problem. According to abolitionism, we should simply get rid of our moral beliefs (Hinckfuss, 1987; Garner, 2007; Ingram, 2015). Conservationism is the view that we should, on the contrary, keep holding moral beliefs—although not necessarily the moral beliefs we currently hold (Olson, 2014). According to revisionary fictionalism, we should replace our moral beliefs with fictional attitudes—make-believe rather than believe that torture is wrong (Joyce, 2001). According to revisionary naturalism, we should replace our moral beliefs with beliefs in natural facts—believe, say, that torture causes significant suffering rather than believe that it is wrong (Husi, 2014; Lutz, 2014; Kalf, 2018). Finally, according to revisionary expressivism, we should replace our moral beliefs with conative attitudes—disapprove of torture rather than believe that it is wrong (Svoboda, 2017).

Hereafter, I will assume that abolitionism is false. Why should we not abolish morality if all atomic moral propositions are untrue? Because morality fulfils functions we deem important. Thanks to the categorical nature of its norms (Joyce 2001: 184), or by making cognitively salient certain attitudes when our will is weak (Olson 2014: 195), it bolsters our self-control. But, chiefly, morality helps us to cooperate and resolve our practical conflicts (Nolan *et al*. 2005: 307). Indeed, insofar as we can agree on the moral question of what we ought and ought not to do, we can more easily agree on the practical question of what to do and not to do. Morality therefore contributes to pacifying human societies. Since we want to overcome akrasia and coordinate our behaviours pacifically, we have as many hypothetical reasons to preserve some sort of moral thought.

Philosophers who are interested in the now-what problem have so far focused their attention on the *kind* of attitudes we should replace our moral beliefs with, on *how* we should accept moral propositions. As a result, they have set aside the *content* of these attitudes. Yet, one may wonder *which* moral propositions we should accept and, more generally, which moral theory—that is which set of moral propositions—we should adopt. And, assuming in line with the error theory that all atomic moral propositions are false, these theories may be called "moral fictions". In a sense, thus, while different solutions to the now-what problem diverge on the kind of attitude we should take towards the moral fiction, they all recommend that we adopt a moral fiction. In this paper, I wonder *which* moral fiction we should adopt as individual members of a community of error theorists with standard desires. And I argue that we should adopt a utilitarian fiction.

Although I want to remain as neutral as possible between non-abolitionist solutions to the now-what problem, I will stick to fictionalism for the sake of presentation. On the one hand, this choice seems natural. Since fictionalism explicitly recommends seeing morality as a fiction, a fictionalist framework lends itself well to our enquiry. On the other hand, I have argued elsewhere that all other solutions to the now-what problem ultimately reduce either to abolitionism or to fictionalism (Jaquet & Naar, 2016; Jaquet, 2020).

According to fictionalism, we should replace our moral beliefs with moral make-beliefs. In other words, we should be disposed both to accept some moral propositions in everyday contexts and to reject all moral propositions in more critical contexts—where a context *C1* is more critical than a context *C2* just in case we question in *C1* a kind of attitude that we hold in *C2* and not vice versa (Joyce 2001: 193). For instance, we should be disposed both to accept that torture is wrong when we deliberate about whether to torture someone—a context in which we assume the existence of moral truths—and to deny that torture is wrong in the metaethics classroom—a context in which we question the existence of moral truths. Moral make-beliefs would thus be similar to moral beliefs in what they would dispose us to do in our everyday contexts (namely, accept their content). But they would differ from moral beliefs in what they would dispose us to do in more critical contexts (namely, reject their content).[1]

[1] Some philosophers take this combination of attitudes to be a form of belief rather than make-belief (Olson 2014). I will not take a stand on this issue. If these philosophers are correct, then the view I will assume is actually a form of conservationism rather than fictionalism.

Before stating my argument in Section 2, and then defending it in the rest of the paper, I will now spell out my main claim. In order to make clear in what sense I think we should adopt a utilitarian fiction, I shall introduce a critical distinction in utilitarian theory, between a criterion for rightness and a decision procedure.

1. Utilitarianism: Criterion and Procedure

Utilitarianism is first and foremost a theory of right action. As such, it puts forward a *criterion for rightness*:

(U_C) An action is right if and only if it maximizes overall well-being.

If my action produces at least as much well-being as any alternative action, then my action is right. If, on the contrary, another action produces more well-being than my action, then my action is wrong. In light of this, the fiction consisting in (U_C) would definitely be utilitarian. Be that as it may, when I say that we should adopt a utilitarian moral fiction, this is *not* the fiction I have in mind. Let me make clear what I mean exactly.

As far as everyday actions are concerned, utilitarians do not advise people to act on principle (U_C). And for good reason: such a policy would have disastrous consequences in terms of aggregate well-being. Most often, when action is called for, we cannot properly weigh the costs and benefits of our options, for lack of time. Not to mention the obvious fact that we are frequently prone to self-deception, as well as all sorts of bias. As Krister Bykvist remarks:

> During the Vietnam War, some US military leaders adopted what they called "utilitarian calculation" when they deliberated about how to fight the war. I do not think it is too wild to guess that these calculations were often skewed in favour of American lives. (2010: 95)

Should we try to act on principle (U_C), we would bring about bad outcomes, meaning that we would act wrongly according to principle (U_C) itself. Utilitarianism consequently advises us not to act on its criterion for rightness. This is why it is sometimes described as a "self-effacing" theory.

Fortunately, utilitarians also provide us with a *decision procedure*, a method we can use in everyday life to decide how to act. For the purpose of moral deliberation, we should accept the set of moral propositions whose acceptance would maximize overall well-being. R. M. Hare (1981) famously developed such a two-level version

of utilitarianism. At what he calls the "intuitive level", where we are when we make practical decisions and need our actions to be guided, we should accept simple and workable principles such as *Torture is always wrong*. By contrast, at the "critical level" that we reach in our more considered moments—e.g. when we do normative ethics—we should reject these straightforward principles and accept (U_C) instead.

Interestingly, this contrast between critical and intuitive levels is reminiscent of the distinction we made between belief and make-belief. The contexts in which utilitarians recommend that we accept simple principles are less critical than those in which they advise us to accept (U_C). Indeed, we make an assumption in the former that we question in the latter: some simple moral principles—of the form *Φ-ing is always wrong*—are true. Hence, in the terminology adopted above, when they say we should accept the moral principles whose acceptance would maximize overall well-being, utilitarians *de facto* recommend that we *make-believe* these moral principles. Accordingly, we should make-believe that an action is right if and only if it satisfies the set of moral principles such that well-being would be maximized if we were to accept it. In other words—since it is not too far-fetched to think of this set of principles as a fiction—the utilitarian decision procedure states that:

(U_P) We should make-believe the moral fiction our acceptance of which would maximize overall well-being.

Suppose we would maximize well-being by make-believing the fiction {*Torture is always wrong*, *Lying is always wrong*, *Killing is always wrong*}. Then, (U_P) entails that we should make-believe this fiction.

I will not argue that we should make-believe the utilitarian criterion for rightness, (U_C). For the reasons just listed, such a policy would have suboptimal consequences in terms of desire satisfaction. Instead, I will directly defend the utilitarian decision procedure, (U_P).[2] It is in this sense, and this sense only, that I think fictionalists should go utilitarian.

---

[2] There are two significant differences between the utilitarian's understanding of (U_P) and mine. First, on the utilitarian reading, (U_P) ascribes a duty to make-believe simple moral principles *to all moral agents*. By contrast, in my understanding, (U_P) ascribes such a duty only to moral error theorists in the process of choosing a moral fiction. Second, on the utilitarian reading, (U_P) ascribes us a *moral* duty to make-believe simple moral principles. We have such a duty because we have a more general moral duty to act so as to maximize overall well-being, with which we will comply only if

Must the fiction whose adoption would maximize well-being be called utilitarian? For all we know, it is very much like a list of deontological principles: it would presumably contain propositions such as *Torture is always wrong*, *Stealing is always wrong*, and *Killing is always wrong*, which are rather characteristic of deontological theories. So, why is the present paper not titled "Deontology for the Error Theorist"?

I presume that this concern rests on the notion that only one fiction deserves the label "utilitarian": that which reduces to the principle of utility. But this cannot be right. To begin with, utilitarians themselves would not recommend accepting this principle as a fiction for everyday contexts. As we just saw, they even think such a policy would be wrong because of its effect on aggregate well-being. Rather, they precisely advise us to make-believe the set of moral principles that I recommend.

Besides, some utilitarians explicitly reject (U_C) even in critical contexts—rule utilitarians, for instance, *believe* that an action is right if and only if it satisfies the set of principles whose general acceptance would maximize overall well-being. They accept this set of principles not only as a useful fiction, but as a criterion for rightness as well. If rule-utilitarianism deserves to be called "utilitarian", then so does our moral fiction.

Moreover, this fiction is unlikely to match any typical deontological theory in terms of its content. Deontologists reject many principles utilitarians defend in applied ethics, even if they agree that wide acceptance of these principles would maximize well-being. Think, for instance, about Peter Singer's view that killing an infant is no more objectionable than having an early abortion or about his claim that we should give ten per cent of our income to charities. While the utilitarian fiction would share many principles with deontological theories, it would also contain principles that are seldom, if ever, defended on deontological grounds.[3]

---

we make-believe simple moral principles. In my understanding of (U_P), by contrast, the "should" is prudential.

[3] Our fiction might well differ from utilitarianism in a crucial respect, though. While the utilitarian decision procedure is meant to maximize well-being *tout court*, our fiction would maximize the well-being of moral agents only. After all, the contractors would be moral agents—from the prudential perspective, there is no point in making a contract with someone who is incapable of moral thought. And, being omniscient, they would know that they are moral agents. This is not to say that our fiction would give us no duties to new-borns, the mentally disabled, and non-human animals. But it would give us such duties only because this would maximize the well-being of moral

Now that my claim is clarified, let us turn to the argument.

2. Outline of the Argument

My defence of (U_P) has a form typical of contractarian arguments. I will justify this approach in more detail shortly. For now, suffice it to note that it makes much sense on the face of it: we are wondering which moral fiction error theorists should choose based on prudential considerations; a contractarian framework seems adequate to answer such a question.

My argument appeals to an original position in which contractors choose a moral fiction that will then be adopted in the real world. In this imaginary situation, the contractors do not know which position they will later occupy, because a veil of ignorance separates the original position from the real world. Other than that, they are omniscient. They know how many people there are in the real world; how happy, healthy, and wealthy these people are; what they desire, fear, and aspire to; which relationships they bear to each other; and so on. The only thing the contractors do not know is which among these people they will be once in the real world. Furthermore—and this will play a critical role in the argument—the contractors are ideally rational.

Now that the stage is set, here is the argument:

(1) We should adopt the fiction that the contractors would choose.

(2) The contractors would choose the fiction whose acceptance would maximize their individual expected well-being.

(3) The fiction whose acceptance would maximize the contractors' individual expected well-being is that whose acceptance would maximize overall well-being.

(U_P) Therefore, we should adopt the fiction whose acceptance would maximize overall well-being.[4]

---

agents, most of whom care somewhat about the fate of all sentient beings. In Kantian terminology, these duties would be "indirect"—they would be derived from the interests of moral agents. Interestingly however, all our duties would be indirect—ultimately, all your duties would be derived from your own interests. Thanks to an anonymous referee for pointing out this issue.

[4] As an anonymous referee pointed out to me, the contractors would be indifferent between two fictions that would produce at least as much well-being as any alternative fiction. I agree. However, since it is very unlikely that two fictions would

The remaining seven sections provide evidence for premises (1), (2), and (3) in alternation with possible objections.

3. The Same-Fiction Constraint and the Veil of Ignorance

Assuming the truth of fictionalism, why should each of us specifically adopt the moral fiction that contractors placed in an original position would choose? In a nutshell, here is why: (i) we should all make-believe the same fiction; but (ii) we will fail to do so unless we make-believe a fiction that is at least minimally attractive to us all; and (iii) we will make-believe such a fiction only if we make-believe the fiction that the contractors would choose in the original position. Let me substantiate each of these claims.

Start with the claim that we should all adopt the same fiction. Considering that we often have conflicting desires, one may doubt that—in the hypothetical sense that is relevant in this context—we should all adopt the same fiction. Maybe each of us should rather make-believe the set of moral propositions that would best satisfy their idiosyncratic desires. Still, there are three important reasons why we should all adopt the same fiction.

The first reason has to do with one of the primary functions of morality. As mentioned above, morality is worth preserving largely because it allows us to satisfy a central desire: that of resolving our conflicts pacifically—moral agreement is a first step towards practical agreement. This is broadly why we should preserve morality even though it turns out to be a myth. If we want a tool for conflict resolution, one that helps us to cooperate, then morality seems to be the perfect fit. Now, this will be true only on the condition that we can agree in moral matters. For, as moral abolitionists often insist, morality can also be a great source of conflict when we disagree morally (Mackie 1980: 154; Hinckfuss, 1987: 45). In such cases, it "inflames disputes because moralizing an issue tends to excite and confuse the parties involved" (Garner 2007: 502).

If we were to make-believe different fictions, we would indeed have a harder time resolving our conflicts. Suppose that Jim accepts a fiction in which torture is always wrong while Pam accepts a fiction in which torture is right when it saves lives. And suppose that Jim and Pam face a situation in which an act of torture would save lives—if they do not torture a terrorist, say, a bomb will explode and many

---

maximize well-being, I will keep talking of "the moral fiction" assuming that this is a tolerable simplification.

people will die. Jim will naturally make-believe that they should not torture the terrorist whereas Pam will naturally make-believe that they should torture the terrorist. As a result, they will disagree on what to do. Worse: each will take his or her opponent to be morally corrupted, which will exacerbate their initial conflict. This suggests that, in terms of cooperation and conflict resolution, we would do better without a moral fiction than with conflicting moral fictions.

Things would be different, however, if we all make-believed the same fiction. Suppose that both Jim and Pam accept a fiction in which torture is always wrong. Facing the same circumstances, they will agree that they should not torture the terrorist. And they will more easily agree on what to do thanks to this moral agreement. Since we want to resolve our practical conflicts, this seems to mean we should all adopt the same fiction.

A second reason why we should accept the same fiction is that lasting moral make-belief is arguably possible only on this condition. According to fictionalism's opponents, in cases of moral disagreement, moral make-believers would tend to question the realist assumption they are supposed to make in everyday contexts. They might then slip into a critical context and reject all moral propositions. Sticking to the moral fiction would therefore be difficult in cases of disagreement (Garner 2007: 508-9). Indeed, it is an everyday observation that moral disagreements often morph into metaethical disputes. Here is a case in point. A vegan argues that eating meat is wrong and, after a couple of failed attempts at rebutting her argument, her interlocutor falls back on the metaethical view that ethics is subjective, that everyone is entitled to have their opinion, or that there is simply no truth of the matter—if not on all these views at the same time. This would presumably happen all the more frequently between people who have previously accepted the error theory and make moral judgments only within a fiction.

How often would make-believers disengage from the moral fiction as a result? This partly depends on how often they would disagree about moral matters. And as we have seen, this should seldom happen provided that they would accept the same fiction. For, then, they would easily agree on specific moral issues. Besides, residual disagreements would transparently result from divergences about non-moral issues (e.g. about whether this specific act of torture would save lives), meaning that the inexistence of moral truths would have nothing to do with there being a conflict in the first place. The parties would realize without difficulty that they ultimately disagree about non-moral facts, so that their conflict would not involve any temptation to disengage from the fictional perspective. Far from suggesting that we should stop

having moral attitudes altogether, these considerations thus indicate that we should all adopt the same fiction.

The third reason why we should all adopt the same fiction relates to another objection often raised against fictionalism. On this line of argument, a fictional attitude to morality would not allow us to have genuine moral discussions: if we were to replace our moral beliefs with moral make-beliefs we would end up talking past each other (Svoboda 2017: 15). Imagine two young siblings playing different games of make-belief: the boy make-believes that their father is a king while the girl make-believes that he is an ogre. If asked, the boy will naturally say, "Dad wears a crown and never eats children" while the girl will naturally say, "Dad eats children and never wears a crown." But they would not disagree. They would not be having a genuine discussion about whether or not their father wears a crown or eats children, not even one inside a fiction. Plainly, they would be talking past each other.

The same would be true of two people make-believing different moral fictions. Suppose again that torture is always wrong in Jim's fiction while torture is right when it saves lives in Pam's fiction. Facing a situation in which they could save lives by torturing someone, Jim would naturally say, "We should not do it" while Pam would naturally declare, "We should do it." Yet, immersed in different fictions as they would be, they would not have a genuine disagreement. In fact, they would not even have a proper discussion about the morality of this act of torture. Just like our two siblings, they would be talking past each other, and they would be aware of this. This will be a problem if they rely on moral discourse to coordinate their behaviour.

By contrast, provided that we would make-believe the same moral fiction, we would be able to have genuine moral discussions. Consider again the case of our siblings stipulating now that they both make-believe that their father is a king. Under this new assumption, they will be able to have a meaningful discussion about what he wears and eats. Likewise, if Jim and Pam adopt the same moral fiction, then they will be able to have meaningful moral conversations. One more reason why we should all accept the same set of moral propositions. Assuming—as we presently do for the sake of argument—that we should replace our moral beliefs with moral make-beliefs, we should therefore all make-believe the same moral fiction.

Claim (ii) will not require as extensive a defence. It is indeed fairly evident that we will all adopt the same moral fiction only if we adopt one that is at least minimally attractive to everyone. If someone has nothing to gain from adopting a given fiction, then they will clearly lack any incentive to adopt that fiction. Either they will adopt another, more appealing fiction or—in case all fictions are equally unappealing to them—they will decline to adopt any. Conversely, then, someone will adopt a fiction

only if it is at least minimally attractive to her. And, more generally, any fiction we would all agree on would be at least minimally attractive to each of us.

Let us finally turn to claim (iii), according to which a moral fiction will be minimally attractive for everyone only if the contractors would choose it in the original position. To appreciate how plausible this contention is, imagine what would happen if we were to select a fiction without this constraint, on the mere basis of who we are and what is in our best personal interests. Under such circumstances, it would be rational for us to choose a fiction that is at our own advantage. I, for one, should choose the fiction according to which an action is right just in case it maximizes my well-being. Unfortunately, for reasons obvious enough, this fiction would not even be minimally attractive to anyone else—with the possible exception of my mother.

To summarize the present section: given that (i) we should all adopt the same fiction, that (ii) this will happen only if we adopt a fiction that is at least minimally appealing to everyone, and that (iii) this, in turn, will happen only if we adopt the fiction that the contractors would choose, we can conclude that this is the fiction we should adopt, in accordance with premise (1).

4. The Objection from Irrelevance

Premise (1) states that we should make-believe the fiction that the contractors would choose. Although the original position involves a constraint of impartiality, embodied by the veil of ignorance, this premise rests on the assumption that we should choose a fiction that is in our interests, one that is good for us. Most fictionalists make this assumption. As Joyce puts it, "The right moral [make-beliefs], I understand to be the most useful ones" (2001: 185). Daniel Nolan, Greg Restall and Caroline West concur: "The question of which fiction [is] to be used is best settled by determining which fiction would be most useful to use" (2005: 327).

Still, not everyone agrees that we should make our decision in this way. Don Loeb, for instance, maintains that most error theorists "would not much care about which moral principles it would be advantageous to accept, but would instead want to know which of them would best accommodate their substantive moral concerns" (1996: 230). Admittedly, rational contractors in the original position would choose the fiction whose general acceptance would maximize well-being. But this is irrelevant to *our* choice of a moral fiction *now*, *in the real world*. Instead, we should base this choice on our values and concerns. If we value freedom, we should choose a fiction in which freedom is good, however doing so would ultimately affect our interests. If we care about equality, we should adopt a fiction in which equality matters, whether or not ideally rational contractors would back up this choice. Which

fiction would best satisfy our preferences is irrelevant to which fiction we should adopt. Call this the objection from irrelevance.

I agree with Loeb that our substantive concerns should play a role in the elaboration of the moral fiction. Yet, I think this is consistent with the claim that we should pick a useful fiction. Indeed, because we care about those concerns, they are partly constitutive of our ends. If we care about them a lot, then they will constitute an important part of these ends. In other words, they will be integrated in our well-being function.

Nonetheless, having accepted the error theory, we must recognize that there are no moral truths that these concerns may track, no moral facts that could make them appropriate or correct. As a result, they should contribute to our moral thinking only as some of the constituents of our well-being. For sure, we care about freedom and equality, but there are other things we care about, and all these things contribute to our well-being.

Moreover, we should recognize that other people have different concerns and that these concerns track moral facts no less reliably than ours. Because these people and we ought nonetheless to make-believe the same fiction, we should therefore consent for these different concerns to be treated on a par with ours. Were we to choose a fiction on the basis of our concerns alone, ignoring theirs, why would they want to make-believe this fiction? After all, we would conversely not adopt a fiction that was chosen in total disregard for our values.

In response to the latter point, an objector might want to reject the same-fiction constraint. She would acknowledge the reasons listed in Section 3 in support of this constraint: we want to resolve our conflicts pacifically, to avoid disengaging from the fictional stance, and to be able to discuss moral issues without talking past each other; and we can satisfy these preferences only if we all adopt the same fiction. But she would highlight that we also have other desires, desires that are often in conflict with the fiction chosen by the contractors—which, granting the rest of the argument, would be utilitarian.

Indeed, utilitarianism is a rather demanding theory, implying for instance that we must devote a large portion of our income to alleviating extreme poverty. In so doing, it conflicts with our appetite for unnecessary consumption and material comfort. The worry is that this appetite may be stronger than the preferences whose satisfaction requires that we all adopt the same fiction. Consequently, it would be rational for us to do whatever it takes not to adopt the utilitarian fiction, including breaching the same-fiction constraint if need be. But then, the argument for premise (1) would be undermined.

This objection rests on a shaky assumption, however. It presupposes that we would unconditionally act on the moral principles we would make-believe. And indeed, many people would rather dispense with a tool for conflict-resolution than be committed to always acting on utilitarian principles. But this is a false dilemma. One can consistently accept a moral fiction without abiding by its principles on every occasion. In order for morality to be useful at all, moral make-beliefs should certainly motivate their bearers to act. But they would do so *only to some extent*, just as moral beliefs do. Hence, by adopting the utilitarian fiction, we would not commit ourselves to acting on it all the time.

Of course, accepting this fiction will not be in our best interests on all occasions. It will by and then prompt us to sacrifice an interest of ours for the greater benefit of others. But this is a price worth paying for the huge advantages of partaking in this whole endeavour—and one that is non-negotiable. Should we accept an idiosyncratic fiction, one discarded by the contractors, we would have to forget about these benefits entirely. In sum, we would do better without a fiction than with different fictions; but, by assumption, we would do better with than without a fiction; hence, we would do better with the same fiction than with different fictions. Pending a better reason to discard the same-fiction constraint, premise (1) seems to escape this objection.

5. Rational Choice Behind the Veil

Premise (2) states that the contractors would choose the fiction that would maximize their individual expected well-being. Let me explain what this means. In the original position, the contractors must choose a fiction to be adopted in the real world. These are some of the things they know: (i) how many positions there are in the real world—say, *n*; (ii) the probability they have of being in each of these positions—1/n; and (iii) which amount of well-being they would get in each of these positions given the adoption of each fiction—W(F, P). Based on this information, they can calculate their expected well-being with each fiction: for any fiction *Fi*, each contractor's expected well-being will be: $1/n\,[W(Fi, P1) + W(Fi, P2) + \ldots + W(Fi, Pn)]$. According to premise (2), they would all agree to choose the fiction that maximizes this value.

This premise rests on the following naïve conception of practical rationality. Imagine an ordinary subject who must choose between two options, *O1* and *O2*. She knows that each option will bring her a certain amount of well-being. But she does not know what amount, for this depends on which of two events, *E1* or *E2*, will take place. Still, she knows three things. First, she knows that, from *O1*, she will get W(O1, E1) in case *E1* takes place and W(O1, E2) in case *E2* takes place. Second, she

knows that, from *O2*, she will get W(O2, E1) in case *E1* takes place and W(O2, E2) in case *E2* takes place. Finally, she knows that *E1* and *E2* are equally likely to take place. She can then *expect* a certain amount of well-being from each option: ½ [W(O1, E1) + W(O1, E2)] from option *O1*, and ½ [W(O2, E1) + W(O2, E2)] from option *O2*.

According to the account of rationality in question, it is rational for our subject to choose the option from which she can expect the largest amount of well-being. In other words, she should choose *O1* if ½ [W(O1, E1) + W(O1, E2)] > ½ [W(O2, E1) + W(O2, E2)], or *O2* if ½ [W(O2, E1) + W(O2, E2)] > ½ [W(O1, E1) + W(O1, E2)]. More generally, it is always rational for a subject to choose the option that maximizes their expected well-being.

This conception of rationality supports premise (2). Indeed, since we characterized the original position in such a way that the contractors are ideally rational, it entails that the contractors would choose the option—in this case, the fiction—that maximizes their expected well-being. A contractor who would choose a given fiction while she could expect more well-being from another fiction would not choose the option that maximizes her expected well-being, and would therefore be irrational on this account. There is no room for such people behind the veil of ignorance.

6. The Objection from Idiosyncratic Reasons

Like the criticism discussed in Section 4, the objection from idiosyncratic reasons is based on our values and concerns. But it is addressed at premise (2)—the contractors would choose the fiction that would maximize their expected well-being. As stated in Section 5, this premise rests on the stipulation that the contractors are ideally rational, combined with the claim that a rational subject always chooses the option that maximizes their expected well-being.

The present objector rejects this account of rationality, arguing that it is sometimes rational to choose an option that does not maximize one's expected well-being (Broome 1991: 54-5; Gauthier, 1982). Here is an example. Suppose that Jim is strongly opposed to torture. Alas, a terrorist has planted a bomb in his house. The bomb will be found and deactivated in time only if Jim tortures the terrorist; or else it will explode and Jim's beloved family will die in the event, plunging him in a state of infinite despair. Whether Jim likes it or not, torturing the terrorist is the option that would maximize his expected well-being. Yet, in light of his attachment to human rights, it may be rational for him not to torture the terrorist.

Our values can contribute to determining the rational decision, alongside considerations of expected well-being. Sometimes, they even override those considerations, in which case we should opt for a course of action that fails to maximize our expected well-being. Contrary to premise (2), the contractors might consequently choose a fiction that does not maximize their expected well-being even though they are *ex hypothesi* ideally rational.

This challenge can be dealt with in either of two ways. First, one might stick to the naïve conception of rationality and argue that, despite apparent counterexamples, it is always rational to choose the option that maximizes one's expected well-being. On this line of argument, whatever his values, Jim should torture the terrorist. Second, one might concede that our values may legitimately play a role in ordinary deliberation, and yet insist that this concession does not affect the contractors' choice because *they* would not have idiosyncratic values. Values and preferences based on them should simply be banned from the original position. I think both strategies are promising, assuming—as we are—the truth of the error theory.

Let us consider the former approach, to begin with. According to the error theory's ontological claim, the only reasons we have are hypothetical. They are reasons to act so as to best satisfy our desires. But this basically amounts to saying that the rational thing for you to do is to maximize your well-being (understood in terms of desire satisfaction). So, the naïve conception of rationality appears to follow rather directly from the error theory. From his perspective, if Jim does not know which options will *actually* maximize his well-being, he should choose the option that maximizes his *expected* well-being, never mind his moral ideals.

Consider now the second strategy. Why should personal values and preferences be excluded from the original position? Because of the same-fiction constraint. If our contractors had idiosyncratic values and preferences—that is, values and preferences imported from the real world—, they would not all choose the same fiction. Suppose that Jim and Pam would import their personal values in the original position, Jim being unconditionally opposed to torture while Pam would be favourable to torture to the extent that it saves lives. Jim would choose a fiction in which torture is always wrong, whereas Pam would choose one in which torture is right whenever it saves lives. And the same-fiction constraint would be violated. Because we need our contractors to converge on a single fiction, we must not let them have idiosyncratic values and preferences. But then, they will pick the fiction that maximizes their expected well-being, in line with premise (2).

7. Expected and Overall Utilities

According to premise (3), the fiction whose general adoption would maximize the contractors' individual expected well-being is the fiction whose general adoption would maximize overall well-being. Why is that so?

The reason is straightforward and will not require much positive argument. On the basis of the information the contractors have at their disposal, we can calculate the overall amount of well-being each fiction would bring about: for any fiction *Fi*, overall well-being will be the sum of the individual utilities each position will involve given the adoption of *Fi*, that is: W(Fi, P1) + W(Fi, P2) + … + W(Fi, Pn). Now, as we saw in Section 5, the contractors' expected well-being with each fiction *Fi* is: 1/n [W(Fi, P1) + W(Fi, P2) + … + W(Fi, Pn)]. Interestingly, this means that overall well-being will be strictly proportional to the contractors' expected well-being. Indeed, for any fiction, overall well-being will necessarily amount to the contractors' expected well-being multiplied by *n* (where *n* is a natural number). This being so, the fiction that will maximize the contractors' individual expected well-being will necessarily be that which maximizes overall well-being, in agreement with premise (3).

8. The Objection from Uncertainty Aversion

Premise (3) meets the objection from uncertainty aversion. As this objection is inspired by a criticism that John Rawls addressed at John Harsanyi's defence of a utilitarian theory of justice, let me briefly present Harsanyi's argument and Rawls's objection.

Like mine, Harsanyi's argument has a contractarian structure (1977a: Chap. 4; 1977b). It appeals to an original position, in which contractors choose institutions that will then be set up in the real world. Harsanyi believes that an institution is just if and only if it maximizes overall well-being because he believes: that (i) an institution is just if and only if the contractors would choose it; that (ii) the contractors would choose an institution if and only if it would maximize their individual expected well-being; and that (iii) an institution would maximize their individual expected well-being if and only if it maximizes overall well-being.

Rawls famously rejected claim (iii).[5] In his opinion, this claim holds only if the veil of ignorance is thin enough to let the contractors know that they have an equal

[5] Rawls's contractors are not so much interested in well-being as they are in "primary goods" (1971: 62). I will ignore this difference, which is inconsequential for my purposes.

chance of occupying each position. Absent this piece of knowledge, their (rational) aversion to uncertainty will prompt them to assume that they will occupy the worst position. As a result, the institutions that will maximize their expected well-being will sometimes fail to maximize overall well-being.

Suppose that the contractors must choose between two institutions, *I1* and *I2*, knowing that they might end up in either of two positions, *P1* and *P2*. Suppose also that they would get 100 units of well-being in *P1* and 20 units in *P2* if *I1* were established, while they would get 40 units of well-being in *P1* and 60 units in *P2* if *I2* were established. If they assume that they will occupy the worst position, their expected well-being associated with *I1* will be 20 (that is, 0 x 100 + 1 x 20), whereas their expected well-being associated with *I2* will be 40 (that is, 1 x 40 + 0 x 60), making *I2* the institution that maximizes their expected well-being. But *I2* is *not* the institution that maximizes overall well-being—overall well-being is 120 (that is, 100 + 20) with *I1* and only 100 (that is, 40 + 60) with *I2*. Hence, assuming that the contractors are denied knowledge of probabilities, the institution that maximizes their expected well-being will not necessarily maximize overall well-being, in contradiction with claim (iii). And, of course, Rawls believes that the veil should be thick enough to make this assumption true.

Is the latter belief correct? Some philosophers have accused Rawls of cheating in this regard. In his rather harsh review of *A Theory of Justice*, Hare writes that Rawls denies his contractors knowledge of probabilities "only because it may help to lead by arguments which [he] finds acceptable to conclusions which he finds acceptable" (1989: 169). A more charitable exegesis is possible, however, in which Rawls denies his contractors knowledge of probabilities because of his reliance on the method of reflective equilibrium.

Applied to the contractarian apparatus, the method requires that we build the original position in such a way that it yields intuitive results (Rawls 1971: 20). But if the contractors were allowed knowledge of probabilities, they would systematically choose the institutions that maximize overall well-being. On such a construal of the original position, utilitarianism can be derived from contractarianism (Rawls 1971: 165-6). The problem, according to Rawls, is that utilitarianism has way too counterintuitive implications to be in equilibrium with our intuitions. Not only could it require important inequalities in the distribution of incomes; it would recommend blatantly wrong institutions, such as slavery, should they maximize well-being (1971: 160).

Assuming with Rawls that the best description of the original position is that which survives in reflective equilibrium, that letting the contractors know

probabilities would have utilitarian implications, and that utilitarianism is irreconcilable with common sense, it follows that we should deny our contractors knowledge of probabilities.

Now, putting aside this debate between Harsanyi and Rawls to focus again on the present argument, one may wonder whether Rawls's objection is effective against premise (3) as well. In other words, are *we* entitled to the stipulation that *our* contractors know the probability they have of ending up in each position? Here, I think it is only fair to reverse the onus of proof. Hare seems to be correct to this extent: the burden of justification lies on those who want to deprive the contractors of probabilistic knowledge.[6] After all, the contractors are otherwise omniscient—apart from the fact that they do not know which position they will occupy, which as we saw in Section 3 can be justified via the same-fiction constraint. The more informed they are, the more relevant their choices will be to ours.

We just saw that Rawls discharges this burden of proof by relying on the method of reflective equilibrium. In an original position fashioned in reflective equilibrium, the contractors would not know the probabilities. So, the question really is, should we rely on the method of reflective equilibrium in the present context as well?

And the answer is no. Admittedly, as long as we are after moral truth, it makes sense to seek a reflective equilibrium. If there are moral truths, then we should trust our moral intuitions, which may well be the only access we have to facet of reality. But, in the present context, we are *not* looking for moral truths. Rather, we are working under the assumption that there are no such things. And on this assumption, our moral intuitions are uniformly deceptive: they present us with a reality filled with ethical norms and values, while ours is a morally empty world. By way of consequence, there is no straightforward reason why we should take them into account while building the original position.

Another justification for denying our contractors knowledge of probabilities would be that they simply *cannot* know that they have an equal chance of being in each position. Because knowledge is factive, this would entail that they actually have an equal chance of being in any position. But, the objection continues, this cannot be the case. As a matter of necessity, the contractors can be only in one position: theirs. Although they do not yet know which position they occupy in the real world because they are temporarily behind the veil of ignorance, there is such a position, and they could not but occupy this position after the veil gets lifted.

---

[6] Rawls concedes this point (1971: 166).

There is some appeal to this reasoning provided that we understand positions as persons—or, more generally particulars. For it is true that a contractor could not turn out to be more than one person. A contractor could not turn out to be Jim or Pam. Either he is Jim and then he is necessarily Jim, or else she is Pam and then she is necessarily Pam. But this is not how positions should be construed—they should be construed as bundles of universal properties. On this understanding, to be in a given position merely involves having a certain set of properties that do not depend on the numerical identity of any object. Now, although a contractor could obviously not have different numerical identities, they could definitely end up with different universal properties—they could be happy or sad, wealthy or poor, healthy or sick. In this sense, they could perfectly turn out to occupy different positions, have an equal chance of being in each of these positions, and know that they have an equal chance of being in each of these positions. There is nothing metaphysically suspect with the claim that they would.

Absent a better reason not to, we should let our contractors know that they have an equal chance of being in each position, and accept whatever choice they make in light of this piece of information.

Conclusion

For some philosophers, metaethics should remain neutral with respect to normative ethics (Sumner 1967; Mackie 1977). Others are very explicit about the normative implications of their metaethical stance. Hare (1981), for instance, famously attempted to derive utilitarian conclusions from his universal prescriptivism. Utilitarianism has also been said to follow from the ideal observer theory (Mill 1961) and from contractarianism (Harsanyi 1977a, 1977b). Some have on the contrary taken contractarianism to have non-utilitarian implications (Rawls 1971; Scanlon 1982). All in all, a number of respected metaethical views ground moral theories in the sense that the former's truth would make the latter true.

The error theory, on the other hand, has as a direct consequence that all theories in normative ethics are false. It is therefore *prima facie* unclear how it could support a moral theory rather than another. As Russ Shafer-Landau puts it:

> If [the error theory] can be vindicated, then obviously the prospects for developing a normative ethics are bleak indeed. If there are no truths within morality—only a truth about morality, namely, that its edicts are uniformly untrue—then the enterprise of normative ethics is philosophically bankrupt. Normative ethics is meant to identify the conditions under which actions are

> morally right, and motives morally good or admirable. If nothing is ever morally good or right, then normative ethics loses its point. (2005: 107)

This worry seems warranted. All moral theories have, for some descriptive property *D*, the form "An action is right if, and only if, it instantiates *D*." Thus, while utilitarianism equates *D* with the property of maximizing well-being, Kantianism equates it with the property of not treating someone as a mere means. Since some actions instantiate these properties, it follows from moral theories that some actions are right, which is incompatible with the error theory. Not only the error theory cannot ground a first-order ethical theory; it is inconsistent with all such theories.

Despite these reasons for pessimism, I have tried to make sense of normative ethics for the error theorist. This paper started with a question: Which moral fiction should we adopt after accepting the error theory? In response, I argued that we should adopt the moral fiction that would maximize overall well-being because: we should adopt the moral fiction that ideally rational and informed contractors would choose behind a veil of ignorance; such contractors would choose the moral fiction that would maximize their individual expected well-being; and the fiction that would maximize their individual expected well-being is that which maximizes overall well-being. Assuming the truth of fictionalism, error theorists should go utilitarian.

References

Broome, John. 1991. *Weighing Goods*. Oxford: Basil Blackwell.

Bykvist, Krister. 2010. *Utilitarianism: A guide for the perplexed*. London: Continuum.

Garner, Richard T. 2007. Abolishing morality. *Ethical Theory and Moral Practice 10*(5): 499-513.

Gauthier, David. 1982. On the refutation of utilitarianism. In Miller, H. B. & Williams, W. W. (Eds.), *The limits of utilitarianism* (pp. 144-63). Minneapolis: The University of Minnesota Press.

Hare, Richard M. 1981. *Moral thinking: Its levels, method, and point*. New York: Oxford University Press.

Hare, Richard M. 1989. Rawls's theory of justice. In *Essays in ethical theory* (pp. 145-174). Oxford: Clarendon Press.

Harsanyi, John C. 1977a. *Rational behavior and bargaining equilibrium in games and social situations*. Cambridge: Cambridge University Press.

Harsanyi, John C. 1977b. Morality and the theory of rational behavior. *Social Research 44*(4): 623-56.

Hinckfuss, Ian. 1987. *The moral society: Its structure and effects*. Canberra: Australian National University.

Husi, Stan. 2014. Against Moral Fictionalism. *Journal of Moral Philosophy*, *11*(1), 80-96.

Ingram, Stephen. 2015. After moral error theory, after moral realism. *The Southern Journal of Philosophy*, *53*(2), 227-248.

Jaquet, François. 2020. Sorting out solutions to the now-what problem. *Journal of Ethics and Social Philosophy* 17(3).

Jaquet, François and Hichem Naar. 2016. Moral beliefs for the error theorist? *Ethical Theory and Moral Practice* 19(1): 193–207.

Joyce, Richard. 2001. *The myth of morality*. New York: Cambridge University Press.

Kalf, Wouter. 2018. *Moral error theory*. Palgrave Macmillan.

Loeb, Don. 1996. Must a moral irrealist be a pragmatist?. *American Philosophical Quarterly 33*(2): 225-233.

Lutz, Matt. 2014. The "now what" problem for error theory. *Philosophical Studies 171*(2): 351-371.

Mackie, John L. 1977. *Ethics: Inventing right and wrong*. New York: Penguin.

Mackie, Jhon L. 1980. *Hume's moral theory*. London: Routledge and Kegan Paul.

Mill, Jhon S. 1961. *Utilitarianism*. New York: Oxford University Press.

Nolan, Daniel, Greg Restall, & Caroline West. 2005. Moral fictionalism versus the rest. *Australasian Journal of Philosophy 83*(3): 307-330.

Olson, Jonas. 2014. *Moral error theory: History, critique, defence*. Oxford: OUP.

Rawls, John. 1971. *A theory of justice*. Cambridge: Harvard University Press.

Shafer-Landau, Russ. (2005). Error theory and the possibility of normative ethics. *Philosophical Issues 15*(1): 107-120.

Sumner, Leonard W. 1967. Normative Ethics and Metaethics. *Ethics 77*(2): 95-106.

Scanlon, Thomas M. 1982. Contractualism and utilitarianism. In A. Sen & B. Williams (Eds.), *Utilitarianism and beyond* (pp. 103-128). Cambridge: Cambridge University Press.

Svoboda, Toby. 2017. Why Moral Error Theorists Should Become Revisionary Moral Expressivists. *Journal of Moral Philosophy 14*(1): 48-72.