Moral Worth: You Can't Have it Both Ways
Nomy Arpaly

Some say that concern for morality de dicto grants right actions moral worth. That is, they say that if you do the right thing because of your concern to do the right thing, your action has moral worth (and you are worthy of esteem for that action). Some say that concern for morality de re grants moral worth - that is, they say that if you do the right action for the reasons that make it right (for example, because it protects wellbeing and respects autonomy) then your action has moral worth. Increasingly, some argue, and many seem to think, that both concern for morality de dicto and concern for morality de re grant right actions moral worth[1] In this paper, I will argue that the last position is false. If de dicto concern grants moral worth, de re concern does not, and vice versa.

Huck Finn Cases: a Few Clarifications
How do you tell a person who consistently acts well and is intrinsically motivated by morality de dicto apart from a person who consistently acts well and is intrinsically motivated by morality de re?

The question is quite schematic, as the good people you know are motivated to some extent by both, and by many other things beside. However, the schematic question has a clear schematic answer. Imagine that both our schematic people – we might call them Diana and Deirdre – have just been convinced by a very clever teacher or book (or a nefarious neuroscientist, if you like this kind of thing) that morality requires F-ing, an action that until now they rightly refused to do. In this case, Diana de Dicto will alter her behavior and start F-ing, while Deirdre de Rey will continue to refuse to F. If asked, she might say something along the line of "screw morality, I am not going to do *that*" or "if this is wrong, I don't want to be right". This is the litmus test.

---

[1] The view that de re - but not de dicto - moral motivation grants moral worth is defended, for example, by Arpaly (2002), (2003), Arpaly and Schroeder (2013), Cunningham (2021), Fearnley (2022), Lockhart (2017), Markovits (2010), (2012), Massoud (2016), Way (2017) and Weatherson (2019). The view that de dicto – but not de re – moral motivation grants moral worth is strongly suggested by as well as strongly suggested by Herman (1993) and defended by Johnson King (2018). Sliwa (2012) defends a connection between moral worth and moral knowledge that also implies the importance of de dicto, but not de re, moral motivation. An influential argument that concern for morality de dicto is actually a bad thing comes from Smith (1994), who thinks of it as "moral fetishism". Many have defended concern for morality de dicto from this charge – for example, Aboodi (2017), Carbonell (2013), Copp, (1997), Lerner (2018), Lillehammer (1997), Olson (2002), Svarvarsdottir (1999), without necessarily rejecting concern for morality de re as having a role in moral worth. Explicit defense of the view that concern for morality de re and concern for morality de dicto are each sufficient for moral worth can be found in Hurka (2014) Isserow (2020) and arguably Johnson King (2019). For recent views of moral worth that are harder to classify see, for example, Portmore (2022) and Singh (2020).

I will argue that when Diana and Deirdre (before being led astray) both act rightly, only one of them, at most, is acting with moral worth. The argument will rely on the implications of the two views – the view that concern for morality de dicto is sufficient for moral worth and the view that concern for morality de re does - when it comes to what I will call *Huck Finn cases* - a term I will use more broadly than philosophers generally seem to use it.

I will use the term "Huck Finn case" to refer to any case in which a person performs her right action for the reasons that make it right but believes that that the action is wrong[2]. In other words, the agent in a Huck Finn case acts out of concern for morality de re - despite his concern for morality de dicto telling him to act differently.

This is a possible interpretation of the literary character Huck Finn, who helps his black friend Jim escape slavery despite not being smart enough to doubt his society's conviction that helping a slave escape is wrong. I have suggested in previous work that Huck, when he finds himself emotionally unable to turn Jim to the people searching for him, is motivated by his gradual discovery of Jim's humanity – his being similar to him and to other friends he might want to help – and that, if so, his action has moral worth. If one does not accept my interpretation of the fictional character's motives, one can still agree that if these were his motives, his action would have moral worth (reminder: in this work I do not mean to further defend the view that Huck is worthy of esteem or defend team de re – my purpose is to argue for the incompatibility of the view that concern for morality de dicto suffices to give right actions moral worth with view that concern for morality de re suffices).

Huck Finn cases, as I define them here, are not always this dramatic. First, Huck Is often portrayed as experiencing "last ditch akrasia" – he finds himself helping Jim right as he tells himself to turn him in, or close enough. There are many cases of gaps between people's actions and their convictions as to what they should do that should be seen as more mundane, as with the immense number of people who think they should go to the gym and never or hardly ever do. An alien who tried to learn about humans from philosophy books would have easily come to the conclusion that most people who think they should go to the gym do go, and there are only a few "deviant" cases that need explaining. If, on the other hand, the alien tried to learn about us from self-help books, the alien would have concluded that *most* people who think they should go to the gym do not, and the question about the remaining people is "what is their secret?". In the same vein, the common Mormon slang term "Jack Mormon" refers to a person who believes the propositions that the church of Later Day Saints endorses, including the normative ones, but does not act as it prescribes. Many value systems have their "Jacks" – and sometimes the Jacks practice *better* than they preach: there are Jack-Christians who do not love their neighbor, but there are also Jack-Ayn Rand fans who do. For me, a Huck Finn need not be akratic – he can be Jack.

---

[2] Perhaps "right" should, strictly speaking, be replaced by "morally desirable", to cover supererogatory actions, but I'll stick to "right" for now.

Huck Finn cases do not have to be dramatic also in the sense that they need not involve strong emotions, which Huck Finn himself feels. A common error in the literature on concern for morality de dicto and de re is mistaking de re motivation for emotional motivation and de dicto moral motivation for motivation through cold resolution[3]. Both concern for morality de dicto and concern for morality de re can be felt vividly, which results in emotional reactions, or non-vividly, through a layer of fatigue or depression or preoccupation with something else, say, which results in cold affect surrounding one's deliberations. Both of them can be wholehearted or in conflict with other motivation. Whether we are talking about concern for morality de dicto or de re, "he cares about morality" can be a true sentence to say about a sleeping person, so it is possible, at least for a while, to have these concerns without feeling anything at all.

So let us look at a relatively undramatic Huck Finn case.

Ted says that he does not care that much about morality. "Maybe I'm not a moral guy. I drink, I gamble, I smoke, I have sex, I don't even mind it that my brother is gay". When asked what he does care about, Ted says, sincerely, that he wants people not to be miserable and he wants his friends to be happy. Furthermore, he wants to live and let live – not to interfere with other people's lives. Ted, let assume, cares about the morally important things de re, but he is a bad ethicist: while he might agree with us, for example, that murder is wrong, his general view of morality is hopelessly false. Now, on one occasion, let us imagine, Ted sees that his friend Clarissa is having a crisis. He knows her well and correctly believes that it would do her good to have a drink with him, which would help her relax a bit and talk about the crisis openly. He wants to protect her wellbeing, so he buys both of them drinks, despite believing that consuming alcohol (or, presumably offering it to another) is immoral. Let us assume that Ted is doing the right thing in buying the drink and that, as he does so to protect Clarissa's wellbeing, he does it for the reasons that make It right. That would make his story a Huck Finn case.

Some doubt the existence of Huck Finn cases because they think Ted, for example, does not really think about morality if his idea of paradigmatic immoral action is promiscuous sex and not murder, or if such sex and murder both look paradigmatically wrong to him. Ted, the objectors say, is using the word "morality" differently from you and me.

The trouble with saying that is that too many people are thus excluded from the community of people who think and talk about morality – including some renowned moral philosophers.

The extent to which Immanuel Kant sympathized with the view that children born out of wedlock do not deserve the protection of the law is debated, but even his taking it seriously makes him feel, to me, at least as alien as Ted. Kant also believed that selling one's hair is wrong, that you should tell the truth to the murderer at the door, and, as befitted his time, that masturbation is a grave wrong, homosexuality is a grave wrong, and women should obey their husbands.

---

[3] As when Lillehammer (1997) seems to assume that a person who does not, at a given moment, *feel* sympathy for her husband cannot act for the sake of sparing his feelings without the mediation of concern for morality de dicto.

Kant's more odious views, with the exception of his view regarding the murderer at the door, are now mostly forgotten –at least partially because, perhaps again with the exception of his view regarding the murderer at the door, these odious views seem very clearly *not* to follow from his theory and, accordingly, not to give his theory any trouble. However, he did have those views, and interestingly to me, knowing that he had them does not lead us to say that he was *not talking about morality* when he did moral theory, or that, if he kept his promises and said he did it for moral reasons, he was somehow talking about something other than what I talk about when I say I kept a promise for moral reasons. We see it as reasonable to learn about morality from the plausible and wise things Kant said about it, and Kant said a lot of such things, even though this fact would have been of no comfort to any gay for whose oppression he might have voted.

So it seems that Kant talked about morality in the same sense of "morality" that we do – in the same way that before Einstein, people already talked about gravity, despite having very different beliefs about it. Gravity for Einstein is quite different from gravity as it appeared to Newton, but they still theorized about the same thing – gravity. Now, I imagine Ted as a contemporary and a non-philosopher, but it's hard to see why he would not count as "really" talking about morality if philosophers in the 18th century and earlier count as having talked about it. They, too, after all, held some actions to be paradigmatically immoral that are not. Nothing Ted believes is any more repugnant than doubting that "bastards" should have as much of a legal right to life as anyone else or any sillier than the belief that masturbation is a grave wrong.

Furthermore, saying that some people do not mean the same by "morality" as others do is to say that some people cannot have a meaningful disagreement with some others about morality: they would be talking past each other. Using this test, we can ask: did garden variety Nazis talk about morality when they said they had a moral duty to kill Jews? Did supporters of slavery talk about morality when they said they had a moral duty to return slaves to their owners? Certainly. A slave owner and an abolitionist could argue about whether or not slavery is wrong without talking past each other, and Nazis could meaningfully disagree with anti-Nazis.

In addition to those who say that those with odious moral beliefs do not refer to morality when they say "morality", there are those who suggest that they are not competent or reasonable holders of moral beliefs. If I were trying to argue that concern for morality de dicto is bad, in a consequentialist kind of way, it would have been tempting to reply, in the spirit of Mill, that it is only a bad thing if one assumes universal idiocy along with it. I have no intention of arguing that concern for morality de dicto is bad, but I must say that it is sadly shown by history and the news that the idiocy required for the development of atrocious moral beliefs is just the idiocy entailed by being human. To argue credibly and without begging the question about the Nazis, the supporters of American slavery, and the supporters of killing young women who "shamed" their family by having sex that they all were, or are, somehow epistemically incompetent with regard to moral belief, one needs to show evidence of incompetence other than the falsity of their beliefs. What would that evidence be? It is not the case that the southern states of America or half of the German population suddenly suffered *en masse* from a mental disorder

or a neurological disability in the relevant time periods. What's true for the more transparently atrocious moral beliefs – those that directly endorse atrocious behavior - is also true for moral beliefs that fall short of that but are no less off-base, like the belief that premarital sex is wrong or that homosexuality is wrong. Were all cultures who univocally believed these things made of incompetent people? If one is not careful, one might come to the conclusion that competent moral believers are a minority.

I will assume, then, that there exist genuine Huck Finn cases, which do not involve incompetence or mistaken use of moral terms, and my argument against the both-ways view of moral worth will rely on intuitions about such cases. Before I start, though, it is important to emphasize: saying that a motive grants moral worth is separate from a lot of other good things you can say about it. One can talk, for example, as Kant did, about motivations that deserve "praise and encouragement, but not esteem". When Kant says that some motives deserve praise he does not seem to mean that the action is what many contemporary writers mean by "praiseworthy" but merely that praise, like encouragement, is a *morally desirable* response to the display of these motives, at least when they result in right actions: we should praise people with those motives, and encourage them – but we must only *esteem* people who act from duty, as only they act with moral worth. For the Kantian, *only* the motive of duty grants moral worth – but one is obligated, just in case, to develop in oneself helpful inclinations, such as compassion and a sense of honor, which are likely to lead one to do the right thing without a need for the motive of duty to kick in.[4]

In a similar way, a person who thinks that only concern for morality *de re* grants moral worth – whether she thinks morality is about utility, universalizability, respect for persons or some other thing or things – might hold that a combination of concern for morality de dicto and some serious amount of moral knowledge is a nice back-up motive, seeing as it would often result in right actions. That a motive results in right actions with good regularity does not make it a moral worth-granting motive. Moral worth comes not from statistical reliability or modal robustness but from the tight relationship between the content of one's motive and morality, a relationship that does not exist if one's motive is the desire for a tax deduction. Team de dicto thinks the requisite relationship is in the agent's *will to obey morality* and team de re thinks that the relationship is in the agent being motivated by *the reasons for which the action is moral*. Thus they attempt to accommodate the same intuition, often described in terms of non-accidentality.

On to the argument.

The Argument: Part One.

Let us start from a Kantian intuition I like to call "the Jewel Intuition", but which can also be referred to as "at least you tried". This is the intuition Kant appeals to when he says that a good will shines "like a jewel" even if it fails to "achieve its goals". Kant says this at the beginning of

---

[4] See Kant *1998/1785*

the *Groundwork,* before laying out his theory, and there is no need to be a Kantian to share his intuition.

Presumably, if you act out of good will and you act rightly, your action has moral worth. But,as my undergraduate students like to ask, what happens when you act wrongly out of good will? That can happen – for example, if your will is good but, through no fault of your own, your intelligence isn't high – and the Jewel intuition suggests that in such a case, you are *not blameworthy* for your wrong action. You are excused. How can you be condemned if "you tried"? Of course, you need to have *honestly* tried: no self-deception, mixed motives, motivated irrationality, culpable ignorance, and so on.

What is true for "the good will" in the abstract is true for any motive that grants moral worth. Imagine for a moment that Team de re is correct: a morally worthy action is a right action performed for the reasons that make is right. Imagine that utilitarians are also correct: what makes an action right is its expected contribution to utility. In this case, the will to maximize utility grants right actions moral worth, and if you did your honest best to maximize utility but failed due to, say, your non-culpable and unmotivated ignorance of economics, you are not blameworthy for your wrong action. Now, what if team de dicto is right? If they are right, the good will, the motive that gives right actions moral worth, is the will to do right. As long as you acted purely out of the will to do right, without self-deception etc., you are excused.

I will take it as my first premise that:

1. A motive that grants moral worth is not the motive of any blameworthy action (barring self-deception, mixed motives etc.)

Now, suppose that concern for morality de dicto grants moral worth. One prima facie difficulty with this view is that concern for morality de dicto is often the straightforward motive of *wrong* actions (without any self-deception etc.). That happens in cases where the agent does something wrong which she honestly believes to be right, and so her will to do right motivates her straightforwardly to do what she takes to be right – but is in fact wrong. If concern for morality de dicto is to be a moral worth granting motive, wrong actions honestly believed to be right are not blameworthy. This is the view known in the literature as "Moral Ignorance Excuses".

I say "honestly believed". By "honest" belief or "honest" ignorance I mean a belief or an ignorance that is not the result of prior culpable action, self deception, or motivated irrationality. I will refer to these three things as "shenanigans". I use the terms "honest false belief" and "honest ignorance" rather than talk about *rational* false belief or *rational* ignorance because there are technically irrational belief states that are not the result of shenanigans. These would be cases of epistemic irrationality that does not result from a desire, will, emotion

or something else motivational. The gambler's fallacy might be such a case: what we call an "honest mistake".[5]

So far, I have said the following:

1. A motive that grants moral worth is not the motive of any blameworthy action (barring self-deception, mixed motives etc.)
2. Concern for morality de dicto is the motive of wrong actions in cases of honest moral ignorance, or at least cases of *honestly but erroneously believing one's action to be right*.
3. Therefore, if concern for morality de dicto grants moral worth, an honest belief that one's action is right excuses from blame.

To get from here to "you can't have it both ways", more steps are necessary. My next step will need more defending. It is to introduce the following premise:

4. If believing an action right excuses from blame, believing an action wrong "excuses" from esteem.

Let me now defend premise 4.

Why the Symmetry?

If honest moral ignorance excuses, it excuses the same way factual ignorance excuses. They are, as Rosen (2003) says, on a par (otherwise, it is quite unclear what we mean when we say it excuses). To see how one kind of moral ignorance – believing your right action to be wrong – excuses from esteem, we need to remind ourselves of the way ignorance excuses in the first place.

Think of the following scenario:

A. Eve gives Sasha poison, honestly believing that she is giving him vitamins.

With some boring assumptions in place, this is a paradigmatic case of factual ignorance excusing from blame. Eve is not blameworthy for poisoning Sasha. Some would even be inclined to say she is worthy of esteem, but much goes wrong theoretically if we assume that there can be a morally worthy action that is not right, which is why philosophers writing on moral worth agree that a morally worthy action has to be right. There might be a right action that Eve does – she *attempts to give Sasha helpful vitamins*– but to for her to be unworthy of blame there need not be such an action, as is the case if she gives him the pill because she wants him to like her, but honestly assumes her action would be helpful, or at least harmless. Simply the fact that she honestly believes she is not harming Sasha exempts her from blame.

---

[5] I am echoing here the distinction made by David Pears (1999) between hot and cold irrationality.

Now think of the following scenario:

    B. Eve gives Sasha vitamins, honestly believing herself to be giving him poison.

In this case, factual ignorance makes it the case that Eve is not esteem-worthy for her action. Some would even be inclined to say she is blameworthy – but much goes wrong theoretically if we allow blameworthiness for right actions, and so almost all philosophers working on free will assume that for someone to be blameworthy, there needs to be a wrong action for her to be blameworthy for[6]. There might be such an action if Eve *attempts to poison* Sasha, but she is not esteem-worthy even if she is attempting no such thing – as would be the case if she gives him the pill because she enjoys watching him consume purple pills, but assumes that her action would be lethal. She is "excused" from esteem simply because she believes that with the pill, she is poisoning Sasha.

So blame for giving someone a harmful pill is eliminated by having honestly believed that the pill was harmless or helpful, and esteem for giving someone a helpful pill is eliminated by having honestly believed the pill was harmful or unhelpful. This is how factual ignorance works.

Now let us look at moral ignorance. Again, if moral ignorance excuses, it excuses in an analogous way to factual ignorance ("on a par" with it), or something very strange is going on. With that in mind, let us look at the following scenarios:

    C. Eve does wrong but honestly believes what she is doing is right.
    D. Eve does right but honestly believes that what she is doing is wrong.

If Moral Ignorance Excuses, Eve at scenario C is excused from blame - and at D, she is "excused" from esteem. For the excuse to apply, it need not be the case that Eve *aims* at doing wrong. Simply having believed that what she is doing is wrong it renders Eve unworthy of esteem for doing it (and the same temptation exists as in 2 to hold her *blameworthy* for her action). If believing right is excusing, believing wrong is, in a way, damning[7].

In arguing that blameworthiness and esteem-worthiness are symmetrical this way I do not intend to argue that that are symmetrical in any other way, nor do I need to do so.[8]

<u>The Argument (Part 2 plus a sequel).</u>

To recapture, here is what we have so far:

    1. A motive that grants moral worth is not the motive of any blameworthy action (barring self-deception, mixed motives etc.)

---

[6] For an exception, see Zimmerman (1997) (2010).
[7] See Zimmerman (1997). Again, I do not agree that "damning" goes all the way to making the agent blameworthy.
[8] For example, I do not think the asymmetries Darwall (2006) and Wolf (1980) argue for, if real, would matter here.

2. Concern for morality de dicto is the motive of wrong actions in cases of honest moral ignorance, at least cases of honestly believing one's action to be right.
3. Therefore, if concern for morality de dicto grants moral worth, an honest belief that one's action is right excuses from blame.
4. If believing an action right excuses from blame, believing an action wrong excuses from esteem.

From 3 and 4, we get: If concern for morality de dicto grants moral worth, believing an action wrong excuses from Esteem. Now:

5. If believing an action wrong excuses from esteem, that agents in Huck Finn cases are not esteem-worthy for their actions.

Huckleberry Finn does right while believing he's doing wrong, and so, if believing wrong "damns", he is "damned" – there is that temptation to call him blameworthy[9] – and he's certainly not worthy of esteem. It does not matter that he does not *aim* at doing wrong. Just like in scenario 2 it does not matter if Eve attempts to poison or just does not care if she's poisoning or not, in Huck's scenario it does not matter if he helps Jim because he thinks it is wrong or in spite of his thinking it wrong: he is not esteem-worthy either way.

If we add up steps 3 through 5, we get:

6. if concern for morality de dicto grants moral worth, agents in Huck Finn cases are *not* esteem-worthy for their actions.

   However,
7. If concern for morality de re grants moral worth, agents in Huck Finn cases *are* esteem worthy for their actions.

Therefore
8. It is false that: both concern for morality de dicto grants moral worth and concern for morality de re grants moral worth.

Another way to get to the conclusion in 8 involves showing that team de dicto and team de re reach conflicting conclusions about blameworthiness. Specifically, they reach conflicting conclusions about the counterpart kind of case to Huck Finn cases – cases of people who *do* follow their false moral beliefs. Imagine a counterpart to Huck Finn who does turn Jim in instead of helping him. More stereotypically, imagine an honest Nazi who kills Jews because he believes it is required, or steals money from a Polish prisoner because he takes it to be permissible. This argument (which I will dub "the sequel") looks like this:

---

[9] Zimmerman (1997) (2010) argues that Huck is blameworthy.

1. If concern for morality de dicto grants moral worth, an honest belief that one's action is right excuses from blame. (repeating step 3 above).
2. If concern for morality de re grants moral worth, the actions in Huck Finn cases have moral worth (the agents deserve esteem for them).
3. If the agents in Huck Finn cases deserve esteem for their actions, honestly thinking one's action wrong does not excuse from esteem.
4. If honestly thinking one's action is wrong does not excuse from esteem, thinking one's action right does not excuse from blame (due to the analogy with factual ignorance used in the previous section: if thinking you are giving poison "excuses" you from esteem for giving vitamins, thinking that you are giving vitamins excuses you from blame for poisoning).
5. But again: if concern for morality de dicto grants moral worth, thinking your action right *does* excuse.
6. In other words: team de dicto implies that the honest Nazi is excused, team de re implies that the honest Nazi is not excused.

Is there Honest Moral Ignorance?

There is one group of philosophers to whom it might seem that all of this discussion is neither here nor there, because there is no such thing as *honest* false moral belief, or, in general, honest moral ignorance. Moral ignorance, they say, always results from some shenanigans[10]. Perhaps the person who believes a false ethical view believes it out of self-deception, as a kind of rationalization for selfish actions she wants do anyway. Perhaps, alternately, her authentic hatred of some individual or group is so powerful that under its influence, her belief-forming apparatus produces irrational beliefs. Or perhaps her ignorance is culpable – she should have worried more about moral matters when she was younger instead of being preoccupied with chess, say, and failing to deliberate about the merits of the moral views her friends favored. One way or another, she did not come by her belief honestly.

Here I must admit to laziness, as I keep using the Nazis, a notorious bunch about whom I have read a great deal, as my go-to example of wrongdoers who think they are right-doers. It sounds like a reasonable thesis that the actual perpetrators of Nazi war crimes, to the extent they were motivated by normative beliefs, got to their beliefs because of shenanigans. These people often did not grow up Nazi. They grew up going to church every Sunday and studying Kant in school just like their friends who did not become Nazis. They became Nazis as teenagers or adults while being exposed to many political views and arguing about them. This is where historians say things like 'German youth were excited about the idea of building a better nation' or 'downtrodden German workers had to take their anger out somewhere'. In other words, there *were* shenanigans: people were motivated to be Nazis by anger or pride or fear, or some unholy mix of emotions. It all sounds reasonable enough.

---

[10] This has been argued famously by Harman (2011). For related discussion see, for example, Guerrero (2007), McGrath (2009), Miller (2021) Sliwa (2012). Wedgewood (2019) and Wieland (2015). For the view that there are even more honest or rational sinister beliefs than I think there are see Begby (2021) and Levi (2021).

However, there are other cases of moral ignorance that seem different – enough of them that the burden of proof rests on the shoulders of those who argue that there is no honest, shenanigans-free moral ignorance. Consider cases in which, unlike Nazism in Germany, a false view is so ubiquitous that a person is not only bound to be taught it as a child but can live a long life without encountering a single person opposing it. Why did Immanuel Kant believe that masturbation is a grave moral wrong? Even the staunchest objectors to Kant's moral theory do not take it to be a consequence thereof. It is easy to make fun of Kant here, and I've heard many people speculate about possible shenanigans that might have gone on in his mind: did Kant hate sensuality? Was he secretly gay and "defensive" about it? I think the most likely story available to us is that Kant believed masturbation was wrong because his mother believed it, his father believed it, his teachers believed it, everybody else in Europe believed it, and Kant never heard anybody cast that belief into doubt. It does not take shenanigans for an average person, or even a genius, never to reach the point of questioning such unquestioned views.

 Some might infer that I am denying that morality can be known a priori. However, I have no need to deny it at this point. Logic is known a priori and we make plenty of perfectly honest logical errors. Mathematics is known a priori and yet our intuitions are forever getting the Monty Hall problem wrong and the Gambler's Fallacy is committed by almost everyone who hasn't heard of it. I do not think every student who fails calculus is somehow motivated to get the wrong answers. She need not unconsciously want to fail nor need she have some other motivation to turn her back to evidence. Even if morality is known a priori, the burden of proof is on anyone who wants to say that there is no honest moral ignorance.

Partial Credit?

When I argue that one cannot have it both ways with regard to moral worth, some minds immediately turn to the possibility that concern for morality de dicto could grant *partial* moral worth even if the de re view is otherwise true. This might seem a strange thought to some other minds, because 'moral worth' is a term we owe Kant, and Kant did not think any agent could somehow receive "partial credit" in terms of moral worth. However, it is still possible to talk of moral worth without accepting Kant's particular views on matters such as mixed motives, nor his normative ethical theory. Recall that according to the de re view of moral worth, an action has moral worth if it is done for the reasons for which it is right, also known as its right-making features. Now imagine a pluralistic, perhaps Ross-like, view of morality is true. For example, suppose it turned out that morality contains a benevolence component, a respect-for-autonomy component, and a fairness component, ranked but not reducible to each other. On such a view there are many pro-tanto moral reasons and the right-making features of actions are complicated. The right action isn't simply "the one that maximizes utility" or "the one that's universalizable", but something resembling "the action that best protects wellbeing without conflicting with anyone's autonomy and without being unfair beyond a certain threshold". If any such view is true, it seems that "partial credit" in terms of moral worth *could* be possible.

Here's how: suppose one day a student of mine is very sad, and I cheer him up considerably by telling him, say, that somebody said she liked his work. What I do, let us suppose, is the right thing. My cheering up my student is my modest contribution to keeping his wellbeing from deteriorating, and that's part of what morality is about: protecting wellbeing. As for the other parts, I am not interfering with his autonomy and I am not being unfair to anyone. I do, on my invented theory, the right thing. But suppose it is true that I am not, in fact, properly moved by all three considerations. I convey gossip to my student with the intention of protecting his wellbeing, but I would have told him the cheering story *even if it were false*. That, let us assume, would be objectionably paternalistic and clash with his autonomy, and so I would have done something wrong. Luckily, the story is true. I have done the right thing. Am I esteem-worthy for my action? Does it have moral worth? To answer "not at all" seems harsh, as it puts me in the same category as the Prudent Grocer, who is acting in order to get money. My benevolence – concern with human wellbeing – is worth *something*. Despite the counterfactual scenario in which I tell a lie, it does not seem like a complete accident that I did something right, as rightness and benevolence are related in a way that rightness and financial profit are not. Hence the idea that when I act rightly on a *pro-tanto* moral reason – "he needs help" - this earns me some credit.

If benevolence can grant partial moral worth, why can't concern for morality de dicto do the same? I have said earlier, as premise 1 of my argument, that a moral worth-granting motive cannot, without shenanigans and such, be the motive of blameworthy action, but that would not be true for a motive that only grants *partial* moral worth. Benevolence can, after all, without shenanigans, be the motive of blameworthy action, as it would have been if I told my student a benevolent lie. Thus, the fact that concern for morality de dicto can sometimes lead to blameworthy actions should not rule out its granting *partial* moral worth.

Except that things don't work this way. It is still true that a motive – or a mix of motives - that grants *full* moral worth cannot, without shenanigans and such, be the motive of blameworthy action. If morality is, to be ridiculously simplistic, 40% about benevolence, 30% about autonomy, and 30% about fairness, and the right action is, let us say, the most benevolent action compatible with protecting agent autonomy without being unfair above a certain threshold, then the right *mix* of concerns, when it leads to a right action, grants it full moral worth, and when it leads to a wrong action, without shenanigans , the action is excused. In other words, the right mix of concerns cannot, without shenanigans, lead to a blameworthy action, even if each ingredient or pair of ingredients can.

So imagine an agent whose motivation to do the right things she does is 40% benevolence, 20% respect for autonomy, 20% fairness, and 20% concern for morality de dicto – a concern which, left alone, leads the agent to follow whatever she takes to be the voice of morality: the agent would probably call that fourth factor simply "morality" or "moral considerations". If these last "wild card" 20% are at all efficacious as part of a motive, then the mix in question can straightforwardly motivate *wrong* actions, as, depending on the agent's views, that important fourth factor can be, say, the extent to which her action promotes the purity of the white race, a "moral" consideration which she "needs" to "balance" with the other three. Are those wrong

actions blameworthy or not? If they are blameworthy, the 4-way mix cannot, in fact, grant full moral worth: it is, without shenanigans and such, the motive of blameworthy actions. Are those wrong actions then un-blameworthy, excused? If so, what excuses them? The only non-question-begging answer is that the agent honestly believes some false view of morality, and *that* excuses her. But if false moral beliefs excuse, Huck Finn cases are also excused from praise and involve no moral worth. Thus, even the view that concern for morality de dicto grants *partial* moral worth is incompatible with the view that Huck Finns deserve esteem for their actions, and thus defeats any ecumenical purpose. If concern for morality de dicto grants *even partial* moral worth, concern for morality de re does not grant *any*.

That makes "hybrid views" that take concern for morality de dicto and concern for morality de re to be different virtues, analogous to justice and benevolence, incompatible with Huck being esteem-worthy. Again, to simplify our example slightly, a view to the effect that a moral person is motivated by, say, 70% benevolence and 30% what she calls "morality" has a problem that does not exist for a view that takes it to be 70% benevolence and 30% justice. "Morality", in conjunction with false moral beliefs, can lead the agent astray, at which point the question of whether she's blameworthy for her wrong actions arises again. If she is blameworthy, the 70%/30% mix cannot grant full moral worth, seeing that it motivates blameworthy actions. If she isn't blameworthy, it seems that what excuses her is her mistaken view about morality (if she had better ones, she would have acted rightly), and if mistaken views about morality excuse, they also excuse from esteem and "excuse" Huck, and so we start again.

It Depends?

After all of this, one might still be tempted to suggest some disjunctive view according to which one gets some kind of esteem for doing right out of concern for morality de dicto and some other kind for doing right out of concern for morality de re.

This move, always tempting to a philosopher, would be costly here. First, note the connection, made explicit in my main argument and even more so in the *sequel* above, between the philosophical discussion of esteem-worthiness and the discussion of moral ignorance excusing or not excusing. If concern for morality de dicto grants moral worth, I have argued, moral ignorance, at least in the form of the false belief about the rightness or wrongness of one's action, excuses, and if concern for morality de re grants moral worth it does not, even in such cases, excuse. It would be non-trivial to defend the view that the debate about the honest or "sincere" Nazi – whether she is excused from blame or not – does not involve a real disagreement, but rather different kinds of blame.

Then we have the fact that there is a clear case in which concern for morality de dicto and concern for morality de re simply lead to mutually exclusive actions. Consider again our two schematic people, Diana de Dicto and Deirdre de Ray. Recall Diana de Dicto is motivated by concern for morality de dicto, and due to her moral beliefs being all true, she always acts well. Deirdre de Rey acts out of concern for morality de re, and she also always acts well, as she cares about all the relevant things in the right order. One day the two of them encounter a clever

teacher or evil neuroscientist who convinces them of a false moral view, or perhaps they encounter such a difficult moral problem that their merely human intelligence does not suffice for them to reach a right answer. Diana de Dicto immediately changes her behavior so as to do all and only things that she takes to be moral. Deirdre de Rey refuses to do what she newly believes morality requires. Would the good person react like Diana or like Deirdre? The relationship between virtue and moral worth is nontrivial, but surely a virtuous person, when she does right, does it nonaccidentally, from a motive that suffices to make her worthy of esteem for her action[11].

Neo-Aristotelians – just about any neo-Aristotelian, I suspect - would be inclined to interject here that there is no sense in asking what the virtuous response would be to acquiring a false moral belief, because a fully virtuous agent simply does not have false moral beliefs. She is too wise to acquire such beliefs, and if one were inflicted on her by an evil neurosurgeon she would have ceased to be fully virtuous.

I think this is wrong, and my thought is, in a way, rather Kantian, in that it latches again to the Kantian sense that a good will is jewel-like even when it does not achieve its goals. A person of perfect good will is a perfectly good person. To be a person of good will you need to have an adult human's intelligence, but you don't need to be smart. How smart you are is a separate question from how good you are, and nobody is ever a morally worse person simply due to a *cognitive shortcoming*. It is true that a cognitive shortcoming can make it harder for a person, on some occasions, to do the right thing, but the same is true of literal blindness and deafness, or even being physically weak or short of money, none of which make a person less good. We hope that a person who has good will would also be strong, because in that way she would be capable of saving more people from drowning. We want her to be rich so that she can give more to the poor, and want her to be smart so that, for example, she will know to vote for the right candidate in a complicated election or be an efficacious lawmaker herself. Still, being smart does not partially constitute being good any more than being strong or rich partially constitutes being good. While wisdom is not the same thing as smartness, it seems to require some smartness at its base, which is why I do not take wisdom to be required for virtue[12].

There is no space left in this paper to defend this view. However, as long as it makes sense to ask what the perfectly good person would do when convinced of a false moral view, Team de Dicto and Team de Re would be providing us with incompatible answers, implying that the disagreement between them cannot be trivially resolved by saying "it depends what you mean by esteem".

Aboodi, Ron (2015). The Wrong Time to Aim at What's Right: When is De Dicto Moral Motivation Less Virtuous? Proceedings of the Aristotelian Society 115 (3pt3):307-314.

---

[11] Such an intuition is, for example, behind Foot's declaration that if utilitarianism were true, benevolence would have been the only virtue. Why wouldn't any character trait that increases utility count as a virtue under utilitarianism? Foot (1976) assumes that a truly virtuous person does the right thing *nonaccidentally*.
[12] See Arpaly (2014).

Arpaly, Nomy (2014). Duty, Desire and the Good Person: Towards a Non-Aristotelian Account of Virtue. Philosophical Perspectives 28 (1):59-74.

Arpaly, Nomy (2002). Moral Worth. Journal of Philosophy 99 (5):223.

Arpaly, Nomy (2003). Unprincipled virtue: an inquiry into moral agency. New York: Oxford University Press.

Arpaly, Nomy & Schroeder, Timothy (2013). In Praise of Desire. New York: Oxford University Press. Edited by Timothy Schroeder.

Begby, Endre.2021. *Prejudice: A Study in Non-Ideal Epistemology*, Oxford University Press.

Carbonell, Vanessa (2013). De Dicto Desires and Morality as Fetish. Philosophical Studies 163 (2):459-477.

Copp, David (1997). Belief, Reason, and Motivation: Michael Smith's "The Moral Problem". Ethics 108 (1):33-54.

Cunningham, J. J. (2021). Moral Worth and Knowing How to Respond to Reasons. Philosophy and Phenomenological Research 105 (2):385-405.

Darwall, Stephen (2006). The Second-Person Standpoint. Harvard University Press.

Fearnley, Laura (2022). Moral Worth, Right Reasons and Counterfactual Motives. Philosophical Studies 179 (9):2869-2890.

Foot, Philippa. 1978. "Virtues and Vices". In Philippa Foot, *Virtues and Vices and Other Essays in Moral Philosophy*. Oxford University Press.1-18

Guerrero, Alexander A. (2007). "Don't Know, Don't Kill: Moral Ignorance, Culpability, and Caution". *Philosophical Studies* 136 (1):59-97

Harman, Elizabeth. 2011. "Does Moral Ignorance Exculpate?" *Ratio* 24:4 443-468

Herman, Barbara (1993) The Practice of Moral Judgement. Harvard University Press

Hurka, Thomas. (2014), Many Faces of Virtue. Philosophy and Phenomenological Research, 89: 496-503.

Isserow, Jessica (2020). Moral Worth: Having It Both Ways. Journal of Philosophy 117 (10):529-556.

Johnson King, Zoe (2018). Accidentally Doing the Right Thing. Philosophy and Phenomenological Research 100 (1):186-206.

Johnson King, Zoë (2019). Praiseworthy Motivations. Noûs 54 (2):408-430.

Kant, Immanuel (1998/1785). Groundwork of the Metaphysics of Morals. Cambridge University Press.

Lerner, Adam (2018). The Puzzle of Pure Moral Motivation. Oxford Studies in Metaethics 13:123-144.

Levi, Neil. 2021. *Bad Beliefs*. Oxford University Press.

Lillehammer, Hallvard (1997). Smith on Moral Fetishism. Analysis 57 (3):187–195.

Lockhart, Jennifer Ryan (2017). Moral Worth and Moral Hobbies. Ergo: An Open Access Journal of Philosophy 4.

Markovits, Julia (2010). Acting for the Right Reasons. Philosophical Review 119 (2):201-242.

Massoud, Amy (2016). Moral Worth and Supererogation. Ethics 126 (3):690-710.

McGrath, Sarah .2009. "The Puzzle of Pure Moral Deference". *Philosophical Perspectives* 23 (1):321-344.

Miller, Daniel J. (2021). "Can Morally Ignorant Agents care Enough?" *Philosophical Explorations* 24 (2):155-173.

Olson, Jonas (2002). Are Desires De Dicto Fetishistic? Inquiry: An Interdisciplinary Journal of Philosophy 45 (1):89 – 96

Pears, David. 1999. Motivated Irrationality. St. Augustin Press.

Portmore, Douglas W. (2022). Moral Worth and Our Ultimate Moral Concerns. Oxford Studies in Normative Ethics.

Rosen, Gideon. 2003. "Culpability and Ignorance". *Proceedings of the Aristotelian Society*

Singh, Keshav (2020). Moral Worth, Credit, and Non-Accidentality. In Mark Timmons (ed.), Oxford Studies in Normative Ethics, Vol. 10. Oxford University Press.

Sliwa, Paulina 2012. In Defense of Moral Testimony. *Philosophical Studies* 158 (2):175-195

Sliwa, Paulina (2015). Moral Worth and Moral Knowledge. Philosophy and Phenomenological Research 93 (2):393-418.

Smith, Michael (1994). The Moral Problem. Cambridge, Mass., USA: Blackwell.

Svavarsdóttir, Sigrún (1999). Moral Cognitivism and Motivation. Philosophical Review 108 (2):161-219.

Way, Jonathan (2017). Creditworthiness and Matching Principles. In Mark Timmons (ed.), Oxford Studies in Normative Ethics, Vol 7. Oxford University Press.

Weatherson, Brian, 2019. *Normative Externalism*. Oxford University Press.

Wedgwood, Ralph 2019. "Moral Disagreement and Inexcusable Irrationality". *American Philosophical Quarterly 56* (1):97.

Wolf, Susan (1980). Asymmetrical freedom. Journal of Philosophy 77 (March):151-66.

Wieland, Jan Willem 2017. What's Special about Moral Ignorance? *Ratio* 30 (2)

Zimmerman, Michael J. (1997). A Plea for Accuses. American Philosophical Quarterly 34 (2):229 - 243.

Zimmerman, Michael J. 2010. *Living with Uncertainty*. Cambridge University Press.