

# Moral Error Theory and the Belief Problem

*Jussi Suikkanen*

Final Author Copy: to be published in *Oxford Studies in Metaethics* 8, 2013.

## 1. INTRODUCTION

According to J.L. Mackie's moral error theory, 'a belief in objective values is built into ordinary moral thought and language' and 'this ingrained belief is false (Mackie 1977: 48–9).' Mackie suggested that, as a consequence, 'all our first order [moral] judgments are false (ibid.).' Mackie's view has become fashionable again. Its new advocates have created more sophisticated formulations of the position, introduced new arguments for it, and solved many of the previous problems of the view.<sup>1</sup>

I will begin by formulating the moral error theory more precisely. I will then claim that moral error theorists face the question of what will and should happen to one's moral thought when one accepts the essential elements of the moral error theory. Posing this question helps us to outline five possible versions of the view that each respond to the previous question in different ways. I set two of the resulting versions of the moral error theory aside, because, currently, we do not have sufficient evidence to evaluate one of their central empirical predictions.

I will then argue in the last two sections that the remaining versions of the moral error theory are inconsistent with the most appealing theories of beliefs.<sup>2</sup> Because of this, the moral error theorists must either (i) defend some new, non-standard accounts of beliefs, or (ii) accept one of the two versions of error theory which come with controversial empirical commitments. That the moral error theorists are forced to make this choice is admittedly not a knock-down argument against their view, but I hope that it will help us to assess the costs of the view in a new light.

## 2. MORAL ERROR THEORY

Moral error theory consists of four elements. The first is:

---

<sup>1</sup> See, for instance, Burgess (2007), Daly & Liggins (2010), Garner (2007), Joyce (2001), Olson (2010), Pigden (2007), and Robertson (2008).

<sup>2</sup> Whilst I was working on this chapter, Matt Bedke (2011) was independently working on a manuscript in which he sketches a resembling argument against all forms of non-naturalistic realism.

**Cognitivism** Utterances of basic indicative moral sentences conventionally express basic moral *beliefs*.

By basic indicative moral sentences, I mean simple, positive indicative moral sentences of the form ‘x is F’ where ‘F’ is a moral predicate. Thus, according to the moral error theorists, if I claim that ‘torture is wrong’, I express my belief that torture is wrong. Cognitivism is therefore a claim about how our moral talk is related to our thought. The role of such talk is to communicate to others how we *believe* the world to be morally speaking. This means that moral error theories are in conflict with the non-cognitivist views in metaethics.

According to the non-cognitivists, moral utterances do not conventionally express beliefs but rather some non-cognitive attitudes such as plans, desires, approvals, acceptances of norms, make-belief, and so on.<sup>3</sup> Roughly, the difference is that beliefs aim at being true whereas non-cognitive attitudes are not considered to do so. Many non-cognitivists believe that attitudes aiming at truth have to be motivationally inert. They then argue that moral utterances must express non-cognitive attitudes because the attitudes expressed by those utterances do play an intrinsically motivating role.

The second element of moral error theories describes the content of the expressed moral beliefs. It is:

**Semantic** Basic moral beliefs ascribe moral properties to objects.<sup>4</sup>

What makes the properties ascribed by the basic moral beliefs *moral properties*? The answer to this question will capture what is ‘conceptually non-negotiable’ in our moral thought (Joyce 2001: ch. 1). That is, one could not have basic moral beliefs unless one’s beliefs ascribed moral properties to objects (but see fn. 4). If our beliefs began to ascribe some other kinds of properties, they would no longer count as moral beliefs. Similarly, if my bachelorhood-belief did not ascribe manhood and unmarriedness to a person, it could not count as a belief about a bachelor. This means that, in order to specify what moral beliefs are, the moral error theorists need to give some criteria for what makes a given property a moral property.

Mackie argued that moral properties must be ‘objective values’ with inbuilt ‘to-be-doneness’ (Mackie 1977: ch. 1). The most charitable way of understanding this is to

---

<sup>3</sup> For representative accounts of non-cognitivism, see Blackburn (1993 and 1998), and Gibbard (1990 and 2003).

<sup>4</sup> By using the term ‘basic’ here again, I limit the scope of the thesis only to beliefs which can be expressed by the simple, positive indicative moral sentences described above. It is less clear whether more complex moral thoughts that can be expressed by negated and disjunctive sentences also ascribe moral properties.

think that moral properties have a constitutive relation to categorical reasons (Joyce 2001: ch. 2, Olson 2010: sec. 2).<sup>5</sup> It is a part their essence that some facts about an object that has such a property provide categorical reasons to agents.

Thus, for example, what is required for torture to have the moral property of being wrong is that the fact that torture is wrong (or the fact that it hurts) provides categorical reasons for agents not to torture others. The talk about these reasons implies that the wrongness of torturing people (or the fact that it hurts) requires not torturing others. That a requiring reason is a *categorical* reason means that its existence does not depend on what agents want (Shafer-Landau 2005: 108).

The third element is:

**Metaphysical** Moral properties are not instantiated.<sup>6</sup>

Given the previous definition of moral properties, the moral error theorists can argue for Metaphysical by showing that there are no categorical reasons. If there were no such reasons, then no object would have properties such that some facts about the bearers of those properties would always provide categorical reasons.

Moral error theorists often argue against the existence of categorical reasons by claiming that they would be metaphysically unacceptable. Allegedly, it would be incompatible with our scientific world-view that facts about what an object is like could require actions from us independently of what we want. As a consequence, it is claimed that, if our best form of investigation to what there is fails to detect any categorical reasons, then we should not believe that they exist. Thus, moral properties fail to be instantiated.

Cognitivism, Semantic, and Metaphysical together entail:

**Falsity** All basic moral beliefs are false.

Falsity means, for instance, that, if you believed that torturing people is wrong, then your belief would be false.

Many error theorists accept Falsity (Mackie 1977: 35, Pigden 2007: 442). This is problematic (Streumer 2012: sec. 3). Only few moral error theorists have reported that they have given up their first-order moral beliefs when they began to endorse the

---

<sup>5</sup> Some realists resist moral error theory by denying this very thesis (Railton 1986).

<sup>6</sup> On many views about properties, properties that are not instantiated do not exist (see, for instance, Armstrong 2010: 15). This would mean that, according to the moral error theorists, moral properties would not even exist. Furthermore, moral error theorists also often make a related epistemological claim according to which, even if such properties existed, we would be unable to have knowledge of them (Mackie 1977: 38).

moral error theory (see sec. 3 below). If they in this situation accepted Falsity, this would easily lead them to a contradiction.

By the standards of epistemic rationality, the moral error theorists who accept Falsity and yet do not give up their moral beliefs should believe that their own moral beliefs are false. Take the moral error theorist's moral belief that *p*. Because of her commitment to Falsity, epistemic rationality would require the moral error theorist to believe that her own belief that *p* is false. But, given the law of the excluded middle, believing that *p* is false also creates a requirement of epistemic rationality to believe that not-*p*.<sup>7</sup> This means that, in this case, a moral error theorist who is disposed to conform to the requirements of epistemic rationality would come to believe that both *p* and not-*p*. To avoid this irrational outcome, the rationally disposed moral error theorists will either have to (i) give up their own moral beliefs and only think that the moral beliefs of other people are false, or (ii) give up Falsity and at least one of the theses that entail it.

This result already begins to show us some of the costs of the moral error theory. However, I want to focus on a slightly different objection. Instead of Falsity, this objection makes use of another consequence of Cognitivism, Semantic, and Metaphysical, namely the weaker thesis that:

**Evidence**      The truth of Cognitivism, Semantic, and Metaphysical means that there seems to be sufficient evidence for the falsehood of our basic first-order moral beliefs.<sup>8</sup>

Note how weak Evidence is. It only says that, if you think that Cognitivism, Semantic, and Metaphysical are true, then it will seem to you that there is sufficient evidence against the truth of your first-order moral beliefs. This much is very difficult to deny. So, I take Cognitivism, Semantic, Metaphysical, and Evidence to be the constitutive elements of the moral error theoretic position. I also want to emphasise that, in themselves, these four claims are at the very least consistent.

### 3. AND THEN WHAT?

---

<sup>7</sup> This does not follow on the epistemic conceptions of truth. That you believe that *p* is not warrantably assertible does not entail that you are required to believe that not-*p* is warrantably assertible (Wright 1992: 19–21).

<sup>8</sup> I assume that the moral error theorist has a stable belief that both Cognitivism and Semantic are true. She is at no stage tempted to give up these beliefs. So, when she considers the arguments for Metaphysical and accepts their conclusion, she will at the same time believe that Cognitivism, Semantic, and Metaphysical are true. In this situation, the moral error theorist can also come to believe Evidence without giving up any of these beliefs. So, the moral error theorist can believe her own view. It might be that one could not believe a global error theory (Streumer 2012).

We can then pose the following question for the moral error theorists: In the case that one accepts the previous four elements of the moral error theory, what will and should happen to one's first-order moral thought?

There are only five possible answers this question. I will argue below that three of the resulting positions are incompatible with the most plausible accounts of beliefs. Furthermore, the remaining two views are committed to the prediction that one will give up one's moral beliefs if one accepts the central elements of the moral error theory. In this situation, anyone who does not answer the previous question will still have to recognise that she will have to either accept the empirical prediction or give an unorthodox account of beliefs.

I admit that not many ordinary people are aware of the moral error theory and the arguments for it. Even fewer people have accepted those arguments. We can still consider what would happen if people were convinced by them.

I will assume that there are no significant psychological differences between how moral error theorists themselves have reacted to accepting their arguments and how others would react if they accepted those arguments. This is because I assume that there are no significant differences between the moral attitudes of ordinary people and those of the moral error theorists – at least before the latter begin theorising.

### 3.1 Descriptive and Normative Abolitionism

Let us begin from ordinary beliefs. When I find out that there seems to be sufficient evidence for the falsehood of an ordinary belief, I seem to (almost?) always directly give up the belief in question. Imagine that I believed that it is currently raining outside on the basis of an earlier forecast. If I then suddenly saw the sun shining, I would recognise that there is sufficient evidence against the truth of my belief. As a result, I would automatically give up that belief.

This observation leads to a philosophical view called *doxastic involuntarism*.<sup>9</sup> It comes in two strengths. The weak version holds that, in fact, no one ever voluntarily continues to believe what they believe they have sufficient evidence to believe to be false. The stronger version holds that it is *impossible* to continue to believe what one believes one has sufficient evidence not to believe.

---

<sup>9</sup> See Alston (1989), Hieronymi (2006), and Williams (1973).

I will call *descriptive abolitionism* the moral error theoretic view which claims that, in a similar fashion, if one becomes convinced about the truth of Evidence, one will give up one's moral beliefs. This claim is a universal descriptive thesis about how human beings react in that situation. To my knowledge, no philosopher has ever defended this view. An even stronger version of this view would hold that it would furthermore be *impossible* to keep one's moral beliefs if one believed Evidence.<sup>10</sup>

In contrast to those views, Richard Garner explicitly states that, if one ceases to believe in moral properties, facts, and truths, this is still not *enough* to get one to give up one's first-order moral beliefs (Garner 2007: 502). He also claims that how difficult it is to give up one's moral beliefs depends on how dedicated one is as a moralist (Garner 2007: 504). I understand these claims to mean that, according to Garner, it is possible for one to keep one's moral beliefs even if one comes to accept Evidence.

This interpretation explains why Garner then goes on to give *prudential normative reasons* for why one *should* give up one's moral beliefs *after* one has accepted the moral error theory (Garner 2007: 506–512). He, for instance, argues that it would be good for one to give up one's moral beliefs because they only tend to inflame disputes and preserve unfair arrangements (Garner 2007: 502, see also Hinckfuss 1987: ch. 3–4). Making this point would hardly make sense if one thought that everyone would automatically give up their moral beliefs when they accepted the moral error theory.

I call Garner's position *normative abolitionism*. It states that, even if one would not automatically give up one's moral beliefs if one thought that there is sufficient evidence against their truth, one would still have sufficient prudential reasons to do so.

### **3.2 Descriptive-Normative and Normative Revolutionary Fictionalism**

Other moral error theorists claim that, when one becomes convinced about the truth of Evidence, one should adopt a slightly modified form of moral thought (Nolan, Restall, and West 2005: 307–314, Joyce 2006: 208). There are two versions of this view.

The first claims that, when one accepts Evidence, it will be impossible not to give up one's moral beliefs. No one would be able to keep their moral beliefs in this situation because doxastic involuntarism is true. The view then claims that giving up one's moral thoughts altogether would be too costly. One would, for instance, lose all

---

<sup>10</sup> I also assume here something like the 'ought implies can' principle here. If we could not keep our moral beliefs, then it is not the case that we should keep them.

the benefits from beneficial co-operation the existence of which requires that everyone complies with certain moral constraints. As a result, one should (prudentially) continue to think *as if* objects would have moral properties.

I call this view *descriptive-normative revolutionary fictionalism*. It states that either no one will or can continue to hold their moral beliefs once they accepted Evidence. It then proposes that, in this situation, everyone should (prudentially) begin to *make-believe* that objects have moral properties.

I will call the second version of revolutionary fictionalism *normative revolutionary fictionalism*. This view holds that, when one becomes convinced about the truth of Evidence, one can give up one's first-order moral beliefs, continue to hold them, or adopt the resembling make-belief attitudes to replace those beliefs. It then offers everyone prudential reasons for why one should choose that last option.<sup>11</sup>

### 3.3 Conservatism

Finally, conservatism believe that, when one accepts Evidence, one can either continue to hold one's first-order moral beliefs, give them up, or adopt the corresponding make-belief attitudes. They then add that keeping the ordinary beliefs would bring about more prudential benefits than giving them up. This is why, on this view, no one should give up their moral beliefs even if they accepted Evidence. This view has been defended, for example, by J.L. Mackie, Charles Pigden, and Jonas Olson.

As quoted in the beginning, Mackie thought both that a belief in objective values is built into our moral thought and that this belief is false (Mackie 1977: 48). According to him, our ordinary moral thought is responsible for our 'first-order moral views', whereas his own moral error theory is a 'second-order view' (Mackie 1977: 16). This enabled him to explicitly note that the first-order and second-order views are completely independent of one another: 'one could be a second-order sceptic without being a first-order one (ibid.).'

---

<sup>11</sup> See Joyce (2001: ch. 8), and Nolan, Restall, & West (2005). Joyce suggests that one should adopt the fictive attitudes instead of the full-blown false beliefs, because 'true beliefs are an extremely valuable commodity (Joyce 2001: 178).' He is sceptical about whether one could retain one's full-blown beliefs once one begins to think that there is sufficient evidence against their truth (ibid.). However, because he thinks that giving up one's moral beliefs altogether would be an option for us, I would classify him as normative revolutionary fictionalist.

Nolan, Restall & West (2005: 310–311), in contrast, think that one probably could not give up one's moral thoughts altogether. Because of this their position does not neatly fit my categories. However, because they recommend the switch from beliefs to make-belief on prudential grounds, their view too is close to my description of normative revolutionary fictionalism.

I understand this claim to mean that, according to Mackie, one can keep one's ordinary moral beliefs even when one comes to accept the higher-order theoretical claims about their falsehood (Mackie 1977: 22, 49). This interpretation makes sense of Mackie's later 'cost-benefit' analysis of moral beliefs. After his moral error theory, he attempted to argue that, for a number of non-moral reasons, it is beneficial to keep one's first-order moral beliefs despite their falsehood (see Mackie 1977: part 2).

Likewise, Charles Pigden argues, on the basis of his and Nietzsche's cases, that it is possible to continue to hold old and adopt new moral beliefs even when one thinks that they are all false (Pigden 2007: 445, Nietzsche 1990: 35–36). According to Pigden's interpretation, Nietzsche then argued that we should adopt new, more life-affirming moral beliefs which have more a 'species preserving' content, even if these beliefs are just as false as the previous moral beliefs (Pigden 2007: 446).

Finally, Jonas Olson explicitly defends conservationism – the view which according to him recommends that 'we stick with ordinary moral discourse with its *false beliefs* and assertions' (Olson 2011: 193, my emphasis). On Olson's view, falsity is not a flaw of moral beliefs. Rather, such beliefs, even if false, should be embraced because of their prudentially beneficial consequences (Olson 2011: sec. 5).

Olson also argues that it is possible to 'compartmentalise' one's everyday moral beliefs and one's theoretical convictions about their falsehood in the philosophical contexts (Olson 2011: 199–200).<sup>12</sup> Because of this, according to Olson, one can and should keep one's moral beliefs even if one accepts Evidence in the philosophical debates. This is because the emotional effects of the morally loaded situations tend to prevent the philosophical views from affecting our ordinary moral beliefs.

I have thus introduced five possible versions of moral error theory.<sup>13</sup> Two of these views (descriptive abolitionism and descriptive-normative revolutionary fictionalism) agree that, if anyone comes to accept Evidence, they either will not or could not retain their moral beliefs. The three other views, in contrast, hold that one could keep one's

---

<sup>12</sup> Note how Olson uses a highly visual metaphor which he has not explained in more basic terms. Because of this, it is difficult to evaluate what the content of his thesis is.

<sup>13</sup> It may sound strange to call these views versions of moral error theory. I do so because they all share the essential commitments of moral error theory, and they each give one possible answer to a question we can legitimately ask from the moral error theorists. Note also that these views only take a stand on the behaviour of moral attitudes with regards to thoughts about evidence. They make no claims about non-moral domains.



moral beliefs even if one accepted Evidence. They then go on to disagree about what one should do in that situation on prudential grounds.

At this point, I will set descriptive abolitionism and descriptive-normative revolutionary fictionalism aside for the following reason. Both the weak and strong versions of these views are committed to the empirical thesis that, if one accepts Evidence, then one will not keep one's moral beliefs.<sup>14</sup> The problem is that we do not yet know whether this empirical thesis is true or false.

Some error theorists have reported that they have kept their moral beliefs (Pigden 2007: 445, Olson 2011: sec. 5). Others have, in contrast, reported that they have given up their moral beliefs (Hinckfuss 1987: ch. 5, Burgess 2007: 437, and Garner 2007). It is not clear how reliable any of these reports are. In addition, from just a couple of cases, we cannot draw general conclusions about what would happen to moral beliefs more widely in certain circumstances.<sup>15</sup> For this reason, we do not yet know whether the empirical commitment of descriptive abolitionism and descriptive-normative is true or false. Because of this, I set these views aside.<sup>16</sup>

I want to emphasise also that these two views are immune to the argument which I will give below against the other three forms of moral error theory. These views were normative abolitionism, normative revolutionary fictionalism, and conservatism. My argument will attempt to show that those views are incompatible with the most plausible theories of beliefs. So, my argument should apply at least to Garner, Mackie, Pigden, and Olson.

Many moral error theorists have not made clear what consequences they believe the acceptance of Evidence to have for our moral thought. One way to read this article therefore is to take it as a recommendation for these philosophers. If one is to be a moral error theorist and accept the most plausible accounts of beliefs, then one should

---

<sup>14</sup> The weak versions predict that one will give up one's moral beliefs, whereas the strong versions furthermore claim that one would necessarily do so. Both of these views make the prediction that no one will ever keep the moral beliefs in the relevant cases.

In contrast, the three other views are only committed to a subjunctive conditional (a modal thesis) of the following form: if we accepted evidence, then it would be possible for us to retain our moral beliefs. Note that this subjunctive conditional can be neither verified nor falsified on empirical grounds. It is compatible with both people keeping their moral beliefs and with people giving them when they accept Evidence.

<sup>15</sup> Of course, to falsify the previous views, all one needs is one case in which someone does not give up their moral beliefs when they come to believe Evidence.

<sup>16</sup> These views admittedly have many other philosophical elements (such as the non-existence of moral properties) the plausibility of which can be assessed a priori. Yet, even if these elements were defensible, the empirical commitment of these views would still make them empirically falsifiable.

be either a descriptive abolitionist or a descriptive-normative revolutionary fictionalist. There are no other consistent options.

#### 4. THE INCONSISTENCY

Two elements of the three error theoretic views specified above are jointly incompatible with the most plausible theories of beliefs. The first of these is Evidence – the thesis that there would seem to be sufficient evidence for the falsity of our moral beliefs. The second is that, even if someone accepted Evidence, it would be possible for them to keep their first-order moral beliefs.

Because of this alleged quality of moral beliefs, moral beliefs are according to these views *systematically insensitive attitudes*. By this technical phrase I mean attitudes that are not given up when one believes that there seems to be sufficient evidence against their truth.<sup>17</sup> However, I do not mean merely attitudes that admit of local failures of sensitive in which one mental state (a belief) is on one or few occasions of irrationality insensitive to thoughts about evidence (see Tenenbaum 2006: 247 and Olson 2011: 200). Rather, I mean mental states that (i) are continuously insensitive to stable explicit thoughts about there being sufficient evidence against their truth, and that (ii) they belong to a whole class of attitudes that are insensitive in this same way. Of course, the defenders of the considered three moral error theoretic views are not only committed to thinking that their own moral attitudes are systematically insensitive attitudes but also to thinking that everyone's moral attitudes are systematically insensitive attitudes in the same way.

The problem is that, according to the most popular theories of beliefs, if a subject *believes* that there seems to be sufficient evidence for the falsehood of her mental state and she can still continue to be in that state (and this happens systematically as defined above), then that mental state cannot count as a belief.<sup>18</sup> On

---

<sup>17</sup> This would seem to make moral thoughts similar to the dogmas held by religious fanatics (or akin to the religious faith one could adopt as a result of Pascal's wager type of arguments). In philosophy of religion, there are interesting analogical debates about whether such dogmas that are insensitive to thoughts about evidence should be thought of as mere faith where having faith consists of having some non-cognitive, practical attitudes rather than or having ordinary beliefs (see, for example, Malcolm 1977: 155).

<sup>18</sup> For a clear formulation of this requirement, see Tenenbaum (2006: 246). In other words, it is constitutive of beliefs that they are 'judgment-sensitive' attitudes (Scanlon 1998: 19, McDowell 1994: 60).

There is a worry that this thesis is too powerful. Consider, for example, Hume's arguments against the existence of external objects (Hume 1978: 187–193). Hume was well aware that, despite these arguments, we could not stop thinking that there are external objects around us (Hume: 1978: 183).

these views, it is constitutive of a belief-state that the state *is* sensitive to the believer's thoughts about the evidence for that belief. These views thus deny that systematically insensitive attitudes could count as beliefs. Thus, according to the standard accounts of beliefs, our first-order moral thoughts (as characterised by the central commitments of the three relevant error-theoretic views) could not count as beliefs.

This would entail that the Cognitivism and Semantic elements of the moral error theory would be false. The utterances of basic indicative moral sentences could not be expressing beliefs, and no one's basic moral beliefs could be ascribing moral properties because there would not be such beliefs. This would also make Falsity false for the same reason.

So, if the error theorist thinks (i) that Cognitivism, Semantic, Metaphysical, and Evidence are true; (ii) that one could continue to hold one's moral beliefs even when one believed (i); and she accepts (iii) the most plausible philosophical accounts of beliefs, then she must also think that Cognitivism and Semantic are false. She would now have to think both that moral utterances express beliefs and that they do not do so. This position would be blatantly inconsistent. The former claim is a central commitment of her view and the latter claim is entailed by her view together with the most plausible accounts of beliefs.

To avoid this inconsistency, the error theorists can do three things. The first option would be to adopt either descriptivist abolitionism or descriptive-normative revolutionary fictionalism. After all, these views deny that moral beliefs are systematically insensitive attitudes.

The second option would be to admit that our first-order moral judgments do not count as beliefs. These theorists could accept instead that our moral judgments are rather belief\*s. Such states are not sensitive to metaethical evidence but rather only to first-order moral evidence. So, even if people do not give up their moral belief\*s because they accept Evidence, they can give up their belief\* that eating meat is permissible on the basis of evidence about how cruel factory farming is. The moral error theorists could also add that belief\*s share many other features of beliefs such as that they play the same practical role in production of behaviour as other beliefs.

---

So, don't we here have a non-problematic case of systematically insensitive beliefs? Not according to Hume. He, of course, went on to argue that our 'beliefs' about the external objects should instead be understood as non-cognitive attitudes (or 'opinions') created by the faculty of imagination (Hume 1978: 198). In this sense, my argument below is Humean.

The problem is that this move amounts to giving up Cognitivism and Semantic which are the fundamental metaethical commitments of the moral error theory. Instead of being moral error theorists, the defenders of this move would now be non-cognitivists.<sup>19</sup> This is because the belief\*s would share all the central features of, for example, fictive attitudes which are typical non-cognitive attitudes. The case of fictive attitudes shows that the kind of internal sensitivity which belief\*s exhibit is not sufficient for being a cognitive attitude. My thoughts about Anna Karenina's character are sensitive to how I think she treats her husband Karenin, and these thoughts still remain at the level of make-belief. Likewise, emphasising the practical role played by belief\*s in the production of behaviour only makes these states more akin to the planning states described by the expressivists. The fundamental problem is that, if the moral error theorists describe belief\*s in these ways, the idea that the moral judgments of ordinary people are in any way mistaken is undermined. It no longer is the case that these states attempted and failed to correctly carve nature at its joints.

The third option would be to argue against the most plausible accounts of beliefs. This would require defending a view according to which systematically insensitive attitudes can count as beliefs after all. I will leave the challenge of formulating such a theory for the moral error theorists themselves. I will, however, explain next why the most plausible accounts of beliefs do not count systematically insensitive attitudes as beliefs.

## 5. THE NATURE OF BELIEFS

Our talk about beliefs and desires constitutes the core of 'folk psychology'. This commonplace theorising consists of our practice of explaining and predicting behaviour by ascribing these basic states to others. In such explanations and predictions, we use simple rules of thumb that connect these states to actions.

An essential property of beliefs and desires is that they are *intentional* states. That is, they are identified by their content, by what they are about. The content of beliefs and desires is usually thought to consist of propositions. A's belief that p thus consists of A being in the believing-relation to the proposition that p whilst A's desiring that p consists of her being in the desiring-relation to that same proposition.

---

<sup>19</sup> The moral error theorist could become expressivists who think that moral utterances express desire-like states (Blackburn 1993 and 1998, Gibbard 1990 and 2003). Or, she could become a fictionalist who thinks that moral utterances express some make-belief states (Kalderon 2005). Both these proposals face the so-called Frege-Geach problem (Oddie & Demetriou 2007, Olson 2011: sec. 3.2).

Propositions are usually understood to be structured semantic entities. This means that, because of the concepts that make up the proposition and the structure between them, a proposition will have conditions under which it would be true. In this situation, as attitudes, beliefs and desires can be distinguished from one another by the different ‘directions of fit’ of the attitudes to the states of affairs that are the truth-conditions of their propositional content (Anscombe 1957: 56, Platts 1979: 257). Beliefs aim at fitting their propositional content to how the world is, whereas desires aim at making the world to fit their propositional content.

This distinction between desires and beliefs in terms of what they aim at is metaphorical. The main challenge in this area is to specify more accurately what is required for an attitude to have the mind-to-world direction of fit of beliefs rather than the world-to-mind direction of fit of desires.<sup>20</sup> One will also have to be able to explain what distinguishes beliefs from the rest of the propositional attitudes such as suppositions, imaginings, hopes, memories, doubts, fears, and so on.

The most popular way of distinguishing beliefs from all the other attitudes is to characterise their unique functional role. I will begin by explaining why such views do not count systematically insensitive attitudes as beliefs. I will then consider whether the moral error theorists could use any of the other views about beliefs to defend their position.

### 5.1 Descriptive Functionalism

Descriptive functionalists believe that we can distinguish propositional attitudes by *their causal roles*. We are caused to have some of these attitudes by the world through perception, the attitudes in different combinations cause other attitudes in us, and finally some of the resulting attitudes cause us to act. It is then constitutive of the identity of any propositional attitude what kinds of other propositional attitudes (or situations) tend to cause one to have that attitude and what kind of other states (and actions) that attitude tends to cause one to have (or do).<sup>21</sup>

A defender of this theory of beliefs must thus be able to describe the causal role of beliefs and to explain how it differs from the roles of the other propositional attitudes. So far, these views have made the way in which beliefs react to our thoughts about evidence a defining feature of beliefs.

---

<sup>20</sup> See Searle (1983).

<sup>21</sup> See Fodor (1968), Lewis (1972), and Putnam (1975).

According to Michael Smith, the difference between beliefs and desires amounts to

a difference in the counterfactual dependence of a belief that *p* and a desire that *p* on a perception with the content that not-*p*... A belief with the content that *p* tends to go out of existence in the presence of a perception with the content that not-*p*, whereas a desire tends to endure, disposing the subject in that state to bring about that *p* (Smith 1994: 115).<sup>22</sup>

So, whether we should ascribe a belief or a desire with the content that *p* to a subject depends on what counterfactuals are true of her attitude. The truth of these counterfactuals is grounded on the different causal dispositions of the two types of attitudes – on what happens to beliefs and desires when we have perceptions with certain contents.

Arguably, in the case of moral beliefs, we will not be able to have straightforward *perceptions* with the negated content of the belief in the same way as in the non-moral cases. If I believe that burning cats for fun is wrong, it is not clear what *perception* I could have with the content that this is not wrong. So, Smith's account does not seem to apply directly to moral beliefs.

Yet, I believe that Smith directs our attention to the importance of the perception with the content that not-*p* in the definition of the causal role of beliefs because he is making a plausible back-ground assumption. According to it, a perception with the content that not-*p* constitutes the best evidence which we think we could have against the truth of the belief that *p*. In the light of this assumption, if we do not give up our belief that *p* when we have a perception with the content that not-*p*, then certainly our attitude with the content that *p* is not responsive to our thoughts about evidence.<sup>23</sup> And, it is just this general causal feature of beliefs – they are responsive to our thoughts about sufficient evidence – that sets them apart from the other propositional attitudes.

---

<sup>22</sup> Brian Loar (1981) also distinguishes beliefs from other attitudes on the basis of unique causal dispositions, which nicely match Smith's criteria. He claims that it is a central part of the nature of beliefs that, when one directs attention to perceptible properties of things, one begins to have the corresponding beliefs about the properties of those objects.

<sup>23</sup> Lloyd Humberstone's first way of understanding the notion of a perception with the content that not-*p* in Smith's view is to think of it as a belief that not-*p*. The problem with this option is that it specifies the causal role of beliefs in terms of beliefs (Humberstone 1992: 64, Sobel & Copp 2001: 46–47). Humberstone's second interpretation of Smith takes the relevant situation to be one in which it appears to the thinker that not-*p*. However, it's not clear whether beliefs are responsive to how things merely appear to be (Sobel & Copp 2001: 47–48). What is required is the further thought about having sufficient evidence for it being the case that not-*p*. This is why my interpretation matches Sobel and Copp's (ibid.: 49) most charitable reading of Smith, and Tenenbaum's reconstruction of Smith (Tenenbaum 2006: 242).

As we saw above, the moral error theorists under consideration think that moral beliefs in general are not sensitive to thoughts about the sufficient evidence for their falsehood. This is why, on their views, we need to consider the prudential reasons for our moral attitudes. So, if we apply Smith's functionalist account of beliefs also to non-perceptual domains, then our moral thoughts as systematically insensitive attitudes could not count as beliefs. This means that the moral error theorists under consideration could not consistently accept Smith's functionalist view of beliefs.

In response, the moral error theorists might attempt to adopt a weaker version of Smith's view. They might claim that a distinguishing feature of beliefs is merely a *disposition* to be given up when subjects believe that there is sufficient evidence against their truth. Thus, they might argue that thoughts about sufficient evidence for the falsehood of a belief merely provide some causal (and perhaps also rational) pressure to give up that belief. And, they might insist that this *disposition* is a feature of moral attitudes even if it will not be instantiated when one accepts Evidence.

This response faces a dilemma. The relevant disposition would either have to be a strong or a weak disposition. A weak disposition would not be enough to distinguish beliefs from other propositional attitudes. After all, hopes too are weakly disposed to be given up when there is sufficient evidence against their content's truth. We give up our childhood dreams when we realise that they will never become reality. Furthermore, the error theoretic views under consideration could not hold that moral attitudes are strongly disposed to be given up when one accepts Evidence. This would hardly be compatible with the idea that we will be able to decide whether to keep them on the basis of the prudential reasons. So, the relevant moral error theorists face a real challenge to give a plausible dispositional story of the way in which moral beliefs are sufficiently sensitive to evidence to count as beliefs such that it is compatible with their other commitments.<sup>24</sup>

---

<sup>24</sup> Perhaps Olson's idea of 'compartmentalisation' could be developed along these lines. On this proposal, the temporal nature of the dispositions matters. It could be argued that, for the first-order moral judgments to count as beliefs, it is enough that they are momentarily sensitive to the philosophical judgments about their falsehood when we consider the arguments against them. The fact that the first-order judgments keep coming back after this still leaves this hallmark of beliefs intact. So, on this view, a judgment that torture is wrong can count as a belief because one can give it up in philosophical reflection. This is so even if the intuition that torture is wrong vividly comes back to mind when one later on sees depictions of torture. One could then claim that beliefs only need to be momentarily sensitive attitudes even if they are systematically insensitive attitudes over time. One problem with this proposal is that it assumes that the naturally recurring and irresistible intuitions prompted by, for example, the depictions of torture really are beliefs rather than merely *sui generis* non-cognitive states which dispose us

Other functionalist accounts pose similar problems for the moral error theorists. On Lloyd Humberstone's view, whether a given propositional attitude counts as a belief depends on whether one has a special second-order background intention with respect to that attitude. He claims that

unless the attitude-holder has ... a controlling background intention that his or her attitudinizing is successful only if its propositional content is true, then the attitude is not that of belief (Humberstone 1991, 73).<sup>25</sup>

On this view, that everyone would have moral *beliefs* would require everyone to have a background intention that their moral attitudinizing is successful only if the content of their moral attitudes is true.

At this point of my argument, what is *not* at issue is whether everyone actually has that background intention. Perhaps everyone has those intentions and cognitivism thus really is true. Instead, what *is* at issue is whether the moral error theorists could consistently hold that everyone has the required kind of background intention in virtue of which our moral attitudes count as beliefs. I claim that the answer to this question is 'No'. This is why this plausible account of what beliefs are is not available for the relevant three forms of error theory.

Let us assume that everyone had a background intention that their moral attitudinizing is successful only if true. What sort of considerations would we in this situation take into account in deciding whether to keep our attitudes, to change them into make-believing, or to give them up – assuming now that we intend to have only successful attitudes? The best way to satisfy our background intention would in this case be to take only into account what we take to be evidence for the truth and falsehood of our attitudes.

In contrast, normative abolitionists, normative revolutionary fictionalists, and conservationists offer us reasons for giving up, keeping, or changing our attitudes that are wholly unrelated to the truth of these attitudes. This conflicts with the idea that, given the content of our assumed shared background intention, we would only take into account the truth-related considerations. So, unless the defenders of these views want to think that we are all choosing instrumentally irrational, inefficient means for satisfying

---

to adopt the relevant belief about the wrongness of torture. For this appealing understanding of intuitions, see Sosa (2007).

<sup>25</sup> For clarification and criticism, see Tenenbaum (2006: 248).



our background intention, they are forced, by their own lights, to give up the idea that we all have that required background intention. Or, in other words, the moral error theorists can ascribe the required second-order background intention to us only if they at the same time assume that we all are instrumentally irrational. This is something we should not assume.

Hence, two plausible standard descriptive functionalist accounts of beliefs fail to either count systematically insensitive attitudes of the considered kind as beliefs or be otherwise available for the moral error theorists. Yet, all the three moral error theoretic positions under discussion would have to be able to find a plausible account of beliefs which would count systematically insensitive attitudes as beliefs. For this reason, the moral error theorists will not be able to accept the standard functionalist views of beliefs.

The moral error theorists who want to be descriptive functionalists could attempt to specify the causal role of beliefs in a way that would enable systematically insensitive attitudes to be beliefs.<sup>26</sup> It is not clear how the resulting account could distinguish beliefs from all other propositional attitudes such as imaginings, assumptions, wonderings, fictive attitudes, and so on. In any case, it will be a real challenge for the moral error theorists to develop such an account.

## 5.2 Normative Theories

There are other accounts of beliefs which the moral error theorists could try to use to show that systematically insensitive attitudes can count as beliefs. I will begin the discussion of these views from two *normative* theories of beliefs.

---

<sup>26</sup> The error theorists could try to emphasise the idea that the causal role of a given attitude is defined by a large set of causal dispositions that form an interconnected network (Lewis 1972). The hope would then be that we can formulate a sufficiently substantial causal role for the beliefs in the network such that it does not require taking beliefs to be attitudes that are responsive to thoughts about evidence. For example, in describing this network, the error theorists could try to emphasise the output side of moral beliefs as a factor that makes these states belief-states. Until the moral error theorists provide a plausible sketch of such an account, I remain sceptical whether one could use the idea of a network of causal dispositions to distinguish beliefs from other propositional attitudes when sensitivity to thoughts about sufficient evidence isn't part of that network. Before we see details of such a view, that we could do so seems to me to be mere wishful thinking.

It could be also argued that Lewis' own version of functionalism allows some individuals to be in a given mental state even if, in that person, the state does not carry out the standard functional role of this state (Lewis 1980). The person who experiences 'Mad Pain' can be in pain without this because he is still in the same physical state as those people in whom that physical state has the standard consequences. This is response is not available for the relevant moral error theorists because they are committed to the claim that everyone's first order-moral judgments are systematically insensitive attitudes.

Both of the views leave room for cases of irrationality in which one does not give up one's beliefs even when one thinks that there is sufficient evidence against their truth (Zangwill 1998: 183). On these views, for believing that *p*, it does not matter whether one would as a matter of fact give up the given belief in such cases. Rather, what matters is what one *ought* to do then.

Call the first theory the *truth-norm account*. According to it, the difference between desires, beliefs, and other propositional attitudes lies in the fact that different norms apply to them.<sup>27</sup> The following norm is then claimed to be constitutive of beliefs:

**Ought<sub>truth</sub>** Believe that *p* only if it is true that *p*.

This norm entails that beliefs are by definition states that we *ought to* give up when their content is false. Desires, hopes and the like are not beliefs because we are not obliged to give them up when their content is false. All these attitudes are guided by some other constitutive norms.

If the truth-norm holds for beliefs and we have no direct way of conforming to it, how should we try to comply with it? It seems that one thing one must do in order to comply with this norm is to only have beliefs supported by evidence. Evidence just means here information that makes the truth of the relevant belief more likely. So, the truth-norm itself entails a further instrumental norm according to which one should believe that *p* only when one thinks that one has some information that makes it likely that *p* and not when one thinks one has some information that makes it unlikely that *p*.

As a result, if one thinks that one has good enough evidence against the truth of a given belief, then one ought to give up the belief. This would be the correct way of trying to comply with Ought<sub>truth</sub>. One would be, in this case, fitting one's beliefs to how likely one thinks it is that they are true.

The truth-norm thus creates an instrumental norm that requires not having beliefs such that one thinks that there is sufficient evidence for their falsehood. This norm entails that beliefs are uniquely vulnerable to criticism under certain standards of justification (Hieronymi 2006: 49–50, Smith 2004: 87). If you believe that *p*, you can be asked what evidence you have for that belief. If you cannot provide any, then you ought to give up your belief. Otherwise, it is not likely that you will believe that *p* only if it is true that *p*.

---

<sup>27</sup> This view is discussed in Zangwill (1998: sec. 1). See also Platts (1979: 257), Shah and Velleman (2005), and Wedgwood (2007: 162).

The second normative theory is Zangwill's *normative functionalism* (Zangwill 1998: sec. 4). It begins from the descriptive functionalist's causal networks for different propositional attitudes. The normative functionalist then claims that the causal connections between the mental states need not hold as a matter of fact. You can believe that p even if you take there to be enough evidence for the falsity of this belief.

However, according to normative functionalists, this causal connection *would* always hold, if you were *rational*. This means that we should think of the distinguishing roles of different propositional attitudes as rational requirements rather than as causal connections. So, if one believes that p, one need not actually give up this belief even if one thinks that there is enough evidence for its falsehood. But, rationality requires that one should give up the belief in that situation.

Thus, both the truth-norm account and normative functionalism agree on the following constitutive feature of beliefs:

**Ought<sub>internal</sub>** If you believe that p and think that there is sufficient evidence for the falsity of that belief, then you ought to give up your belief that p.

Ought<sub>internal</sub> seems to be compatible with the versions of the moral error theory under discussion. They only claim that one could continue to believe that p even when one accepts Evidence. As such, this central element of those views is neutral about what one *ought* to do in that situation.

Furthermore, both normative abolitionism and normative revolutionary fictionalism agree with Ought<sub>internal</sub> that we should give up our moral beliefs when we think that there is sufficient evidence against their truth. Of course, these views disagree with the Ought<sub>internal</sub> principle about why exactly we should give up these beliefs. Normative abolitionism and normative revolutionary fictionalism claim that we should do so for prudential reasons, whereas Ought<sub>internal</sub> suggests that we should give up our moral beliefs on the basis of the evidence we have against their truth.

Initially, it might seem less clear how conservatism could be compatible with accepting Ought<sub>internal</sub>. After all, conservationists reject the thesis that we ought to give up a moral belief when we think that there is sufficient evidence for its falsehood. So, it seems as if they could not accept Ought<sub>internal</sub>.

In response to this challenge, conservationists could claim that the ought in the Ought<sub>internal</sub> principle is 'an ought of a constitutive norm for the state', whereas when they discuss the prudential reasons for keeping the first order moral beliefs, they are

talking about wholly different kind of ‘practical oughts’. In this way, conservationists could try to argue that there is no conflict between their position and Ought<sub>internal</sub>.

I believe that these normative proposals unfortunately fail to save the relevant versions of the moral error theory for a simple reason. There must be some explanation for why a normative principle such as Ought<sub>internal</sub> would apply to some propositional attitudes with the content that p but not to others (Zangwill 1998: 175, Sobel & Copp 2001: 52). Furthermore, the standard explanations available for the moral error theorists would all seem to entail that systematically insensitive attitudes could not count as beliefs. This means that, unless the moral error theorist is able to come up with a completely new kind of an explanation for why Ought<sub>internal</sub> applies to some attitudes but not to others, we still lack a plausible account of beliefs that would classify systematically insensitive attitudes as beliefs.

It is widely accepted that the normative supervenes on the non-normative.<sup>28</sup> There cannot be a difference in the normative properties of two things without a difference in their non-normative properties. So, take an agent’s belief that p and her desire that p. The current proposal is that what makes the first attitude a belief and the second a desire is that the agent ought to give up the first attitude but not the second when she thinks that there is sufficient evidence for not-p.<sup>29</sup>

Given supervenience, that what the agent ought to do differs with the respect to the two attitudes requires that there is some more basic non-normative difference between them. Something non-normative about the attitudes themselves must explain why one ought to give up one of the attitudes but not the other in the identical situations. This means that Ought<sub>internal</sub> can distinguish beliefs from desires only if there is already some prior non-normative difference between the two kinds of attitudes.

One appealing way of specifying why Ought<sub>internal</sub> applies to a given attitude but not to another with same content is to refer to Humberstone’s higher-order intention for having this particular attitude only when it is true. Having such an intention directed

---

<sup>28</sup> See, for instance, Blackburn (1993: chs. 6–7), Smith (1994: 22 and 40), Jackson (1998: 118), and Shafer-Landau (2003: 77).

<sup>29</sup> As suggested above, the ought in question here is an ought of the constitutive norms for the state and not a practical or an overall ought. I am sceptical about whether there are different kinds of oughts. It seems to me that there is only one kind of oughts and only many different sources of oughts. For this reason, Ought<sub>internal</sub> would perhaps be best formulated in terms of reasons. We would then have that it is constitutive of beliefs that, if you believe that p and also believe that there is sufficient evidence against p, then you have strong epistemic reasons to give up your belief that p. This formulation allows us to also recognise that there can be pragmatic reasons for beliefs too.

towards a given attitude would explain why one ought to give up a given attitude when one thinks there is enough evidence against its truth.<sup>30</sup>

Now, return to the moral error theories. As we saw in the previous section, they are unable to hold that we would have the required sort of higher-order intention with respect to our moral beliefs. This would entail that, for them, Ought<sub>internal</sub> could not apply to these states and thus they would not count as beliefs under the present proposal. So, the resulting normative theory of beliefs would not be compatible with the relevant forms of moral error theory under discussion.

Of course, the moral error theorist could try to find some other non-normative explanation for why Ought<sub>internal</sub> applies to some states but not to others. Obviously, this explanation could not just state that Ought<sub>internal</sub> applies to moral beliefs because they are beliefs. This would be a blatantly circular way of explaining why these states have the status of being beliefs. This explanation should also enable the moral error theorists to count the first-order moral insensitive attitudes as beliefs without making desires, hopes, imaginings, guesses and the like also beliefs. It is hard to imagine what such an explanation would be like. So, it does not seem like the normative accounts of beliefs help the moral error theorists' case.

### 5.3 Other Views

Of course, the various forms of functionalism and the normative theories are not the only accounts of beliefs which the moral error theorists could try to use in response to my challenge. I will finish off by briefly commenting on three remaining standard accounts of beliefs.

*Representationalists* believe that, metaphysically, beliefs consist of a representation (a structured proposition) being physically instantiated in the brain in some way. They think that brains are 'biological machines' that manipulate these representations. The brain will, however, manipulate the same representation in different ways depending on the form in which the representation is stored. Thus, on Fodor's view, believing that *p* consists of the sentence of language of thought *that p* being stored in a 'belief-box' rather than in a 'desire-box' (Fodor 1975 and 1987).

---

<sup>30</sup> A view along these lines is defended in Millar (2009). According to him, believers must have the higher-order intention to believe only truths as a result of recognising that the Ought<sub>internal</sub> applies to beliefs.

The defenders of this account need to explain what distinguishes the same representation being tokened in a belief-box from it being tokened in a desire-box. So far, the representationalists have based such explanations on the different functional roles which a given representation can have in the brain (see *ibid.*, Millikan 1984, and Dretske 1988). These functional descriptions have thus far resembled the functionalist accounts already discussed. Thus, a sentence that *p* being in a belief-box has been thought to require that the representation is sensitive to our environment and also to our thoughts about the evidence for the given representation.

A moral error theorist drawn to representationalism could not accept this. Her view would require that a representation could be in a belief-box even if it were systematically causally insensitive both to one's environment and to one's representations about the evidence for that state. So, she would need to provide a new account of the conditions under which a mental representation is in a belief-box.<sup>31</sup> Thus, representationalism itself will not help the moral error theorists.

The views discussed so far assume that beliefs exist in minds in some robust sense. *Interpretationalists* deny this (Davidson 1984, Dennett 1987 and 1991). They think that propositional attitudes are merely posits which we use to understand the linguistic and non-linguistic behaviour of others. They claim that there are no hard facts that would make ascriptions of the propositional attitudes robustly correct in some further evidence-transcendent sense.<sup>32</sup>

Let us begin from the utterances of another person. According to Davidson, in order to understand them, I need to be able to know the conditions under which the uttered sentences would be true. However, I could only know the truth-conditions of these utterances if I would be able to decipher which propositional attitudes the utterer has. This is because the truth-conditions of her utterances depend on what she intends

---

<sup>31</sup> Admittedly, on Millikan's view, states are individuated by the function which they played in the historical context of evolutionary selection (Millikan 1984). It is compatible with the fact that a state evolved to play a given functional role that it no longer plays that role in our psychology. So, on this view, it could be that our moral judgments count as beliefs because they played a belief-like functional role in the past even if they no longer do so. This view faces the famous swampman problem. According to it, a newly created physical duplicate of me could not have the same moral beliefs as I do.

<sup>32</sup> This is a consequence of the 'indeterminacy of interpretation' according to which there are always many different, empirically equally good interpretations of agents' behaviour in terms of their propositional attitudes (see below). Davidson thought that this means that the different interpretations are merely 'notational variants' of one another, and so therefore there is no further question of which one of them is correct. When we consider Crispin Wright's other standards of anti-realism (see Wright 1992), the interpreters who come to hold different but empirically equally accurate interpretations cannot be accused of cognitive shortcomings either. Given the role of the interpreters' attitudes in applying the principle charity, facts about propositional attitudes would also be judgment-dependent.

to communicate, what she takes her words to mean, and how she believes the world to be. The problem is that I will not be able to know what propositional attitudes the other person has by asking her unless I already know what her utterances mean.

Davidson thought that, to solve this problem of how to understand simultaneously what the utterances of the other person mean and to know what attitudes she has, I must interpret the other person according to the principles of charity. I must assume that she satisfies the basic requirements of rationality and coherence, that most of her beliefs are true, and that most of her desires are for good things.

Without these assumptions, there would be infinitely many options of how to interpret others that would be equally compatible with all our empirical evidence about their behaviour. Even with these assumptions, we can often come up with many equally sensible interpretations of others in terms of the different combinations of propositional attitudes which they could have. In these cases, according to Davidson, there is no further fact about what attitudes others really have.

Could the moral error theorist use this notion of beliefs to save the idea that the systematically insensitive attitudes could count as beliefs? Unfortunately not. If the principle of charity is a constraint for understanding others, then we cannot generally ascribe to others both the belief that *p* and also the thought that there is sufficient evidence against the truth of the belief that *p*.<sup>33</sup> Yet, the three error theoretic views under discussion are all committed to thinking that one could continue to hold a moral belief even if one came to think that there is sufficient evidence against that belief. So, the possibility of that interpretation should always be available.

If such interpretations were acceptable, then, if we observed a biologist seeing a black swan and uttering ‘that’s a sufficient reason to believe that not all swans are white’, nothing would yet determine whether we should ascribe to him the thought that not all swans are white. We still could not make sense of what this person believes. This seems very counterintuitive. Hence, making sense of others requires that we assume that their beliefs are sensitive to their thoughts about sufficient evidence. If the moral error theorists reject this assumption as a constraint of interpretation, they need to provide us some alternative way of interpreting others that would fit their view.

---

<sup>33</sup> Admittedly, the moral error theorist is already violating the principle of charity in ascribing a large class of false moral beliefs to all of us. Moral error theorists tend to respond to this worry by claiming that other metaethical views too ascribe many false beliefs to individuals. Non-cognitivists, for example, seem to have to claim that people falsely believe that they have moral beliefs. Non-cognitivists can deny this for a good reason. They can claim that the minimal folk concept of beliefs correctly applies to moral beliefs even if these states do not meet the philosophical, functional criteria of belief-states.

Perhaps the best way for the moral error theorists to explain how the systematically insensitive attitudes could count as beliefs is *dispositionalism*. On this view, whether a given mental state counts as a belief that p has nothing to do with what other propositional attitudes the agent has – not even with what evidence she takes there to be for or against the belief.

Rather, according to dispositionalism, whether a person believes that p depends only on whether she possesses one or more of the behavioural dispositions pertaining to p (Marcus 1990). These dispositions include the disposition to utter that p when prompted, to be surprised should the falsity of p become evident, to assent to q when that p entails that q, and to depend on the truth that p in one's plans. The moral error theorists could argue that we can have all these dispositions even when we also think that there is sufficient evidence against the truth of that p. This would enable one to have moral beliefs even if one accepted Evidence.

Unfortunately, such accounts of beliefs are implausible. People with the very same belief can act in completely different ways depending on what other attitudes they have (Chisholm 1957). For example, one might think that, if one believes that it is hot outside, one will be disposed to wear cool clothing. However, a person who believes that it is hot outside will only be disposed to wear shorts if he doesn't want to get hot. But, someone could want to be hot and thus wear warm clothing even when he thinks that it is hot outside. The dispositionalists would get the beliefs of such persons wrong on the basis of their behaviour.

Because of such basic problems, most philosophers reject dispositionalism. A moral error theorist who would want to save her view by understanding moral beliefs in the atomistic dispositional way would need to solve these problems. She would also have to be able to describe the behavioural dispositions of those who hold a particular moral belief in a way that would allow distinguishing that state from all other possible propositional attitudes. These behavioural dispositions would furthermore have to remain unaffected no matter what other beliefs and desires the agent came to have. Presumably finding a suitable theory of moral beliefs of this sort will be overwhelmingly difficult.

## 6. CONCLUSION

In the second section, I explained that moral error theorists accept Cognitivism, Semantic, and Metaphysical. They must also think that these allegedly true theses



together seem to be sufficient evidence against the truth of moral beliefs. As a result, moral error theorists must accept Evidence.

In the third section, I suggested that the moral error theorists can react in five different ways to the question of what happens to our moral beliefs when we accept the previous four theses. Two of the resulting positions (descriptive abolitionism and descriptive-normative revolutionary fictionalism) predict that, when one comes to accept Evidence, one will give up one's first-order moral beliefs. I set these views aside, because, currently, we have no way of knowing whether that empirical commitment of these views is true or false.

The three remaining options (normative abolitionism, normative revolutionary fictionalism, and conservatism) all hold that, when one accepts Cognitivism, Semantic, Metaphysical, and Evidence, one option one will still have will be to keep one's first-order moral beliefs. These views then go on to debate whether, in this case, there are more prudential reasons to keep the moral beliefs, to give them, or to adopt some kind of new make-belief attitudes.

In the last two sections, I argued that this makes these three remaining moral error theoretic positions problematic. Four of the standard philosophical accounts of beliefs (descriptive functionalism, the normative theories, representationalism, and interpretationalism) are not available for the moral error theorists because they will not count systematically insensitive attitudes as beliefs. Yet, all these moral error theoretic positions require that they would count as such. So, these views will not be able to accept the standard philosophical positions of beliefs on the pain of inconsistency.

This means that the moral error theorists who accept any one of these three views will either have to reformulate radically one of the standard accounts of beliefs, solve the problems dispositionalism, or introduce some new theory of beliefs. All these options will have their problems.

#### **References:**

- Alston, William (1989), 'Concepts of Epistemic Justification', in his *Epistemic Justification*. Ithaca: Cornell University Press.
- Anscombe, G.E.M. (1957), *Intention*. Oxford: Blackwell.
- Armstrong, D.M. (2010), *Sketch for a Systematic Metaphysics*, Oxford: Oxford University Press.
- Bedke, Matt (2011), 'Normative Judgments Are Not Beliefs about Non-Natural Properties', an unpublished article manuscript.
- Blackburn, Simon (1993), *Essays in Quasi-Realism*. Oxford: Oxford University Press.
- Blackburn, Simon (1998), *Ruling Passions*. Oxford: Oxford University Press.

- Burgess, John (2007), 'Against Ethics', *Ethical Theory and Moral Practice*, 10: 427–39.
- Chisholm, Roderick (1957), *Perceiving*. Ithaca: Cornell University Press.
- Daly, Chris and Liggins, David (2010), 'In Defence of Error Theory', *Philosophical Studies*, 149: 209–230.
- Davidson, Donald (1984), *Inquiries into Truth and Interpretation*. Oxford: Clarendon Press.
- Dennett, Daniel (1987), *The Intentional Stance*. Cambridge, Ma.: MIT Press.
- Dennett, Daniel (1991), 'Real Patterns', *Journal of Philosophy*, 87: 27–51.
- Dretske, Fred (1988), *Explaining Behaviour*. Cambridge, Ma.: MIT Press.
- Fodor, Jerry (1968), *Psychological Explanation*. New York: Random House.
- Fodor, Jerry (1975), *The Language of Thought*. New York: Crowell.
- Fodor, Jerry (1987), *Psychosemantics*. Cambridge, Ma.: MIT Press.
- Garner, Richard (2007), 'Abolishing Morality', *Ethical Theory and Moral Practice*, 10: 499–513.
- Gibbard, Allan (1990), *Wise Choices, Apt Feelings*. Cambridge, Ma.: Harvard University Press.
- Gibbard, Allan (2003), *Thinking How to Live*. Cambridge, Ma.: Harvard University Press.
- Hieronymi, Pamela (2006), 'Controlling Attitudes', *Pacific Philosophical Quarterly*, 87: 45–74.
- Hinckfuss, Ian (1987), 'The Moral Society: Its Structure and Effects', *Discussion Papers in Environmental Philosophy*, 16.
- Humberstone, Lloyd (1993), 'Direction of Fit', *Mind*, 101: 59–83.
- Hume, David (1978) [1739–1740], *A Treatise on Human Nature*, edited by L. A. Selby-Bigge and P. H. Niddich. Oxford: Clarendon Press.
- Jackson, Frank (1998), *From Metaphysics to Ethics*. Oxford: Oxford University Press.
- Joyce, Richard (2001), *The Myth of Morality*. Cambridge: Cambridge University Press.
- Joyce, Richard (2006), *The Evolution of Morality*. Cambridge, Ma.: Harvard University Press.
- Kalderon, Mark (2005), *Moral Fictionalism*. Oxford: Oxford University Press.
- Lewis, David (1972), 'Psychophysical and Theoretical Identifications', *Australasian Journal of Philosophy*, 50: 249–58.
- Lewis, David (1980), 'Mad Pain and Martian Pain', in N. Block (ed.) *Readings in the Philosophy of Psychology, Vol. 1*. Cambridge, Ma.: Harvard University Press.
- Loar, Brian (1981), *Mind and Meaning*. Cambridge: Cambridge University Press.
- Mackie, John (1977), *Ethics – Inventing Right and Wrong*. London: Penguin Books.
- Malcolm, Norman (1977), 'The Groundlessness of the Belief in God', in S.C. Brown (ed.) *Reason and Religion*. Ithaca: Cornell University Press.
- Marcus, Ruth (1990), 'Some Revisionary Proposals about Belief and Believing', *Philosophy and Phenomenological Research*, 50: 132–153.
- McDowell, John (1994), *Mind and World*. Cambridge, Ma.: Harvard University Press.
- Millar, Alan (2009), 'How Reasons for Action and Belief Differ', in S. Robertson (ed.) *Spheres of Reason*. Oxford: Oxford University Press.
- Millikan, Ruth (1984), *Language, Thought and Other Biological Categories*. Cambridge, Ma.: MIT Press.
- Nietzsche, Friedrich (1990), *Beyond Good and Evil*. Trans. by R.J. Hollingdale. Harmondsworth: Penguin.
- Nolan, Daniel, Restall, Greg, and West, Caroline (2005), 'Moral Fictionalism Versus the Rest', *Australasian Journal of Philosophy*, 83: 307–30.

- Oddie, Graham and Demetriou, Daniel (2007), 'The Fictionalist's Attitude Problem', *Ethical Theory and Moral Practice*, 10: 485–498.
- Olson, Jonas (2010), 'In Defence of the Error Theory', in M. Brady (ed.) *New Waves in Metaethics*. Basingstoke: Palgrave MacMillan.
- Olson, Jonas (2011), 'Getting Real about Moral Fictionalism', *Oxford Studies in Metaethics*, 6: 181–204.
- Pigden, Charles (2007), 'Nihilism, Nietzsche, and the Dobbelganger Problem', *Ethical Theory and Moral Practice*, 10: 441–56.
- Platts, Mark (1979), *Ways of Meaning*. London: Routledge and Kegan Paul.
- Putnam, Hilary (1975), *Mind, Language, and Nature*. Cambridge: Cambridge University Press.
- Railton, Peter (1986), 'Moral Realism', *Philosophical Review*, 95: 163–207.
- Robertson, Simon (2008), 'How to Be an Error Theorist about Morality', *Polish Journal of Philosophy*, 2: 107–25.
- Scanlon, T.M. (1998), *What We Owe to Each Other*. Cambridge, Ma.: Harvard University Press.
- Searle, John (1983), *Intentionality – An Essay in Philosophy of Mind*. Cambridge: Cambridge University Press.
- Shafer-Landau, Russ (2003), *Moral Realism – A Defence*. Oxford: Oxford University Press.
- Shafer-Landau, Russ (2005), 'Error Theory and the Possibility of Normative Ethics', *Philosophical Issues*, 15: 107–20.
- Shah, Nishi and Velleman David (2005), 'Doxastic Deliberation', *Philosophical Review*, 112: 497–534.
- Smith, Michael (1994), *The Moral Problem*. Oxford: Blackwell.
- Smith, Michael (2004), 'A Theory of Freedom and Responsibility', in his *Ethics and the A Priori*. Cambridge: Cambridge University Press.
- Sobel, David and Copp, David (2001), 'Against Direction of Fit Accounts of Desires and Beliefs', *Analysis*, 61: 44–53.
- Sosa, Ernest (2007), 'Intuitions: Their Nature and Epistemic Efficacy', *Grazer Philosophische Studien*, 74: 51–67.
- Streumer, Bart (2012), 'Can We Believe in Error Theory?', forthcoming in *The Journal of Philosophy*.
- Tenenbaum, Sergio (2006), 'Changing One's Direction of Fit', *Oxford Studies in Metaethics*, 1: 235–64.
- Wedgwood, Ralph (2007), *The Nature of Normativity*. Oxford: Oxford University Press.
- Williams, Bernard (1973) 'Deciding to Believe', in his *Problems of the Self*. Cambridge: Cambridge University Press.
- Wright, Crispin (1992), *Truth and Objectivity*. Cambridge, Ma.: Harvard University Press.
- Zangwill, Nick (1998), 'Direction of Fit and Normative Functionalism', *Philosophical Studies*, 91: 173–203.