

Moral Beliefs for the Error Theorist?

François Jaquet & Hichem Naar

This is a penultimate draft.

Please cite the published version that can be found [here](#).

Abstract: The moral error theory holds that moral claims and beliefs, because they commit us to the existence of illusory entities, are systematically false or untrue. It is an open question what we should do with moral thought and discourse once we have become convinced by this view. Until recently, this question had received two main answers. The abolitionist proposed that we should get rid of moral thought altogether. The fictionalist, though he agreed we should eliminate moral beliefs, enjoined us to replace them with attitudes that resemble to some extent the attitudes we have towards pieces of fiction. But there is now a third theory on the market: conservationism, the view that we should keep holding moral beliefs, even though we know them to be false. (According to a fourth theory, ‘substitutionism’, we should modify the content of our moral claims in such a way that they become true.) Putting abolitionism (and substitutionism) aside, our aim is to assess the plausibility of conservationism as an alternative to the – relatively dominant – fictionalism that we find in the literature. Given the difficulty of finding a conservationist view that is both (i) plausible and (ii) not merely a terminological variant of fictionalism, we will argue that conservationism fails to constitute a plausible alternative to fictionalism, at least insofar as it purports to be an alternative view as to what we should do with our moral thoughts.

Keywords: prescriptive metaethics, error theory, fictionalism, belief, make-believe.

Introduction

The moral error theory holds that moral discourse – roughly, judgments predicating moral properties of acts, people, or states of affairs – is irremediably error ridden. The reason is that, just like discourse about witches or phlogiston, moral discourse either conceptually entails or else presupposes the existence of entities that fail to exist or the instantiation of properties that fail to be instantiated. Error theories vary however according to the specific error they attribute to moral judgments. On the most popular of these variants, moral judgments conceptually entail or at least presuppose the existence of categorical reasons, but, as it happens, the world is devoid of any such reasons (Garner, 2006; Joyce, 2001; Mackie, 1977).¹ For the

¹ Other errors one may attribute to moral discourse include, *inter alia*, the commitment to the existence of God, to the existence of free will, or to the possibility of a convergence in desires between fully rational agents (Smith 1994).

sake of the discussion, we will assume the truth of the error theory in at least one of its versions.

One important reason why one might want to resist the error theory is the undesirable consequences its adoption is frequently supposed to have. Indeed, if they come to believe that all moral claims are false, what will prevent people from breaking their promises and lying to, or even killing, each other? Morality seems to be an effective means to the achievement of social cohesion and stability. To the extent that these are things we value, we seem to have hypothetical reasons to value morality as well.² Some error theorists, the abolitionists, reject that contention (Garner, 2007; Greene, 2002; Hinckfuss, 1987). In their view, morality secures inequalities and elitism, it makes practical conflicts difficult to resolve, each participant being convinced she is on the side of the good, and it yields unpleasant feelings, such as guilt and resentment (e.g. Hinckfuss, 1987: Chaps. 3 and 4). Instead of using moral language and thought, whether in our intrapersonal deliberation or while discussing practical matters, we should focus on what we desire and feel, and thus try to find compromises. In short, we should get rid of moral thought and talk altogether. In the present paper, we will assume without argument that abolitionism is false and that morality is a useful practice. We will be interested in those theories that address the worry that acceptance of the error theory may threaten social cohesion and stability.

One such theory is fictionalism, the main alternative to abolitionism. Although the fictionalist agrees with the abolitionist that we should get rid of moral beliefs proper, she nonetheless thinks that we should keep having moral attitudes of some sort: we should adopt, towards moral propositions, attitudes that resemble to some extent the attitudes we have towards pieces of fiction.

It is important to distinguish two types of attitude that pieces of fiction generate. Suppose I comment on the movie *The Blair Witch Project* and say, “There is a witch in the woods.” Quite obviously, my attitude towards the relevant proposition is not best understood as the belief that there is a witch in the woods. What I believe is that, *in the movie*, there is a witch in the woods. My attitude is a genuine belief, but a belief whose content contains the fictional operator ‘in the movie’. Consider now what happens when actress Heather Donahue says, in character, “There is a witch in the woods.” As mine, her attitude is not the belief that there is a witch in the woods. However, neither is it best understood as the belief that, *in the movie*, there is a witch in the woods. Donahue is *playing* the fiction with a standpoint that is somehow internal to it; she is not talking *about* it from an external viewpoint. In some sense, what she does is *make-believe* that there is a witch in the woods. Thus, her attitude is different from an ordinary belief but its content is literally that there is a witch in the woods.

Accordingly, there are two types of moral fictionalism.³ *Content fictionalism* is

² The typical error theorist has no problem with reasons as such. What he finds objectionable are *categorical* reasons. As a consequence, the error theorist will be happy to recognize that we have hypothetical reasons to value the most effective means to our ends, whatever they are.

³ The distinction is often construed as one between two proposals about how we should use moral *discourse*. As Jonas Olson’s critique of Richard Joyce’s proposal, which will be the focus of this paper, is principally at the psychological level, we prefer to construe the distinction as primarily a

the view that we should *believe* that *in the (best) moral fiction* (e.g.) killing is wrong. The attitudes that should replace our present moral beliefs are beliefs as well, but with a fictional operator in their content. Moral thought and talk, revised as prescribed by content fictionalism, would be thought and talk *about* the moral fiction. By contrast, *force fictionalism* holds that we should *make-believe* (in some sense to be made explicit hereafter) that killing is wrong, in a way more or less analogous to that in which Heather Donahue make-believes that there is a witch in the woods. Moral thoughts and talk, revised as prescribed by force fictionalism, are thoughts and talk *in* the moral fiction. In this paper, we will be interested in Richard Joyce's version of the view (Joyce, 2001).

In order to understand what Joyce's proposal that we should stop holding moral beliefs and replace them with moral make-beliefs amounts to, we need to make clear the way he distinguishes belief from make-belief (which he sometimes calls 'pretense' and 'fictive judgment'):

(Belief) S believes that p only if (i) S has assented to p in her most critical contexts, and (ii) S is disposed to assent to p in her most critical contexts. (2001: 192-3)

(Make-belief) S make-believes that p only if (i) S is disposed to assent to p in some contexts, (ii) S has assented to not- p in her most critical contexts, and (iii) S is disposed to assent to not- p in her most critical contexts. (2001: 193)⁴

All this would not be very helpful without an explanation of what it is, for a context, to be (more/less) critical. Here is how Joyce defines the notion:

(Criticality) For any pair of contexts C1 and C2, C1 is more critical than C2 if and only if C1 involves scrutiny and questioning of the kind of attitudes held in C2 but not vice versa. (2001: 193)

The case of acting seems to satisfy Joyce's characterization of make-belief: while playing her character, Heather Donahue assumes the witch's existence, which she questions (and even rejects) when she discusses the movie. Thus, the context of acting is less critical than the one in which she comments on the movie. Now, she is disposed to dissent from the proposition *There is a witch in the woods* in the latter context, so that, according to Joyce's conception of belief, she does not believe that there is a witch in the woods. She is also disposed to assent to that proposition in the context of acting. Hence, according to Joyce's conception of make-belief, she make-believes that there is a witch in the woods. Let's call a subject who make-believes that p , where p is a moral proposition, a 'make-believer' with respect to p , to be contrasted with the 'believer', that is someone who adopts the attitude of belief towards p .

According to Joyce, we should be make-believers with respect to moral propositions, if we are believers with respect to the error theory. Joyce's moral

psychological one.

⁴ These are not intended to be definitions but a characterization of one crucial difference between belief and make-belief (Joyce, personal communication, May 2014). For the sake of simplicity, we will nonetheless call these 'theories'.

fictionalism is thus the view that we should be disposed both to assent to (e.g.) *Killing is wrong* in some contexts and to dissent from that very proposition in our most critical contexts.⁵ Now, the context in which we do metaethics is more critical than the one in which we do practical ethics: when we discuss the morality of abortion or the death penalty, we assume that there is a truth of the matter; but when we do metaethics, we question that assumption.⁶ And it is not the case that when we do practical ethics we question the assumptions we make, or the conclusions we reach, while doing metaethics. Hence, according to Joyce, we should be disposed to both utter first-order moral claims when we do practical ethics and deny them when we do metaethics. And when he says that we should get rid of our moral beliefs, what he means is that we should not be disposed to assent to moral propositions when we do metaethics.

While abolitionism and fictionalism have dominated the debates since the error theory first appeared, a new contender has recently emerged.⁷ Conservationism is the view that we should keep holding moral beliefs just like we used to before we accepted the error theory (Olson 2011).⁸ In line with fictionalism, and against abolitionism, conservationism recommends that we adopt various cognitive, affective, and behavioural dispositions in relation to moral propositions that are highly similar to the ones we had prior to becoming error theorists. However, in contrast with fictionalism, conservationism goes further in recommending that we keep our old moral *beliefs*, and the associated dispositions, as opposed to acquiring attitudes and dispositions that are only similar in various respects to them; as Jonas Olson says, “Moral belief is to be embraced rather than resisted” (Olson 2011, p. 198).

In Sec. 1, we will discuss some objections Olson raises against Joyce’s fictionalism. We will argue that they do not succeed if we grant Joyce his theory of (make-)belief. Then, in Sec. 2, we will examine what Olson has to say in favour of conservationism. As we shall then see, Olson’s case is mainly defensive, focusing as it is on objections Joyce had made in anticipation. Without taking a stand on these issues, we will argue that Olson’s rejoinders are far from decisive.

⁵ Since fictionalism, as we construe it here, is a view about what we should do *once we have accepted the error theory*, there is no need to mention that we should have dissented from the proposition *Killing is wrong* in the past.

⁶ Metaethics is here not construed as a discipline only philosophers are concerned about. Thinking that there might not be objective moral truths *is* doing metaethics, a similar way thinking that there might or might not be a God *is* engaging in metaphysics (or philosophy of religion). It is clear to us that the folk do philosophy every once in a while.

⁷ Another contender is substitutionism, recently defended by Lutz (2014), according to which, roughly, we should modify the content of our moral claims in such a way that they end up being true. Interestingly, Mackie himself may have envisaged just this possibility in his 1976 book, *Problems from Locke*: “A similar conceptual reform [to the one about personal identity I am proposing], rather than mere analysis of our present concepts, is, I believe, needed for ethics. I hope to discuss this topic in another book” (Mackie 1976, p. 196n27). Thanks to an anonymous referee for drawing our attention to this passage.

⁸ See also Brown 2011, for a similar defence.

Finally, in Sec. 3, we will raise a challenge for the conservationist, that of proposing a notion of belief that does not cover Joyce's make-beliefs and makes it plausible that we should keep having moral beliefs. Were the only difference between fictionalism and conservationism to lie in how they name the mental states they both suggest that we should have towards moral propositions, the latter would fail to constitute a distinctive *prescriptive* view. And what we *should* do with our moral talk and thought is what this whole debate is about.⁹

1. Olson's objections to Joyce's fictionalism

In "Getting real about moral fictionalism" (2011), Olson rejects Joyce's fictionalism and argues that we – as error theorists – should keep believing in moral propositions; we should be believers with respect to *both* the error theory *and* moral propositions. Before defending this view, Olson addresses two objections to Joyce's brand of fictionalism. First, he argues that there is a tension in the make-believer's psychology, one that we do not find in the (error-theoretic) believer's psychology. Second, he rejects Joyce's claim that having moral make-beliefs would help us fight weakness of the will. We will see that, as long as we keep Joyce's theory of (make-)belief in mind, Olson's attack is unsuccessful.

First, Olson thinks that the one who develops moral make-beliefs will be in an unstable state of mind:

One should [...] constantly be on one's guard not to slip from moralized [make-belief] into moral belief. Such self-surveillance seems to involve occasionally reminding oneself that morality is fiction. But this reveals a deep practical tension in moral fictionalism, for it also seems that in order for moral precommitments to be effective in bolstering self-control, beliefs to the effect that morality is fiction need to be suppressed or silenced. But suppressing or silencing these beliefs while at the same time acquiring physical and psychological dispositions to behave in accordance with the fictional moral norms makes it all the more likely that one slips from moralized [make-belief] into moral belief. (Olson 2011, p. 197)

Olson's argument here may be interpreted as putting forward a dilemma.¹⁰ On the one hand, (i) the make-believer may never remind herself of the truth of the error theory in her daily life, but then she is likely to slip into full-blown moral belief and thereby cease to be an error theorist (a bit perhaps like a smoker who, after having decided to quit smoking, would automatically light up a cigarette if not

⁹ In a recent article (Suikkanen 2013), Jussi Suikkanen argues, among other things, that *both* fictionalism and conservationism are problematic given that the conceptions of belief on which they rely are inconsistent with the most appealing theories of belief that we find in the literature. While we think that Suikkanen's point is ultimately successful – that the prescriptive views in the debate make implausible assumptions about the nature of belief – our purpose in this paper is much narrower in scope: it is to assess the possibility of defending a genuine conservationist alternative to a relatively influential fictionalist position. This project, we take it, can be pursued even if we think that both fictionalism and conservationism are ultimately inadequate.

¹⁰ Thanks to an anonymous referee for suggesting this interpretation to us.

reminded of her decision). On the other hand, (ii) she may occasionally remind herself of the truth of the error theory in her daily life, but then, given that she is less likely to be motivated to act in accordance with the demands of the moral fiction she accepts in those situations where acting requires self-control on the part of the agent (e.g., when acting would imply a certain sacrifice), she must make sure to toughen up in some way her attitude towards morality, a process that is likely to lead her to acquire full-blown moral beliefs. Either way, Olson argues, the make-believer is likely to end up with moral beliefs, which is inconsistent with fictionalism.

We do not think that Joyce's fictionalism faces Olson's dilemma. But first, given that Olson's target is Joyce's brand of fictionalism, we need to reformulate the dilemma in Joyce's terms. The dilemma is the following. On the one hand, (i*) the make-believer may never remind herself of the truth of the error theory in the *critical contexts* of her daily life,¹¹ but then she is likely to *stop assenting to the error theory in her most critical contexts* (e.g., when doing metaethics), thereby stopping to be an error theorist. On the other hand, (ii*) she may occasionally remind herself of the truth of the error theory in the *critical contexts* of her daily life, but then, given that she is less likely to be motivated to act in accordance with the demands of the moral fiction she accepts in those situations where acting requires self-control on the part of the agent (e.g., when acting would imply a certain sacrifice), she must make sure to toughen up in some way her attitude towards morality, a process that is likely to lead her to assent to moral propositions (and therefore reject the error theory) in her most critical contexts. Now that we are clear about the dilemma that Joyce's fictionalist supposedly faces, let's tackle each of its horns.

The first horn seems to tell us that, assuming Joyce's theory of (make-)belief, a make-believer is likely to stop being an error theorist if she stops entering the critical contexts at which she, as an error theorist, is supposed to reassert her error-theoretic commitment. If, for whatever reason, the make-believer never finds herself in the relevant critical context, the story goes, she is likely to stop believing the error theory. At first sight, it is difficult to see how this claim could follow from Joyce's particular brand of fictionalism. After all, all Joyce's view says is that, in order for you to continue believing in the error theory, it must be true of you that, if you *were* again in the relevant critical context, you *would* reassert your error-theoretic commitment, not that it must be the case that you sometimes reassert your commitment in a critical context. However, it may be the case that, in the absence of a sufficient number of occasions to reassert one's error-theoretic commitment, one is likely to lose one's disposition to do so over time should a critical context present itself. Consider the following situation. As a philosophy major, Bob decides to write a thesis in metaethics, and, after much thought and discussion, comes to the conclusion that the error theory is true. Despite his philosophical talents, Bob decides to become an accountant, never thinking about philosophy (hence metaethics) again. Looking at him twenty years later, we can ask: "What is

¹¹ We take it that, on a plausible interpretation of Joyce's theory of belief, a context at which one reminds oneself one's theoretical commitment on some matter counts as a critical context. It is not because the make-believer does not engage in some serious metaethics when reminding herself of the truth of the error theory that she thereby fails to be in a critical context.

his take on morality?" And it is not clear what the answer would be. In fact, it is an entirely empirical matter whether or not he still believes in the error theory. It seems to be a fact of life that, if you stop reasserting certain commitments, you might lose them over time. The question is whether this fact gives us a good reason to reject fictionalism.

We do not think it does. Granted, an error theorist might over time stop being an error theorist if she stopped thinking critically about morality. The sort of error theorist we have in mind here, however, is not any old student of philosophy who, after having defended a view in a term paper, goes on to live their life as if they never worked on that paper. The error theorist at issue is someone who is genuinely concerned about the personal and social implications of her commitment, about how she should conduct her life, about how she should behave with others, including believers, about how widespread she wishes the error theory to be, and so on. In short, the error theorist at issue is *not* likely to stop thinking about metaethics over such a long stretch of time that it becomes reasonable to suppose that she is not an error theorist anymore.

Perhaps Olson's first horn makes a stronger empirical hypothesis than the one just discussed and rejected. On this hypothesis, it is not simply the case that the absence of critical contexts is likely to lead to a change in belief; having *moral* make-beliefs, with their associated cognitive, affective and behavioural dispositions, is likely to lead to the adoption of full-blown moral beliefs in the absence of a sufficient number of critical contexts over a significant period of time. Although this hypothesis is plausible as far as it goes – in fact, we take it that even Joyce himself may be happy to embrace it – we do not think it ultimately constitutes a problem for Joyce's fictionalism. First off, it may seem that the fictionalist could again appeal to the plausible claim that the make-believer that we are interested in is likely to be concerned about metaethics to such an extent that she is likely to encounter a sufficient number of critical contexts for her belief in the error theory to be preserved. This reply would be insufficient, however, as it ignores the key thought underlying the objection, namely that the nature of the dispositions involved in moral make-beliefs is such that it might over time influence the make-believer's metaethical views. Consider Maria who, because she accepts the error theory, intends to merely make-believe moral propositions. Being a highly emotive person, however, she is regularly overwhelmed by emotions, such as guilt, pity, or indignation, to such an extent that she comes to express doubts about the error theory. There are at least two possible variants of the case, none of which threatens Joyce's proposal. One possibility is that Maria is *not* in a critical context when she expresses her doubts, as long as she feels emotional – she does not genuinely question the assumption that there are moral facts and truths. Once this moment is over, Maria may be able to enter in a critical context again, and reassert the error theory. Second, Maria's emotions might lead her to doubt the error theory even when she is *in a critical context*. In this case, given that she is in a genuine critical context, she would be taking her emotions as (potential or actual) *evidence* that the error theory is false. Nothing on Joyce's proposal prevents the make-believer from *changing her mind* with respect to metaethical issues, of course, and to become a moral believer again. The question at issue here is, what should we do with our moral thoughts and discourse once we have accepted the error theory? And the answer is presumably *not* 'stay away from considerations that may ultimately make us change our minds'. The error theorist, as far as she is

aware of her epistemic fallibility, should certainly be open-minded about this. We can conclude that the fictionalist is not impaled on the first horn of the dilemma.

Is the fictionalist impaled on the second horn of the dilemma? The second horn tells us that, if the make-believer is to reassert her error-theoretic commitment in her daily life, she is likely to lose her motivation to act in situations where some degree of self-control is required by the moral fiction. Tempted to steal the money in front of him, Frank desperately searches for reasons for stealing until he hits what he takes to be a conclusive reason to indulge in the act: since no act is really wrong, he will not do anything wrong if he steals the money. Reminding himself of the truth of the error theory led him to perform an act he would not have done if his first-order moral attitude were stronger. This is a possible case, to be sure, but the question is not whether moral make-belief can come in motivationally weak forms; the question is whether moral make-belief is *by nature* motivationally weak. If moral make-beliefs could sometimes be motivationally strong (after some training, perhaps), then the fictionalist would be able to reject the second horn of the dilemma by saying that the sort of make-belief that we should have towards moral propositions is of the motivationally strong sort. Whether or not this is a plausible reply turns on whether the fictionalist has an adequate response to Olson's second objection to fictionalism, to which we now turn.

On Joyce's view, moral make-beliefs help us fight weakness of the will, just like our current moral beliefs do. According to Joyce, make-beliefs can have a significant weight in our practical deliberation in that they are capable of generating motivational states. He gives three examples of make-beliefs that may be thought to generate motivation: "[W]atching *The Blair Witch Project* may lead one to cancel the planned camping trip in the woods" (2001, p. 303); thinking I ought to do fifty sit-ups every day may help me do the number of sit-ups I really ought to do in order to stay in good shape – more-or-less fifty more-or-less each day (2001, p. 215); and, after the brutal death of a friend due to his alcohol abuse, a drunkard may think, "If I drink this will happen to me" (2001, p. 216).

Olson does not think that make-beliefs are motivational in the way Joyce suggests, however. Interestingly enough, he only discusses the first example. The make-beliefs that *The Blair Witch Project* elicits, he argues, do not generate motivational states; however, the movie yields a genuine belief that does:

[U]pon watching *The Blair Witch Project* I may come to *believe* that (there is at least a possibility that) a crazy serial killer roams the woods. And these *beliefs* about *reality* [...] may serve as partial explanations of subsequent behaviour. But the moral fictionalist's contention is that engagement with moral fiction can have a bearing on motivation and behaviour without prompting false moral belief about reality. (2011, p. 195)

Make-beliefs, even when conjoined with desires, are never sufficient for motivation, or so the objection seems to be. A genuine belief is needed for this.

However, that my attitude that there may be a killer in the woods is best construed as a belief seems to depend on what its content precisely is. Suppose its content is that there *possibly* is a killer in the woods, as Olson's very formulation suggests. There are two problems with this reading. First, it is unlikely for this belief to be instilled by the movie, because this is a belief that everyone is likely to

have in advance of watching the movie. It is certainly possible for there to be a killer in the woods. The movie will not be responsible for generating a belief with this particular content. Second, and more importantly, it is doubtful that the attitude would generate in me the relevant motivational state. Most people go camping even if they know there might be a killer in the woods. If we are to cancel our camping plans it must be because we think there is more than a mere possibility that there is a killer in the woods. We need that to be *likely*, to some extent. Suppose then that this is the content of my thought. By watching the movie, I form the thought that there *probably* is a killer in the woods and, as a consequence, cancel my camping plan. This seems more plausible. But then my thought does not seem to be a genuine belief, at least if we share Joyce's conception of beliefs. Indeed, although I am disposed to assent to the proposition *There probably is a killer in the woods* in the present context, there is a more critical context – a context at which I am able to question the assumptions that I now make as a response to the movie – in which I would dissent from it. Hence, my attitude is best described as a make-belief.

That Olson does not discuss Joyce's workout example or that of the drunkard is interesting because it seems to indicate that he takes the 'fiction' bit in 'fictionalism' too seriously. Admittedly, these two examples have little to do with pieces of fiction or even make-belief as it is ordinarily construed. They nonetheless satisfy Joyce's characterization of make-belief. When the time comes to work out, I am disposed to assent to the proposition that *I ought to do fifty sit-ups every day*. Still I am disposed to dissent from it in a more critical context – where I question (and reject) the assumption that there are a precise number of sit-ups that I ought to do on a regular basis. In such a context, I only accept that I ought to do more-or-less fifty sit-ups more-or-less every day. Thus, I believe that I ought to do more-or-less fifty sit-ups more-or-less every day, and I make-believe that I ought to do precisely fifty sit-ups every day. Now, since I am motivated to do *precisely* (rather than more-or-less) fifty sit-ups, the attitude that motivates me must be my make-belief (rather than my belief). Similarly for the drunkard case: although he assents to *If I drink this will happen to me*, he is disposed to dissent from that proposition in a more critical context. In such a context, he will only assent to a proposition like *If I drink, there is a 10 per cent chance of this happening to me* (Joyce 2001, p. 216). Thus he believes that if he drinks, there is a 10 per cent chance that he will die brutally, and he make-believes that if he drinks he will die brutally. Having the latter thought is a much better way to quit drinking. Thus, *pace* Olson, according to Joyce's theory of (make-)belief, make-beliefs *do* generate motivational states.

2. Olson's defence of conservatism

In light of Joyce's theory of (make-)belief, Olson's objections against his brand of fictionalism fail. This does not mean that conservatism should be rejected, of course, as it may be the case that it gives us a better story as to what we should do with our moral thought and talk. The question we would like to ask now is whether it does. Olson certainly cannot cite the fact that conservatism does not face the problems faced by fictionalism seen above, for we have seen that the fictionalist has ready responses to them. Given that conservatism is such a novel view, we need at this stage reasons to think that it should count as a serious

alternative to views such as abolitionism and fictionalism. In other words, we need some motivation for taking seriously the thought that it would be a good idea to retain our moral beliefs if we became error theorists. In his paper, Olson can be interpreted as aiming to provide the needed motivation.

Olson partly builds his case for conservatism by criticizing Joyce's case against it. To begin with, Joyce makes the general point that true beliefs have instrumental value: "In the vast majority of cases having a true belief to act upon is more likely to bring satisfaction of desire than having a false belief on the matter" (2001, p. 179). Suppose, to use an example he uses in another context (2006, p. 135), that two lions chase me one of which retracts. I form the belief that I am not chased anymore, as a consequence of which the remaining lion catches me and kills me. The belief was false and harmful. Moreover, it was harmful because it was false. Suppose that instead I had formed the belief that one lion was still chasing me, and that as a result, I had kept running and escaped it. That belief would have been true and useful. And it would have been useful because true.

Now, eloquent as such an example may be, there are counterexamples to the claim that only true beliefs are useful. Olson (2011, p. 193-4) cites some research in empirical psychology:

Overly positive self-evaluations, exaggerated perceptions of controls and mastery, and unrealistic optimism are characteristic of normal human thought [and] these illusions appear to promote other criteria of mental health, including the ability to care about others, the ability to be happy or contented, and the ability to engage in productive and creative work. (Taylor & Brown 1988, p. 193)

In addition, he mentions beliefs in personal identity, free will, responsibility and desert, whose usefulness seems to be quite independent of their truth. It seems that some false beliefs have instrumental value after all.

Another problem with Joyce's example is that a *make-belief* that no lion remains would arguably be as damaging as my corresponding belief is in his example. Indeed, whatever I am disposed to assent to in a more critical context, what makes me run or not in the eventuality of my being chased by a lion is what I assent to *in that context*. Whether I believe or make-believe that there is no lion, I will not run and the remaining lion will devour me. What Joyce needs is to show that it is better to have a false make-belief than the corresponding false belief. And we agree with Olson that he fails to do this.

As Olson notes, Joyce is likely to respond that he does not claim that true belief is always instrumentally valuable, only that we need a general cognitive *policy* recommending the acquisition of true beliefs and the avoidance of false ones (Joyce 2001, p. 179). According to Joyce, failure to adopt such a policy can have two unwanted consequences. First, it will lead to a tension in the believer's psychology: "A seemingly useful false belief [...] will require all manner of compensating false beliefs to make it fit with what else one knows" (2001, p. 179). In the present case, we think it is safe to understand 'what else one knows' to include the error theory. After all, we are interested in what we should do with our moral thoughts *qua* error theorists. Thus, the idea is that if you believe that all moral claims are false, you will naturally believe (e.g.) that it is not the case that lying is wrong. And the latter belief is inconsistent with the belief that lying is

wrong, thereby implying a certain tension in the believer's psychology. Let's call this the 'no-compensation rationale'. Second, Joyce argues that to accept to have mutually inconsistent beliefs on moral matters is to put a foot on a slippery slope. Quoting Charles Peirce, he says that, when it comes to beliefs, anything but a policy of seeking the truth leads to "a rapid deterioration of intellectual vigor" (2001, p. 179). Once the error theorist accepts moral beliefs, the thought seems to be, it is hard to resist having inconsistent beliefs in other areas. Let's call this the 'no-deterioration rationale'.

Let's start with the latter rationale. We agree with Olson that Joyce has not done enough to motivate the claim that failing to adopt a truth-seeking policy in one area will lead to the failure to adopt it in other areas. According to Olson, Joyce's rationale

rests on an underestimation of our ability to discriminate in our cognitive policies. It is unclear why we should expect a pragmatic cognitive policy of sticking to what one at some level recognizes as false beliefs about some particular matter to infect one's general commitment to a truth-seeking policy in any intellectual endeavour. (2011, p. 194)

It may indeed be possible for us to adopt, as it were, localized pragmatic policies that would allow us to assent to mutually contradictory propositions in critical contexts in one area of thought without doing so in other areas of thought. It certainly seems possible for a subject to adopt such a pragmatic policy regarding free will, says Olson, while adopting a strictly truth-seeking policy in mathematics. More concretely, such a subject would allow herself to assent to the proposition *My will is free* in some critical contexts and dissent from it in others. But nothing indicates that this would prevent her from consistently assenting to the proposition *Two and two make four* in all contexts. This is an empirical claim that may or may not be true, and Joyce's hypothesis is merely one among others. As a result, we cannot take it to pose a problem to conservationism.

Regarding the no-compensation rationale, Olson finds it odd that Joyce puts forward such a claim, as he arguably accepts the thesis that the moral is autonomous from the non-moral or empirical (Joyce 2001, pp. 153-158), "at least in the sense that no moral conclusions follow from purely non-moral or empirical premises" (Olson 2011, p. 194). If we accept the autonomy of the moral, therefore, it should be legitimate to at once believe the error theory and believe first-order moral claims, as we would be unable to derive from the error theory – a non-moral thesis – first-order moral claims that are inconsistent with the ones we initially accept. As a result, it seems that believing at once the error theory and first-order moral claims would *not* demand the sort of compensation Joyce talks about.

We think one plausible line of response is open to Joyce, a response which does not seem to involve a wholesale rejection of the autonomy thesis. The response is this. There are various versions of the autonomy thesis, some of which are stronger than others. According to a modest but appealing version, no *positive* moral claim – namely a moral claim predicating a moral property of things – follows purely from non-moral propositions, though it may be that moral claims introduced with a negation operator do. This weak autonomy thesis, we suggest, is especially attractive for fictionalists like Joyce.

Consider the following line of reasoning: (1) the error theory is true; hence, (2)

no (positive) moral proposition is true; hence, (3) the proposition *Lying is wrong* is not true; hence, (4) it is not true that lying is wrong; hence, (5) it is not the case that lying is wrong. Whereas it is immediately implausible that the proposition *Lying is wrong* directly follows from the proposition *Lying involves treating someone as a mere means*, there does not seem to be anything so absurd to the view that (5) directly follows from (1). The strong interpretation of the autonomy thesis excludes both types of inferences, whereas the modest one excludes only the latter. Now, notice that the weak thesis is clearly compatible with there being a tension in the believer's psychology. Indeed, (5) contradicts the claim that lying is wrong. Thus, if the inference from (1) to (5) is allowed (and sufficiently easy to draw), there will plausibly be a tension in the one who *believes* in the error theory and *believes* that lying is wrong. Hence, it seems that Joyce could accept the weak version of the autonomy thesis instead of the strong one and maintain that, for an error theorist, being disposed to assent to positive moral claims when she does metaethics would require quite a lot of compensating false beliefs. As a result, there is a tension in the believer's psychology, at least in light of Joyce's theory of (make-)belief.

Of course, one may disagree at various points of the response. One may, for instance, defend the strong autonomy thesis, or maintain that, even if the error theory entails the negation of all positive moral claims, this would not have the unwanted consequences Joyce talks about. Conservatism is therefore not off the table yet. We think however that the debate between the fictionalist and the conservationist is driven by a more fundamental disagreement to which no clear resolution may be forthcoming and the result of which may fail to pick out a form of conservatism that is both plausible and not merely a terminological variant of the sort of fictionalism Joyce defends. This claim will be defended in the final section of this paper.

3. A challenge for the conservationist

It is one thing to show that conservatism should not be discarded out of hand; it is quite another to show that there is a version of it that is plausible and constitutes a genuine alternative to fictionalism. The view enjoins the error theorist to keep her moral beliefs. But what does this mean? What does the conservationist's prescriptive claim amount to?

One version of the conservationist view could be that we should continue to assent to moral propositions in our most critical contexts, that is, when doing metaethics. We could, for instance, accept the error theory in some critical contexts and accept first-order moral propositions (hence reject the error theory) in some other critical contexts. Assuming this is a possible state of affairs, we could then say that this scenario is in some way preferable to the scenario where, as error theorists, we assent to moral propositions in non-critical contexts only. We do not have a knockdown argument for this claim, but it seems to us that such a view, though apparently coherent, is too implausible to be taken seriously by those attracted to conservatism. At any rate, the burden of proof is on the conservationist to show us that this is indeed a viable alternative. Note however that it is a view that we can clearly distinguish from the view Joyce defends: whereas Joyce's fictionalism recommends that we stop assenting to moral propositions in critical contexts (i.e., when doing metaethics), the conservatism

under discussion recommends that we continue to assent to moral propositions in the same contexts.

So, what is Olson's version of conservatism? It is presumably not the one just discussed. Rather, Olson seems to put forward the claim that it is in fact possible to have beliefs fixed by *non-critical contexts*, and as a result that Joyce is wrong to claim that one's beliefs are only fixed by critical contexts. It is possible, says Olson, to believe – in the seminar room – that no act is really right or wrong while believing – out of the seminar room – that some acts are right and some acts are wrong. Against Joyce's conception, Olson cites evidence to the effect that beliefs are not as Joyce construes them. First, it does seem possible to simultaneously have an occurrent belief that *p* and a dispositional belief that non-*p* or to "temporarily believe things we, in more reflective and detached moments, are disposed to disbelieve" (2011, p. 200). Listening to a charismatic politician, we might be taken in by what they say to such an extent that we might take ourselves to believe them, if only temporarily. Furthermore, there are theories, which, although we might doubt their truth for substantial reasons, we do not take to be conceptually confused and which seem to presuppose a conception of belief along the lines we just sketched. Consider Pascal's Wager, the argument to the effect that we ought to believe that God exists and act on that belief in everyday life even if we ought to deny His existence in the philosophy room. This seems to be a perfectly coherent position. Similarly for two-level versions of utilitarianism, according to which, in everyday contexts, we ought to believe that killing is always wrong but disbelieve this proposition when we do normative ethics. Were the belief that *p* a disposition to assent to *p* in critical contexts, these theories would be far less plausible (and perhaps conceptually confused). And again, we tend to say things such as "I knew she was lying, but hearing her speech and the audience's reaction last week, I really believed what she said" (2011, p. 200). Nothing indicates that we are conceptually confused there, but Joyce's theory seems to entail that we are. And if we are not, one might legitimately wonder, why would it be confused to say "I knew there are no moral truths, but seeing him beat his son last week, I really believed it was wrong"?

We do not think that Olson has succeeded just yet in motivating conservatism. What he needs to show is not simply that there is an alternative conception of belief the acceptance of which allows us to defend a plausible conservatism position (as opposed to the implausible one discussed at the beginning of this section). What he needs to show is that the conservatism position in question constitutes a genuine alternative to Joyce's fictionalism. After all, it could be the case that the sort of moral beliefs Olson recommends that we keep are in fact the moral make-beliefs that Joyce recommends. It could be the case, in other words, that Joyce's fictionalism and Olson's conservatism are mere terminological variants of the same prescriptive view.

The reason why we should take this possibility seriously is that fictionalism and conservatism (of the kind under discussion) both seem to recommend that (i) we dissent from moral propositions in our most critical contexts and that (ii) we assent to some first-order moral propositions in our non-critical contexts. The only difference between the two is that the conservatism, by contrast with the fictionalist, takes some of the attitudes we hold when we assent to propositions in non-critical contexts to be genuine beliefs. The conservatism therefore faces a challenge: that of filling in the details of her theory of belief in such a way that it

clearly distinguishes belief from the make-belief that Joyce recommends that we have towards the moral fiction. Let us call this problem the ‘conservationist’s challenge’. We will end this paper by outlining a strategy that the conservationist may employ in order to meet this challenge.

Recall Joyce’s characterization of make-belief:

S make-believes that p only if (i) S is disposed to assent to p in some contexts, (ii) S has assented to not- p in her most critical contexts, and (iii) S is disposed to assent to not- p in her most critical contexts. (2001, p. 193)

One striking feature of this characterization is that it does not draw any distinction between different kinds of make-belief. If one is disposed to assent to a proposition in a non-critical context, but not in a critical context, one thereby holds the attitude of make-belief towards that proposition. So Joyce would count as make-belief virtually any cognitive attitude that fits this description. Surely, this is not because he thinks that there is no difference between (say) the make-beliefs that we hold at the movies and the make-beliefs that he recommends we hold towards the moral fiction. Rather, he does not draw any distinction between make-beliefs because he thinks that such a distinction would be *irrelevant* in the context of providing an answer to the question of what we should do with our moral beliefs once we have accepted the error theory. As long as we adopt make-belief in one (sufficiently motivationally strong) form or another towards the moral fiction, we are doing exactly what Joyce’s fictionalism recommends that we do.

Now, a three-step strategy for meeting the conservationist’s challenge suggests itself. First, the conservationist could argue that, in refusing to discriminate between different kinds of make-beliefs, Joyce in fact fails to take notice of a crucial dimension of difference between two kinds of attitudes – call them ‘belief*’ and ‘make-belief*’ – that are relevant to the present context. The conservationist could then argue that, in light of this distinction, the type of attitudes one should have towards first-order moral propositions is one of those types – say belief*. The conservationist could finally argue that the fictionalist should restrict his proposal to the other type – *viz.* make-beliefs* – for it to have any plausibility or to be any interesting.

Here is a way the conservationist should *not* want to implement this strategy. First, she would substantiate the distinction between beliefs* and make-beliefs* by saying that belief*, as opposed to make-beliefs*, are motivationally strong. Second, she would insist that motivational strength is an important feature of any attitude towards moral propositions we should recommend. (This is especially plausible if, as Joyce thinks, one of the main functions of moral beliefs is that of helping us resist weakness of the will.) But, in order to complete the strategy, the conservationist would then have to argue that fictionalists should be committed to recommending make-beliefs* – construed as attitudes that lack motivational force – rather than beliefs*. And this is why the manoeuvre would fail, for, as we have seen in Sec. 1, motivational force is a central feature of the attitudes the fictionalist recommends, and one the fictionalist is entitled to posit. The conservationist therefore cannot appeal to motivational strength in order to substantiate the distinction between belief* and make-belief*.

Based on Olson’s remarks on the nature of belief, we believe the conservationist

could carry out the strategy in the following way.¹² First, the difference between beliefs* and make-beliefs* lies in their relative ‘seriousness’. While beliefs* are fixed by an assent of a serious sort, make-beliefs* are fixed by an assent of a non-serious sort – or some story in the vicinity. Second, the attitudes towards moral propositions we should recommend would be serious in this sense. This makes sense since morality is, and should arguably remain, a serious matter. Finally, the fictionalist is committed to recommending the having of make-beliefs*, namely attitudes fixed by an assent of a non-serious sort.

We are sceptical that the last claim can be properly defended. In order for the strategy to be carried out, the conservationist at this stage must propose an elucidation of the notion of seriousness such that the fictionalist would have to recommend that we have non-serious attitudes – rather than serious ones – towards first-order moral propositions. We doubt that this can be done. One mark of a non-serious type of assent which Olson’s text suggests¹³ is the accompanying higher-order disposition to occasionally remind oneself that one’s state of mind is directed towards false propositions. A serious type of assent, by contrast, is not accompanied by such a disposition. On this understanding of seriousness, beliefs* in false propositions are more serious than make-beliefs* in the same propositions in virtue of the fact that they are not accompanied by a disposition to occasionally remind oneself that the propositions in question are false, whereas make-beliefs* are. Although this may be an attractive way of spelling out the notion of seriousness in other contexts, it still does not help here. As we have seen when we discussed Olson’s objections to fictionalism, the fictionalist is free to argue that the sort of make-beliefs she recommends that we should have towards first-order moral propositions need not be, and are typically not – given the sort of error theorist she has in mind – accompanied by the disposition to occasionally remind oneself that the moral propositions in question are indeed false. She would therefore be happy to say that such attitudes are serious in Olson’s sense (and hence beliefs*) while denying that they are beliefs. Of course, Olson could stipulate that such attitudes are beliefs, but then he would only be proposing a terminological variant of the fictionalist’s prescriptive story.

Furthermore, the claim that the disposition to occasionally remind oneself that one’s attitude is fictive is the distinguishing feature of make-belief* is not an attractive option for the conservationist who – reasonably¹⁴ – wants to make sure that paradigmatic forms of make-belief come out on the make-beliefs* side of the distinction. Consider make-belief towards fictional characters and situations – a paradigmatic form of make-belief if any is. It does not seem to be an essential feature of it that it is accompanied by a disposition to occasionally remind oneself that the relevant characters and situations are not real. In fact, in much of our engagement with fiction – in particular pleasurable fiction – we are taken in to such an extent that we are *not* disposed to have the relevant thoughts. If the conservationist wishes to count paradigmatic forms of make-belief as make-

¹² Thanks to an anonymous referee for suggesting the following way of construing Olson’s account of belief.

¹³ See in particular, his remarks on ‘compartmentalization’ (2011, p. 199).

¹⁴ Because she wishes to accommodate the ordinary notion of make-belief, or simply because she takes it that the fictionalist would adopt this constraint in drawing a distinction between belief* and make-belief*.

beliefs*, she therefore seems to have a counterexample to the claim that make-beliefs* cannot be serious attitudes.¹⁵

The conservationist is free to count such an engagement as belief*, to be sure, but given that the kind of attitude in question is one falling under the fictionalist's recommendation, she would then be unable to put forward a conservationist position that constitutes a genuine prescriptive alternative to fictionalism. The conservationist thus still faces the challenge, and it is not clear how she could meet it. Until this is done, conservationism should not be seen as a serious alternative to the extant positions in what we may call 'prescriptive metaethics'.

Conclusion

What should we do with our moral thoughts and discourse if we became error theorists about them? Until recently, the most prominent answer was fictionalism, the view that we should, in some sense, do as if our moral thoughts and claims are or could be true, as if we were not error theorists, at least in ordinary circumstances. A new solution has recently been put forward, one that deserves careful assessment; it is the view – called 'conservationism' – that we should carry on believing and asserting moral propositions. This solution, its proponents argue, promises to have the advantages of fictionalism without its flaws. In this paper, we have argued that the particular way conservationism has been defended by Jonas Olson fails to make it superior to at least one brand of fictionalism, the one defended by Richard Joyce. Joyce's fictionalism presupposes a partial account of beliefs. First, we have shown that Olson's objections to Joyce's fictionalism fail if we assume the truth of that account. Second, we have seen that Olson has not yet provided a conclusive defence of conservationism against Joyce's objections. Finally, we have presented a challenge to the conservationist: that of providing a conservationist position that is both plausible and not merely a terminological variant of fictionalism. And we have expressed doubts about the possibility of meeting such a challenge.¹⁶

References

- Brown P (2011) The error in moral discourse and what to do about it. PhD dissertation, University of Manchester
- Cova F & Naar H (manuscript) 'What's the real question?' Recasting the paradox of fiction as a psychological puzzle
- Garner R T (2006) On the genuine queerness of moral facts. *Australasian Journal of Philosophy* 68(2):137-146

¹⁵ Other paradigmatic examples include children's games of make-believe, videogames, and role-playing games. For a discussion of make-believe in its paradigmatic forms, see Walton (1990). For a discussion of the emotional and motivational aspects of our engagement in such activities, see [Cova & Naar, manuscript].

¹⁶ We would like to thank Richard Joyce, Richard Dub, Florian Cova, audiences in Geneva and Ovronnaz, Switzerland, and two anonymous reviewers for this journal for helpful comments on an earlier draft of this paper. We also would like to thank the Swiss National Science Foundation (SNSF) and the Center for Research on Ethics (CRE) for their generous support at various stages of the writing.

- Garner R T (2007) Abolishing morality. *Ethical Theory and Moral Practice* 10(5):499-513
- Greene J D (2002) The terrible, horrible, no good, very bad truth about morality, and what to do about it. PhD dissertation, Princeton University
- Hinckfuss I (1987) The moral society: Its structure and effects. Australian National University, Canberra
- Joyce R (2001) The myth of morality. Cambridge University Press, Canberra
- Joyce R (2006) Metaethics and the empirical sciences. *Philosophical explanations* 9:133-148
- Lutz M (2014) The 'now what' problem for error theory. *Philosophical Studies* 171: 351-371
- Mackie J L (1976) Problems from Locke. Oxford Clarendon Press, Oxford
- Mackie J L (1977) Ethics: Inventing right and wrong. Penguin Books, New York
- Olson J (2011) Getting real about moral fictionalism. In: R Shafer-Landau (ed), *Oxford Studies in Metaethics* (Vol. 6:182-204). Oxford University Press, New York
- Smith M (1994) The moral problem. Blackwell
- Suikkanen J (2013) Moral error theory and the belief problem. In: R Shafer-Landau (ed), *Oxford Studies in Metaethics* (Vol. 8:168-194). Oxford University Press, New York
- Taylor S E & J D Brown (1988) Illusion and well-being: A social psychological perspective on mental health. *Psychological Bulletin* 103:193-210
- Walton K. L. (1990) Mimesis as make-believe: On the foundations of the representational arts. Harvard University Press.