

Discussion of Nomy Arpaly's *Unprincipled Virtue* for Philosophical Studies

Elizabeth Harman

Published online: 10 May 2007
© Springer Science+Business Media B.V. 2007

Nomy Arpaly's *Unprincipled Virtue* is a brilliant work of philosophy. Arpaly argues effectively against many central tenets of moral philosophy and action theory, and advances a new view of her own. The book is a joy to read—clearly written, clever throughout, and compelling in its descriptions and analyses of some imperfect agents who have been widely misunderstood.

Arpaly argues for a view about moral worth, or praiseworthiness. Her central claim is that praiseworthiness is a matter of acting in response to the moral reasons: whatever features in fact make an action right, it is acting in response to those features that makes an agent praiseworthy. For example, consider Huckleberry Finn who believes that turning in the escaped slave Jim to the authorities is the right thing to do. Huck does not turn Jim in; he is surprised at himself and does not know why he acts this way. Arpaly argues that there is a version of this case (indeed, she claims it is the version Twain intended) in which Huck acts in response to the fact that Jim is a person, just like Huck, who deserves to be saved from slavery. Because this fact is what makes Huck's action right, Huck is praiseworthy—even though Huck acts against his best judgment, thinks what he is doing is wrong, wishes he would act differently, and does not articulate to himself the reasons on which he acts. As Arpaly understands the case, it undermines the following commonly held claims: (i) an agent acts on reasons only if she articulates those reasons to herself and has them in mind as she acts, (ii) an agent is morally responsible only if she is agent-autonomous; for example, an agent who is moved to act by motives she wishes did not move her is not morally responsible, (iii) an agent acts irrationally if she acts against her best judgment. While Arpaly has much more to say against (i), (ii), and (iii), the case of Huck does nicely illustrate that they all fail.

Of course, there is much of the book I will not be able to discuss. To give some of its overall flavor, I will list three of my favorite lines.

E. Harman (✉)
Department of Philosophy, New York University, 503 Silver Center, 100 Washington Square
East, New York, NY 10003, USA
e-mail: elizabeth.harman@nyu.edu

“We want the moral person to have self-control, but only for the same reason we want the moral person to have money—namely, because self-control helps the moral person attain her goals (just as it helps the immoral person attain hers)” (100).

With this one line, Arpaly elegantly makes the case (also made by Kant) against the common belief that self-control is in itself a good thing. Surely she is right: the self-control of the assassin lacks any worth at all—just like the courage of the 9/11 hijackers.

“In a world in which people are rarely allowed to act on all their hearts’ desires, it is little wonder that one’s actions do not always reflect one’s concerns very well” (98).

This line culminates a discussion in which Arpaly argues that views which assign importance to *character* have been too quickly dismissed by taking people’s characters to be read easily off their actions. Arpaly argues that there is a notion of character as *what an agent deeply cares about*, and that facts about this type of character are more subtle than opponents to talk of character acknowledge.

“Stress may cause people to act out of character, but it may also truly *change their characters*” (166, Arpaly’s emphasis).

Here Arpaly explains why she thinks that cases in which agents act after “brain-washing” or science fiction neurological manipulation are not all on a par. Some such agents are not responsible for the actions these transformations induce in them: they become good people with compulsions to do bad things. But sometimes good people are turned into bad people, Arpaly claims, using Patty Hearst’s conversion by her kidnappers (and subsequent conviction) as an example. Arpaly thinks we are right to hold people responsible when they act after actually being transformed in this way.

I will make two points about Arpaly’s book. First I will raise a worry about her view of moral worth. Then I will argue that, contrary to Arpaly’s own assertion, her view provides some useful advice about how to live.

1. Arpaly’s central view is that a right action has moral worth to the extent that it was done in response to the features of the action which in fact make it right.

Arpaly’s view raises two questions: (1) how extensive are the features of a right action that make it right? and (2) in order to be praiseworthy for doing a right action, does an agent need to be acting in response to *all* the features of the action that make it right?

Consider the following case: Olivia has a cookie that she doesn’t want; her options are to throw it away or give it to Annabelle; seeing that Annabelle wants the cookie and that it will make Annabelle happy, Olivia gives the cookie to Annabelle. Suppose there are no further relevant details. In this case, it is clear that Olivia does the right thing, and she is praiseworthy.

A central part of the explanation of what makes this a right action is that it makes Annabelle happy. Furthermore, Olivia acts because acting this way will make Annabelle happy. So Olivia is acting in response to a feature of her action that plays a central role in making the action right. Is that enough to make her praiseworthy?

Note that Olivia’s action is not right *merely* because it makes Annabelle happy. Rather, it seems that Olivia’s action is right because it makes Annabelle happy, it

has no cost for Olivia, it has no cost for anyone else, Olivia has not made any promises that she will be unable to keep if she performs the action, and etc. The full explanation, it seems, will be long and elaborate: there are plenty of ways an action can be wrong although it makes Annabelle happy. Olivia's action is right because it makes Annabelle happy and also because it has other features: not being costly to Olivia, not being costly to anyone else, etc.

Does Olivia act in response to *all the features* of her action which make it right? It seems not. She may well not know exactly which features of her situation ensure that this is a case in which making Annabelle happy is the right thing to do, and so it seems she cannot be acting because *exactly those features* are present. Of course, as Arpaly points out, one can act in response to the right-making features of an action without thinking of them as the right-making features and without consciously thinking about them at all. Nevertheless, it seems that the right-making features of Olivia's action are so extensive and subtle that it is extremely unlikely Olivia is acting in response to *exactly those features*. It is simply psychologically implausible that Olivia is acting in response to all the features of her action that make it right.

Therefore, it seems that Arpaly should take this view: acting in response to a feature of one's action that plays a central role in making the action right is sufficient to be praiseworthy. That Olivia's action makes Annabelle happy does play a central role in making the action right; Olivia does act for this reason; so she is praiseworthy.

But we might raise an objection to this view. Consider another case, in which Trudy has a cookie she doesn't want; her options are to throw it away or give it to Annabelle; seeing that Annabelle wants the cookie and that it will make Annabelle happy, Trudy gives the cookie to Annabelle. While Trudy is moved to make Annabelle happy, she is *obsessively* moved in this direction. Annabelle's well-being motivates Trudy, but other people's interests do not move her at all. Trudy gives Annabelle the cookie because it will make Annabelle happy; and she would do whatever would make Annabelle happy, even if doing so would involve hurting other people. Just like Olivia, Trudy does the right thing. But we are inclined to say that Trudy is not praiseworthy for giving Annabelle the cookie, precisely because she acts with utter disregard for whether there is any cost to making Annabelle happy. Trudy does act because her action will make Annabelle happy; so she does act because of a feature of her action which plays a central role in making it right. On the view we offered to Arpaly in the last paragraph, Trudy comes out praiseworthy. So Arpaly should not adopt that view.

It seems that in this case, we think Trudy should be acting with regard to *all* the features of her action that make it right. The mere fact that Trudy is moved by a feature that is central to the rightness of her action does not show that Trudy is at all moved by the action's *rightness* if she is blind to the further features that make it right: for example, there being no one who will be hurt by the action. This understanding of the case seems to reinforce Arpaly's view: an agent is praiseworthy only if she's responsive to the rightness of her action. However, we are forced to the version of Arpaly's view on which agents are praiseworthy only if they act because of *all* the features of an action that make it right; and it's psychologically implausible that agents do tend to be motivated in that way.

I have argued that Arpaly's view faces a dilemma. Either (I) an agent is praiseworthy for performing a right action if and only if she acts because of a feature of that action which plays a *central role* in explaining its rightness, or (II) an agent is praiseworthy for performing an action if and only if she acts because of *all the*

features of the action which explain its rightness. (I) gets the case of Olivia right, but gets the case of Trudy wrong—so (I) must be false. (II) must be false because it is implausible that agents ever do act because of *all the features* of their actions that explain their rightness; so (II) implies that agents are never praiseworthy.

Arpaly might respond by claiming that (I) is true, but agents are *more praiseworthy* the more of the real explanation of the rightness of their actions is moving them to act. Indeed, it is part of Arpaly's view that praiseworthiness comes in degrees: one is more praiseworthy the more one is responsive to the features of one's action that make it right. Some of Arpaly's comments suggest that one way to be more responsive is to act on more of the features that make the action right. Plausibly, Olivia responds to more of the features that make her action right than Trudy does; so Olivia is more praiseworthy than Trudy. On this view, Trudy is praiseworthy, though less so than she might be. But surely Trudy is not praiseworthy at all. So this view must be wrong.

Alternatively, Arpaly might embrace (II) but deny my claim that the *full explanation* of the rightness of Olivia's and Trudy's action includes facts such as that no one is hurt by the action. Rather, the full explanation of the rightness of the action is that it makes Annabelle happy. It's true that if other factors were present, these factors would *defeat* the reason for the action given by its making Annabelle happy, but that doesn't mean *the absence of those factors* is part of the explanation of the action's rightness. On this view, saying that the full explanation of the rightness of Olivia's action is that it has certain features does not imply that any other action with those features is also right. This is a possible view about full explanations. Relying on this view, Arpaly can claim that Olivia does act in response to the full explanation of the rightness of her action. So (II) does not turn out to yield the conclusion that agents are never praiseworthy. However, Trudy's case now raises a problem for (II). Trudy is also responding to the full explanation of the rightness of her action: she is also responding to the fact that her action makes Annabelle happy. So on this view, Trudy is praiseworthy. Again, this view must be wrong.

Arpaly might offer a very different response, rejecting the dilemma and saying that she needn't embrace (I) or (II). Rather, she can take a middle view: (III) an agent is praiseworthy for performing a right action if and only if she acts because of *all* those features of the action that *centrally explain* its rightness. On this view, to be praiseworthy an agent must act in response to enough of the features of her action that she is responsive to the *whole central explanation* of its rightness, though she need not be responsive to the whole explanation. Regarding Olivia and Trudy, Arpaly could claim that the central explanation of the rightness of giving Annabelle the cookie includes both that it makes Annabelle happy and that it hurts no one else. It is plausible that Olivia acts in response to both these features; Trudy does not; therefore, Olivia is praiseworthy while Trudy is not.

This response to the dilemma has some intuitive appeal, but it raises two problems. The first problem would be in spelling out what makes some features part of the central explanation of an action's being right; I am not sure how this would be done. The second problem is a problem for Arpaly given her broader goals. She does not just want to be able to handle simple cases like Olivia's straightforwardly praiseworthy action, done for the right reasons with no inner conflict. Arpaly also wants to accommodate the praiseworthiness of Huckleberry Finn, who saves Jim against his better judgment. Huck is reacting to Jim's humanity and is for that reason praiseworthy. But is Huck reacting to all of the *central explanation* of the rightness of

his action? It seems to me that Huck needn't be reacting to all of the central explanation to be praiseworthy—and I think Arpaly would agree.

Finally, Arpaly might respond to the dilemma by embracing (II), that an agent is praiseworthy if and only if she acts because of all the features of her action which explain its rightness, and saying that the whole explanation of the rightness of Olivia's and Trudy's actions is that (a) the action makes Annabelle happy, and (b) there is no reason not to perform the action. Arpaly might claim that it is plausible that Olivia actually acts in response to (b), though she needn't articulate (b) to herself. It is more plausible that Olivia acts in response to *the fact that there are no reasons against performing her action* than that she acts in response to *all the specific facts that make this general claim true*: all the absences of factors that would mitigate against performing the action. Arpaly can claim that Olivia acts in response to both (a) and (b), so Olivia is praiseworthy. By contrast, Trudy acts in response to (a) but not to (b). Trudy does not care whether (b) is true and does not act in response to its truth.

I think this final response is the most attractive of the responses I have offered to Arpaly. However, this response would require Arpaly to modify her view, and it would thereby lead her into a problem we saw with the last response. Arpaly would have to say that agents are praiseworthy only if they respond to the *whole explanation* of the rightness of their actions. This enables her to get the cases of Olivia and Trudy right. But Arpaly wants to find Huck praiseworthy too. It's implausible that Huck is reacting to the *whole explanation* of the rightness of his act. In Huck's case, there are some reasons in favor of turning Jim in: for example, it is illegal not to turn Jim in. The whole explanation of the rightness of Huck's failing to turn Jim in involves factors on both sides, along with the fact that the harm to Jim of being a slave outweighs the reasons in favor of turning him in. It's implausible that Huck is acting in response to all of that explanation—and yet, it seems Arpaly would still find him praiseworthy.

2. Early in the book, Arpaly makes a helpful distinction between a theory that tells us which actions have moral worth and a theory that tells us *how to act* so that our actions have moral worth: the latter theory is an instructional manual; but Arpaly is offering the former theory. By making this distinction, Arpaly seeks to make it less counterintuitive that she thinks agents can often be praiseworthy although they act against their best judgment. She says that a correct instruction manual could never tell one to act against one's best judgment, but that doesn't mean that only actions done in accord with the agent's best judgment have moral worth.

This seems right to me. Furthermore, the cases she gives of actions that have moral worth although they are done against the agents' best judgments are compelling. A man embroiled with the mob finds himself calling the police despite having decided not to do so; unbeknownst to himself, he has slowly been noticing how wrong his complicity with the mob is, and the motivation to break away has been building inside him in light of this perception. Surely he is praiseworthy, although he never makes an explicit decision to break away, he is surprised by his own action, and he thinks he's making the wrong choice. In another example, a woman quits her chemistry PhD program because she has slowly been noticing the ways chemistry does not suit her. She has not articulated these ways to herself, and she thinks she is probably making a mistake, being weak-willed, in quitting the program. But she is right that chemistry is wrong for her, and she realizes later that she acted

out of an unarticulated perception of this fact. Surely this woman is rational in leaving—more rational than she would be in staying, at least—although she acts against her best judgment at the time. In both cases, the agent’s best judgments are formed irrationally, without taking into account some facts they have noticed without consciously considering them; their instincts do a better job of responding to everything the agents know and want.

My quibble with Arpaly is a small one. It is with her claim that she is not trying to give an instruction manual, and the corresponding implicit suggestion that her view does not provide us with any useful instructions.

Arpaly says that a correct instruction manual would always tell us to act in accord with our best judgments. Some of her cases might seem to suggest this is wrong. I might be trying to decide which of two job offers to take. When I articulate the reasons for and against each job, they seem to clearly favor job A over job B. Nevertheless, I have a strong “gut feeling” in favor of job B. It seems that I could think as follows: I may be like the girl in the chemistry program, aware of reasons I haven’t articulated to myself; taking job B may in fact be the right thing, though my best judgment favors job A. It may seem that an instruction manual that said you may go with your gut instinct over your best judgment would be correct.

However, it seems that the instruction manual would be allowing me to *conclude* that my gut is right. In that case, I would be revising my best judgment: although the reasons I can articulate favor job A, my best judgment favors job B *because my gut leans in that direction*. So, this case does not pose a counterexample to Arpaly’s claim that a correct instruction manual will always tell us to act in accord with our best judgment.

Nevertheless, there are some closely related claims that Arpaly’s view does call into question. Besides telling us *to act on our best judgment*, it seems that good instruction manuals would also tell us *not to act without reasons*. As Arpaly points out, it’s very common to conflate acting for reasons and acting with an articulated reason in mind. So the intuitive plausibility that the correct instructions tell us *not to act without reasons* leads to the mistaken claim that they tell us *not to act without an articulated reason in mind*. Indeed, people commonly make this mistake. If one notices oneself about to act with no clear reason in mind, or no clear *adequate* reason in mind, one may infer that one is acting irrationally. On the common views Arpaly criticizes, acting without a reason in mind is simply acting without a reason, and so must be irrational. Arpaly points out that this is a mistake. So, her view suggests the following piece of advice:

When you notice that you’re about to act without an articulated reason in mind, don’t thereby conclude that you’re about to act irrationally; allow at least for the possibility that you’re acting on good reasons you haven’t articulated to yourself.

It also suggests this advice:

When you feel inclined to act against your “considered judgment”—the judgment you make on the basis of the reasons you can articulate—don’t automatically conclude that acting on this inclination would be irrational; allow at least for the possibility that you’re acting on good reasons you haven’t articulated to yourself.

This seems like good news, or at least important and surprising news, particularly to the more neurotic and over-analytical among us. While we may caution ourselves that we tend to “overthink” things and thereby get into “muddles,” Arpaly lets us

know that all hope is not lost in these cases. We may manage to act rationally even when we cannot reason our way to a conclusion about how to act, and even when we have reasoned to the wrong conclusion. Knowing our tendencies to overthink, we needn't just try to *think harder*; we should seriously consider giving our instincts more of a chance, though their source may be opaque to us.

Arpaly would caution us in several ways. First, she would point out that gut instincts can be off-track as well as on-track; blindly following them is not advisable, and trusting them will be hit and miss. Second, she would point out that there is much to recommend articulating to oneself the reasons one has and the reasons one is acting on. The more complicated and removed an issue is, such as the question of how to vote, the more useful conscious deliberation may be. Finally, she might note that if we are aware of particular powerful motives within us that we do not endorse, then instincts that fall in line with those motives should be subject to strict scrutiny; for example, someone who knows herself to be unhealthily power-hungry should be suspicious of her inclination to accept a burdensome position as department chair. (But Arpaly's book also teaches us to be suspicious of these judgments about our motives. One person may be unhealthily power-hungry; another may have too little self-esteem and see her desire to have an important position as unhealthy power-hunger, when it is really a reasonable desire.)

I have made two points about Arpaly's book. First, I have argued that despite the intuitive attractiveness of the idea that agents are praiseworthy just in case they respond to the features of their situation that in fact make their actions right—that Huck, for example, is praiseworthy—there is a puzzle about *how many* of the right-making features of their actions agents must be responding to. If responding to any right-making feature will do, then the view gets Trudy's case wrong, saying she is praiseworthy. If responding to all right-making features is required, then the view implies no one is ever praiseworthy, because it's implausible that agents ever respond to the full explanation of their actions' rightness. Second, I have argued that Arpaly's view provides us with some useful advice, despite her protestations to the contrary. We should take our unexplained gut instincts more seriously than we do; we should not dismiss them out of hand.