Moral Worth And Normative Ethics

Nomy Arpaly


According to a number of theorists (Arpaly 2002, 2003; Arpaly and Schroeder 2014; Markovits 2010), a morally right action has moral worth if and only if it is performed for the right reasons, which are the reasons for which it is right, or the right-making features of the action. I have referred to morally worthy actions as “praiseworthy actions”, though, as we will see, perhaps “esteem-worthy actions” would be more precise, if one were to use Kantian terminology.1

The idea central to these theories, of acting for right reasons through acting on the right-making features of actions, deserves particular attention. My purpose in this paper will not be to defend the view of moral worth in question, which I have already done in a paper and two books, but to argue that it has some implications for normative ethics, especially that it provides support for a pluralistic, rather than a Kantian or utilitarian, view. However, it would be useful to provide a few reminders of the nature of the view before exploring its implications, and I shall do so briefly in this section.

1 It should be added that what holds for right action also holds for supererogatory action, so really “morally desirable” might be more accurate than “right”. For ease of reference, though, I’ll talk about morally right actions.
Right reasons for action do not need to be reasons that the agent believes to be the right reasons for action. Huckleberry Finn’s action of helping Jim has moral worth if his reasons for helping Jim are the right-making features of helping Jim (that it treats Jim with respect, perhaps), however wrong he takes his action to be. Likewise, a person who performs the right action because she thinks it is right might not be acting for the right reasons at all. A person might think, for example, that the moral thing to do is to promote the interests of Aryan people over people of other races. If this person hires the better job candidate, not because she is better, but because she is of the “right” race, she acts rightly and acts because she wants to do the right thing – but still for the wrong reasons. She performs the right action, but not for its right-making features, and thus her action has no moral worth.

Another way to put this point, found in Michael Smith (1994), is to hold that acting out of a commitment to the right de dicto, as opposed to de re, is a kind of moral fetishism. On my view, being committed to the morally right whatever it is (when “maximize the beauty of ducks” is a possible candidate for the right moral theory in the agent’s mind) is not praiseworthy in the least. More precisely, a morally worthy action stems from a commitment to the right and the good correctly conceptualized. If utilitarianism has the right account of the features that

2 Arpaly (2003) defends this thesis at length.
make actions right then the agent performing a morally worthy action conceives of her action as maximizing utility, and is committed to maximizing utility so conceived; if Kantianism is correct she conceives of her action as respecting persons or as acting in a universalizable fashion, and so on. Whether she also conceives of her action as “the right action” is immaterial.

Very few people act on exactly the right reasons, however. Consider a person who does the right thing from a commitment to respecting persons but does not respect women, homosexuals, or blacks. This agent is not worse than most people in history, including Kant himself. If respect for persons is what acting rightly is about, it would be very harsh to hold that on those occasions in which Immanuel Kant worked hard to keep his promises his promise-keeping had no moral worth at all, and neither did the right actions of any other 18th century European male. Such a person, who is not committed to the complete moral reasons, might be committed to at least some partial, pro tanto moral reasons, and act upon them while doing the right thing – and that, as far as moral worth goes, should count for something. Kant thought of moral worth as an all-or-nothing affair, but my view is open to the possibility of degrees of moral worth (here Markovits differs from me – she denies that there are such degrees)³.

³ Following some of her arguments, I no longer accept my old (2003) view of degrees of moral worth. For the purpose of this paper I am not committed to what Timothy Schroeder and I say about degrees of moral worth, either (2014).
In Arpaly (2002 and 2003) and Arpaly & Schroeder (2014), I have also said that blame and vice involve a sort of “negative moral worth” – a failure to respond to the moral reasons one has for taking a certain course of action. Sometimes they also involve being motivated to perform an action by its “wrong-making features”.

Julia Markovits and I hold that there is an important connection between moral worth and moral reasons. If one grants that the morally worthy action is done for its right making features, then one can make inferences from claims about moral worth to claims about what the right making features of actions are. From the latter claims one can then make inferences about which normative ethical theories might be the correct ones.

Markovits (2010) has already applied the idea to utilitarianism in arguing that, if utilitarianism were correct then people who keep promises and are not motivated by considerations of utility would act with no moral worth. One way to put it is to say that she would act with no more moral worth than Kant’s prudent grocer. Kant’s prudent grocer, recall, acts honestly because he reasons that acting honestly will bring him money. He is a paradigm of a person whose right action has no moral worth as his motive is not based on the right-making features of his action. But despite what one would expect if utilitarianism were true, it is implausible to think that everyone who keeps promises for non-
utilitarian reasons is equivalent to the grocer in the moral worth of his or her actions. Yet, if utilitarianism were correct, it would follow that these people, just like the grocer, act for morally irrelevant reasons and not for the right-making features of their actions, and so act without moral worth. This, Markovits argues, is a reason to reject utilitarianism. It is of course open to the utilitarian to reject the intuition that people act with moral worth for non-utilitarian reasons, but that would be a costly choice. So would the choice to hold that any motive, regardless of its connection to utility, grants moral worth if it increases utility (i.e. if an invisible hand were to cause greed to result in prosperity for all, actions motivated by pure greed could be morally worthy). I cannot explore all of Markovits’ replies to objections here, but her line of argument strikes me as plausible. I do not wish to argue that in every case in which a normative claim (such as utilitarianism) clashes with an intuition about moral worth we should follow the latter, but sometimes the price of conflicting with an intuition about moral worth can be too high for a normative theory (and here it is worth noting that Kant starts the main text of the Groundwork with intuitions about moral worth - the good will - before getting to normative ethics - the categorical imperative).

In the next section I will discuss some general implications that facts about moral worth have on normative ethics if my view of moral worth is true. In the following section I will argue at some length for one such particular implication:
That there is more than one kind of moral reason, or, in other words, more than one right-making feature for actions. It is possible for two right actions to have different right-making features. This claim, I will argue, follows from facts about the moral worth of altruistic action – a topic on which a well-known view, usually identified with Kantianism, is wrong.

**General Implications**

What are the right reasons for a moral action? When asking this question, I will assume as a starting point that there are actions that have at least some moral worth. Kant doubts that any morally worthy action had ever been performed, but I see no need to share his brand of pessimism. An important thing to remember about agents who have acted in a morally worthy manner is that most of them are not ethicists. I refuse to refer to these agents as ordinary people, as some of them are quite extraordinary (see, for example, people who saved Jewish people from the Nazis). I will call them “philosophically unsophisticated”. It is unlikely that when acting well they follow, with any precision, the categorical imperative or the principle of utility (or Peter Railton’s specific consequentialism,4 or...). That alone is not enough to rule out either the categorical imperative or the principle of utility as moral truths. This is because it does not rule out the possibility that most praiseworthy moral actions are

---

4 See, e.g., Railton (1988).
committed for *pro tanto* moral reasons rather than for the complete moral
reasons, and the complete moral reasons *are* captured by the categorical
imperative or the principle of utility.

However, given my view of moral worth, the correct moral theory implies that
the reasons for which the philosophically unsophisticated morally worthy agents
act are either the complete moral reasons, or *pro tanto* moral reasons, or at the
very least be trivially derivable from such reasons. Commonsense morality is not
necessarily the right moral theory, but some commonsense moral reasons – the
ones acted upon by praiseworthy agents – are at least *pro tanto* moral reasons.

What moral reasons do philosophically unsophisticated praiseworthy agents,
whether ordinary agents or philosophically unsophisticated saints and heroes,
act from? I will not try to provide an exhaustive list of reasons, but it is easy to
think of some. “I promised” is one such reason: Quite often people explain why
they kept a promise simply by appeal to the fact that they have promised. The
fact that *someone is in need* is cited just as often as a simple reason for acts of
benevolence. In all likelihood these are reasons that genuinely motivate many
praiseworthy acts of promise-keeping and of benevolence, respectively, and thus
a true normative theory must account for their status either as complete moral
reasons or *pro tanto* moral reasons for action, or as at least derivable instances of
such reasons.
The Arpaly-Markovits thesis, again in conjunction with the assumption that there exist morally worthy agents who are philosophically unsophisticated, has another implication for the search for the right-making features of actions. Acting on the reasons behind uncomplicated right actions should not require an amount of reflection few are capable of or cognitive abilities that only a few possess. When I say “uncomplicated” moral action I am allowing for the fact that some moral actions, such as voting for the right economic legislation, do require reflection and intelligence due to their very nature. Some, however, do not. For lack of better words, it is an advantage for one’s theory of the right not to be too lofty.

One has to be careful not to underestimate the amount of complexity of which people are capable while at the same time being unable to articulate this complexity. People are capable of learning complicated dances without being able to provide purely verbal instructions as to how to perform them, and they are capable of learning complicated languages without being able to articulate their grammars (but this special ability to act for grammatical reasons, an ability learned in such a special way, is probably not representative of our abilities to act for complex reasons in general).
However, there is something suspicious about a moral theory that justifies uncomplicated right action in a manner that is foreign and hard to explain to anyone but the brightest university students. For example, explaining to all but these students what “contradiction in conception” means is remarkably hard, and so I hold in suspicion the view that people who are not the brightest students routinely act to avoid such a contradiction (pending a report from cognitive science that we have a special contradiction-in-conception detection ability – we who contradict ourselves so often!). One must note quickly that this disadvantage of the “contradiction in conception” view is far from a big problem for Kantianism in general. After all, even children ask “what would happen if everyone did that?” and “what would you say if someone did it to you?”, and these considerations are at least Kantian in spirit.

I do not wish to argue that one should expect all unsophisticated agents to fully understand one’s moral theory. For one, there is more to a moral theory than specifying the right-making features of actions. There is, for example, explicating conceptual connection between different features (e.g. a Kantian might ask what the relationship is between universalizability and respectfulness and a virtue ethicist might ask whether there is, in fact, a unity of the virtues). There is also, of course, the task of figuring out the implications of basic normative truths for the right thing to do in deeply controversial or complicated situations. However, it
speaks well for one’s theory if a university-level course is not required to understand one’s description of the reasons for which moral agents act.

**Altruism**

In this section I would like to defend the view that some actions have moral worth when they are motivated by non-instrumental concern for the wellbeing of others. I will also argue that, if we assume my view of moral worth, it follows that there are at least two kinds of moral reasons, altruistic reasons being one such kind. In other words, there are at least two different right-making features that different actions can have, one of which is the fact that an action protects wellbeing (or, if you wish, prevents or alleviates ill-being). The distinction between alleviation of ill-being and increasing wellbeing is blurry, but I will talk mostly about protecting wellbeing and alleviating or preventing ill-being since it is plausible that it is sometimes morally required to protect wellbeing, or prevent or alleviate ill-being (and, more controversially, that it is ever morally required to increase the wellbeing of people who are already doing well). There is no precise word in English for concern to protect wellbeing, though it is related to compassion, kindness and sympathy. As I don’t think it is precisely identical to any of these, I will refer to it, following Lawrence Blum (2009) as “altruism”.

For the purpose of this paper, when I am talking about altruism, I am talking about genuine non-instrumental concern for the wellbeing of one’s fellow humans, and also concern for a specific person that is a “realizer” of such a non-instrumental concern. Let me explain what I am not talking about, which will be equally important for what follows.

I am not talking about the emotion that Kant talks about when he says that “We love everything over which we have a decisive superiority, so we can toy with it, while it has a pleasant cheerfulness about it: little dogs, birds, grandchildren”.\(^5\) This statement, offensive to grandparents everywhere, is not about non-instrumental concern for another’s wellbeing, which has nothing to do with the joy of toying with the weak. I am also not talking about what Bennett talks about when he talks of “sympathy” as something that would motivate one to avoid taking a child to the doctor because one’s heart is melted by the child’s crying. This is not concern for the child’s wellbeing, unless, to invoke Mill for a moment, one assumes idiocy along with it. True altruism is also distinguished from what psychologists currently call “agreeableness”, which is sometimes all we refer to when we say casually that a person is “nice”. Agreeable people “get along” with people around them, but absence of aggression and aversion to causing conflict does not amount to caring non-instrumentally about anyone’s wellbeing.

---

As I am using altruism as a term of art, I would like to stipulate the following: when a person is motivated by altruism, her reason for action is “this person needs help”. Her motivation does not include the extra consideration sometimes known as the avoidance of a contradiction in one’s will. Christine Korsgaard (1996, 60) rephrases that consideration when she says that the morally worthy helper, according to Kant, has a motive to help that is deeper than that of the sympathetic helper because he has a “further stretch of motivating thought.” That further stretch “concerns the sort of world that this would be if no one helped, or better still if no one perceived the need for help as a reason to help (….). Such a world would be unacceptable because we regard our own needs as reasons why we should be helped.”

I would like to argue that the view summed up by Korsgaard (whether or not it is exactly Kant’s view) is false. One can be a morally worthy helper without having a “further stretch of motivating thought” of this sort.

First, on pain of loftiness, it is fairly safe to say that one can be a morally worthy helper without having exactly the thought Korsgaard describes. I doubt that many philosophically unsophisticated people wonder what the world would be like if no one “perceived the need for help as a reason for help”! However, there is nothing unusual about wondering what things would be like if everyone acted (or failed to act) in a certain way, and more specifically, there is nothing unusual
about a person wondering how everyone acting (or failing to act) a certain way would affect her. I am adding this specification because it is arguably possible to wonder what would happen if “everyone did that” or if “no one did that” in a way that is more consequentialist or rule-consequentialist than Kantian, which happens when you simply consider the effects of hypothetical universalized action on the world. However, it is also common to wonder how some universalized course of action would affect you. A simple example would be the statement “I must pay my taxes, because if no one paid taxes there would be no state, and without a state, where would I be? It’s only fair”. This is a Kantian “stretch of thought”. It is extra-Kantian if the implicit answer to the rhetorical question is something like “I would not be able to have an income” or “I would not have the comfortable life I want” (the latter desire being the basis of your wish to keep your money).

But consider a person – let’s call her Jennifer – who helps another individual – let’s call him John – who has serious difficulties writing his dissertation. John is not Jennifer’s student, nor is he a close friend, nor has he done her a favor in the past, nor is he likely to do her a favor in the future – there is no obvious issue of fairness in the background. It is simply that Jennifer notices John’s ill-being and is in a good position to help him. She helps him, perhaps at some cost to herself. Her reason? A human being needs help. It never occurs to her to ask herself how she would feel in a world in which nobody helped. No thought of a hypothetical
need for help, expecting to be helped, doing her fair share etc. plays a role in her motive as she helps John.

We might even imagine that she does not take her needs as reasons to help her. Perhaps Jennifer is really tough and values independence so deeply that she does not want to be helped – not in general and certainly not for free - or think she should be helped. We can also imagine Jennifer as a super-hero who is rationally confident that no difficulties she might encounter are such that she cannot surmount them without help. But the Jennifer I have in mind is a real human who simply is not motivated by a belief about her hypothetical needs as she helps John with his.

Jennifer’s motivation, it is worth remembering, is not hedonistic – she cares about other people’s wellbeing non-instrumentally rather than for the sake of some pleasure it might bring her. Neither is her motivation narcissistic. It is not the case that she helps John because of an inference along the lines of “do whatever Jennifer wants, Jennifer wants to help John, therefore I’ll help John”, but rather she helps John because John needs help. Some Kantians I have talked to seem to assume that a motivation other than the categorical imperative is always narcissistic, but people who act purely out of love for their children are surely not always motivated by the categorical imperative when they do, and are often far from narcissistic. If Robert cares about his daughter’s wellbeing and
helps her as a result, it need not be the case that Robert is motivated, narcissistically, by the thought “Robert cares about Martha’s wellbeing”. Robert might be a self-hater who does not care very much about himself or the fact that he cares about something. Luckily, self-help books are wrong when they tell you that you first have to love yourself in order to be able to love another. Robert the self-hater can be motivated by the thought “Martha needs help” – his not caring about himself and his own cares does not mean that he does not care about Martha. Jennifer cares about John’s wellbeing in a different way – she cares about it as an instance of human wellbeing – but the same reasoning applies to her.

To recapitulate: Jennifer helps John because, as stereotypical liberals like to say, she cares. She does not think, at any level, whether in lofty or pedestrian terms, about whether or not she would take her own needs as reasons for people to help her. Despite this fact, agents like her often receive praise, gratitude, esteem, and, in dramatic cases, even admiration for their actions. Some would have high regard for Jennifer especially because, when she helps, the thought of her own needs does not occupy her mind at all, not even hypothetically. Even if one thinks that a healthy or rational person has some concern for her needs and accepts help at least occasionally, one can still admire a helping action that is not motivated by any thought associated with one’s needs (in the case of the super-

---

6 See Pettit and Smith (1990)
hero it is hard to see why it would be irrational of her to reject help). One can admire such an action in a basically healthy and virtuous person who thinks of her needs when they arise, but not when helping others, and even in a person whom one would criticize as suffering from significant irrationality because she refuses help. The “tough” version of Jennifer, the one who values her independence above all may seem silly – like the stereotypical man who does not ask for directions – but that does not lower our regard for her helping John. Even if we imagine a Jennifer who is clinically depressed and who quite irrationally fails to take her needs as reasons for others to help her, we imagine a person who would normally be esteemed for helping John, as long as she does it because of genuine concern with wellbeing and not as a desperate gambit for attention, say.

It is possible for a person to have a disorder and also perform morally worthy actions directed at others (and Kantian intuitions tell us that there is something noble about a depressed person doing just that). Similar things are true for a Jennifer who, morally speaking, does not respect herself enough. It is possible for a person to have a vice (lack of self respect) and still perform morally worthy actions.

Robert Louis Stevenson said: “It is the history of our kindness that alone makes this world tolerable. If it were not for that – for the effect of kind words, kind looks, kind letters (….) I would be inclined to think our life a practical jest in the
Korsgaard’s morally worthy helper, the one who attends to another’s needs because she would want others to take her needs as reasons to help her, is arguably acting not out of kindness in the ordinary sense but out of a type of fair-mindedness. Now, not everyone values kindness quite as much as Stevenson does, but Stevenson’s sentiment is not unusual, and many have some version of it. It seems that, if one has Korsgaard’s view, one is committed to thinking that Jennifer, who acts out of a type of kindness that is not related to fairness, is someone whose actions have no more moral worth than those of the prudent grocer when he acts for profit. Does that mean that the advocate of this Kantian view must alienate herself from an intuition that, for many, runs deep?

Some contemporary Kantians – I am thinking of Wood (forthcoming) and Johnson (forthcoming), would say that’s not necessarily so. They would invoke Kant’s concept of moral *merit*, which they interpret as distinct from his concept of moral worth. Kant’s concept of merit is quite complex, but for the moment, it suffices to say that it is plausible that Jennifer’s action has merit. If one’s actions have merit, one deserves praise and encouragement for one’s action, even if one does not deserve esteem (the latter being reserved to morally worthy actors). Wood and Johnson point out that for Kant, actions performed out of pure self-interest do not deserve praise and encouragement, but actions performed out of compassion can deserve them under some conditions, as well as actions

---

7 This quote is used in a moving way by Kay Jamison in her autobiography *An Unquiet Mind*.
performed out of concern for honor. Thus, the Kantian does not need to lump Jennifer’s case together with the case of the prudent grocer. Can a Kantian who holds that even though Jennifer’s action has no moral worth, though she still deserves praise and encouragement, accommodate the common intuition that there is something valuable about Jennifer’s action and motivation when she helps?

To answer this, we need to ask the question: What is the difference between deserving esteem for one’s action, which seems to be equivalent to one’s action having moral worth, and simply deserving praise?

One clue is provided by the juxtaposition of praise and encouragement. These are both actions, whereas esteeming is not an action (the question “what are you doing?” can be answered with “praising and encouraging my child” but not with “esteeming Mozart”\(^8\)). Usually, encouragement is deserved the way a reward is deserved, whereas esteem is deserved the way an evaluation is deserved. The word ‘praise’ in English refers to something that can be deserved in both ways.

To see this, consider the following conversation:

Louis: I think he deserves a lot of praise.

---

\(^8\) I esteem James Dreier for coming up with this approximate test of action-ness.
Clara: Why?

Louis: Because he is a very good painter.

Clara: I don’t think he deserves to be praised as much as he is praised these days.

Louis: Why?

Clara: Because he never, ever, praises anyone else.

Louis and Clara are talking about different ways in one can deserve praise. Louis is talking about whether praise is warranted in the case of the man they are talking about, which is really asking whether admiration for him is justified. Whether admiration for a person’s ability as a painter is justified is a question to which it is simply irrelevant whether or not he praises other painters: Bringing this up would be appealing to the “wrong kind of reasons”. Clara, on the other hand, is talking about a practical question: whether one should perform an act consisting in giving the man praise. She wants to know if he deserves the benefit of hearing people praise him, in the same way in which one can ask whether someone deserves the money that people pay for his paintings. Whether or not the painter she is discussing praises other painters may or may not be relevant to the question, but it is not patently absurd to suspect that it is.

Talk of deserving praise can go both ways, then. Sometimes we say a person deserves praise and mean to say that admiration and esteem for her are warranted. At other times we say that a person deserves praise and mean that it
would be good to say a few good words about him, just like a person can
“deserve encouragement” in the sense that it is good to encourage him. To say
that Jennifer deserves “praise and encouragement” but not “esteem” is to say
that praising her is desirable, the way encouraging her is. This might be true, and
it may well be one difference between her case and that of the prudent grocer.
However, those who think Korsgaard is wrong about people like Jennifer do not
think that the disagreement is about how we should act towards Jennifer, but
rather exactly about how admirable or esteem-worthy she and her action are. We
might agree that it is desirable to praise the honor seeker, at least in a society in
which the right things are honored, perhaps because such honor-seeking, unlike
the grocer’s greed, is likely to bring forth good actions. Yet in the case of Jennifer
we think there is something that “shines like a jewel” about the agent herself and
the way she has acted – something beyond the desirability of praising the agent
(or the desirability of encouraging her or of any other behavior towards her).

To change the subject a little bit, can a broadly Kantian thinker be more
accommodating towards Jennifer if he rejects Kant’s view of the emotions?
Korsgaard points out what she sees as Kant’s idea of emotions and feelings as
“stupid”, which would be a great adjective to describe sympathy if it in fact
prevented otherwise smart people from taking children to the doctor (of if it lead
otherwise smart people to help an art thief who is struggling with a heavy load,
as Herman (1989) suggests). Korsgaard (1996) raises the possibility of a generally
Kantian view that does not include a Kantian view of emotion and feeling but is instead coupled with a more sophisticated, Aristotle-style view of these things. Indeed, there seems to be nothing to prevent us from being roughly Kantian and still holding a view according to which emotions and feelings can embody value judgments or respond to reasons. This could imply that some actions that seem guided by emotion and feeling, as opposed to some kind of explicit deliberation, have moral worth.

I agree that it is open to Korsgaard, and to many who hold any number of mainstream Kantian normative views, to hold that an agent who seems to act out of simple sympathy can be responding in a sophisticated way to reasons or values. However, the challenging thing about Jennifer’s case for such a philosopher is not that her motive might be “emotional” but that her motive seems to have nothing to do with universalizability: She does not have that additional “stretch of thought”. As I have described her, Jennifer’s motive does not even have to be “emotional” in the ordinary sense: Perhaps, when she helps John, she is too tired to have warm feelings of compassion towards him, but she tells herself coldly and explicitly that John needs help, and that cold and explicit thought motivates her – a thought that has nothing to do with how she would have felt about a world in which nobody considers helping her. This version of Jennifer’s case would still have to be regarded by Korsgaard as involving no moral worth.
I should mention that it seems perfectly possible to me for someone to act on a “visceral” motive and still be responding to reasons involving universalizability: Take, for example, a person who quickly declines a proposition involving tax evasion (or even tax avoidance) because he recoils, as viscerally as can be, at the thought of being a “free rider”. This person does not consciously deliberate about universalizability, and yet his feeling of revulsion at the thought of making an exception for himself could betray a sensitivity to universalizability-based moral reasons. By hypothesis, though, Jennifer is different.

**Misguided Altruistic Actions and Types of Reasons**

On my view, the esteem-worthiness of motives is not about reliability or “modal robustness”. The “non-accidentality”, as Markovits calls it, of the connection between the moral worth-granting motive and the right action done from it is in the fact that the right action is done from the reasons for which it is right. To be more precise, the non-accidental connection is not so much between motive and action but between motive and morality. If universalizability makes actions right and a person does the right action because it is universalizable, the connection between his motive and rightness goes beyond matters of probability.
However, a fact that needs to be explained by anyone who holds that altruism grants moral worth is the fact that altruism sometimes leads to a wrong action. I am not talking about cases in which a person helps another not knowing that the other is a thief, or fails to take a crying child to the doctor because of not knowing the elementary fact that it would be in his best interest to go there. I am referring to cases in which there is no complication in the form of lacking information or false factual beliefs, but the altruistic agent does the wrong thing anyway. If altruism is a moral motive, how can it lead, without factual ignorance, to even one wrong action? Also in need of explanation is the fact that even though people who are not philosophers know full well that there is such a thing as a wrong action motivated by altruism, they often resist the conclusion that right actions motivated by altruism have no moral worth.

Consider again Jennifer, who helps John with his dissertation in order to alleviate ill-being. Imagine that after getting to know Jennifer a little better you discover that she has at some point lied to her roommate. Jennifer, it turns out, faced the situation described by Thomas Hill in “Autonomy and Benevolent Lies.” Her roommate had barely recovered from an extremely painful love affair with a person with whom she was obviously incompatible. One night she asked Jennifer whether she thought her ex-boyfriend would be open to getting back together, and Jennifer said “no” despite knowing that the ex-boyfriend had, in

---

fact, wondered aloud about the rightness of the breakup and expressed a desire
to get back together. Suppose Jennifer said “no” to her roommate for the same
reason for which she helps John with his dissertation: Her roommate risks utter
misery and she wants to protect her wellbeing. Let us assume that Hill is correct
that this action is wrong.

Thus it appears that if my view of moral worth is true it should be expected that
if we were to learn about Jennifer and her roommate, we would reevaluate her
action vis à vis John and his dissertation and regard it as morally worthless. We
would feel the same kind of disappointment we feel when we discover that a
person’s good deeds were in fact motivated by money or by honor or by a need
to see himself as a good person. But while Jennifer’s lie to her roommate is a sign
of an imperfect character, many of us would keep having some esteem for her
and for her action of helping John. We can imagine John, upon discovering
Jennifer’s paternalistic lie, saying something like “Look, Jennifer, you are
basically a good person and I really appreciate what you have done for me. But
you’ve got to get it into your head that sometimes you have to let people make
their own decisions”. Our attitude to Jennifer is not pure disillusionment but
rather ambivalence.

To account for this complex attitude, I propose the following. There is more than
one type of moral reason – more than one right-making feature that different
actions can have – and it is possible to act in a way that reveals a commitment to only part of morality. One right-making feature some actions have has to do with wellbeing, but it is not the only one. Altruism fails to deliver right actions every time because considerations that have to do with wellbeing present a moral reason to help that is sometimes overridden by other moral reasons, reasons to do something else. Accordingly, Jennifer’s helping John might not have quite as much moral worth as such an action could have – perhaps the right-making features of helping John, specified more fully, include not just “alleviates a person’s ill-being” but also, say, “does not violate a person’s autonomy” or “is universalizable”, in which case Jennifer was only moved by a part of that which makes her action right. Her doing the right thing is not completely accidental – her reasons for action are some of the right reasons. It is, however, partially accidental – it is lucky that in the case of helping John no overriding reason exists to refrain from helping. However, even if this is true, I see no reason to insist, with Kant, that moral worth is an all-or-nothing affair. When a writer of short stories who is good at character development but not at creating page-turning plots writes a story in which the former is essential and absence of the latter does not really matter, she shows a real aesthetic sensibility even though there are writerly tasks in which she would fail. Moral sensibility can be similar.
More questions are raised by what I have said than I have space to answer, but I shall tackle two of them.

I have argued that wellbeing provides a moral reason for action that is not based on universalizability. A question that one might raise is the following: We have an idea of what an agent is like who is motivated to protect wellbeing but is missing at least one additional consideration, thus doing wrong. What would an agent be like who is motivated by universalizability, but who is missing at least one additional consideration, thus acting wrongly?

A full answer to the question requires another paper. It would require presenting a putative example of such an agent and defending the claim that he is in fact such an agent. Here, I will only provide the putative example and the basic argument. Strictly speaking, I will argue not that such an agent can exist, but that such an agent can exist if my view of moral worth is true. The relevant part of my theory is the part that states that a person acts viciously only if her course of action manifests some indifference to one or more moral reason.

Imagine the following two agents:

1). The Tough Guy: Akhil thinks of himself as a tough guy. He is a person who does not want anyone to help him, and will not accept free help under any circumstances. He values his self-reliance to the point that
offers of free help are perceived by him as a nuisance, and he is rather vain about the subject. One day Akhil sees Jardish, whom he knows from work, in the parking lot of a store, struggling with a heavy box that is he trying to carry the short distance to his car. Jardish has to take frequent breaks to catch his breath, but Akhil, of course, has spent a lot of time lifting weights. It would be remarkably easy for Akhil to help Jardish, but he does not offer help. He, Akhil, would never have wanted anyone to do him such a favor, even if the box was heavy enough to give him trouble. With a lack of imagination typical of human beings, Akhil thus assumes that surely Jardish feels the same. “Maybe not everyone is quite as tough as me”, Akhil would say if asked why he did not help, “but surely anyone, especially a man, would be insulted if offered help with that situation. How could I offend Jardish by implying that he can’t find a solution himself?” In truth, Jardish, who is different from Akhil in values and temperament, would have gratefully accepted the offer.

2) The Other Tough Guy: Kabeer also considers himself a “tough guy”, develops his physical strength and values self-reliance in the same extreme way that Akhil does. He sees Jardish struggling with the box. He knows Jardish does not have macho values or macho vanity and guesses correctly that he wants help. He does not help, even though it would have been easy for him. If asked for his motives, he would say, “Why should I
help? I don’t want help from anyone, so it’s only fair that I help no one”. He would say that honestly and sincerely. He always avoided, and will avoid, courses of actions that he would, in fact, regard as non-universalizable.

There is a difference between our two tough guys: Kabeer is a lot worse than Akhil. In contemporary parlance, Kabeer is behaving and thinking like an asshole, or at the very least a jerk, whereas Akhil displays no such tendencies, though he might show himself to be more naïve than is desirable. The difference between vicious behavior and well-intentioned behavior gone wrong due to naiveté is, on my view, the following: Naïve behavior goes wrong due to a factual mistake, whereas vicious behavior originates in a failure to respond to a moral reason. If morality were all about universalizability, and Kabeer’s only crime was misapplying the universalizability criterion, we would perceive him as regrettably naïve, bumbling, or simply as someone who made an error in calculation – we would see him, in other words, as we see Akhil. However, this is not how we see him: We see him as vicious. There is a moral reason to which he is failing to respond when he shrugs off Jardish’s need for help.

On to a different question. The Kantian view I have criticized in this paper is a view based on the idea of universalizability and thus rests on a common way to interpret the universal law formulation of the categorical imperative. There are,
however, other formulae. Can there be a plausible view that anchors both considerations of wellbeing and fairness-style considerations in the popular Kantian idea of refraining from treating one’s fellow human as mere means and always treating her as an end? Perhaps there is a duty, grounded in the formula of humanity, to take on other people’s ends as one’s own, and Jennifer’s action of helping John seems esteem-worthy because she is acting on this duty.

It is implausible, *prima facia*, that Jennifer acts from the formula of humanity. The duty to further the ends of other persons, at least as normally conceived by Kantians, is strictly incompatible with paternalism, which is an imposition on the other person of ends set for him by the agent. Though ordinary English makes it sound strange, Kantians regards paternalism of the sort expressed in benevolent lies as a paradigmatic way of treating persons merely as a means, or at the very least as a way of disrespecting the human ability to set ends as present in the victim of the paternalistic action. Jennifer, according to the most natural interpretation of her story, acts from the same motive when she helps John and when she lies to her roommate, and a motive that leads, without factual errors, to paternalism cannot be the formula of humanity. There is no reason to assume that Jennifer’s decision to lie to her roommate is based on a factual error, such as literally not knowing that the roommate is neither psychotic nor a child, but rather a competent, rational-enough human capable of setting her own ends, if not always very wisely. It is extremely unlikely that every case of paternalism
involves such a factual error (though unfortunately such errors do happen, for example when the victim of paternalism is very old or very unusual). Instead, it seems that often the paternalistic agent sees the victim’s ability to make autonomous decisions, but disrespects that ability, or, in more old-fashioned terms, sees the humanity in the victim but treats it merely as a means – and it is hard to use something as a means when you do not know it is there. If we assume that Jennifer’s case is one of these common cases, we have to conclude that she is not acting on the formula of humanity when she lies to her roommate. By hypothesis, the motive for Jennifer’s action is the same when she lies to her roommate and when she helps John. An advocate of the formula of humanity may, of course, deny that is possible and insist that Jennifer helps John from a radically different motive from the one that leads her to lie to her roommate. She can hold that all virtuous help is motivated by one thing – the formula of humanity – whereas all misguided paternalistic help is motivated by another – perhaps simply concern for wellbeing. That would be costly, as rather often the agent does not experience things that way, nor do the people around him - who think, for example, that the same good intentions that he always has were sometimes appropriate and sometimes misguided.

In addition, whether one perceives the different formulae of the categorical imperative as versions of the same directive or as related in a more complex way, it is an awkward position for a Kantian view to hold that it can be said, of the
same agent, that she follows one formula but not the other. Jennifer, by hypothesis, is not motivated by the thought of universalizability, and so, if the moral law is about univeralizability, it seems that she is not motivated by the moral law. If we say that she is motivated by the formula of humanity, we imply that she is, in fact, morally motivated. Usually, one would think that for a Kantian theory, either one is morally motivated or one is not – and the “either-or” is an exclusive one.

The question under discussion here is not whether the correct normative theory can be rightly called “Kantian”. This is partially a question of Kant interpretation and I will leave that to Kant scholars. The question is whether the correct normative theory involves one right-making feature or more than one. Hopefully I have shown reasons to suspect that there is more than one right-making feature that actions can have. It would take more to vindicate the suspicion, which I hope to attempt on another day.

Works Cited


Johnson, Robert. 1996. “Kant’s Conception of Merit” *Pacific Philosophical Quarterly* 77: 310-34


