


The Out of Character Objection to the Character Condition on Moral Responsibility

Robert J. Hartman 
Ohio Northern University

Benjamin Matheson 
University of Valencia

Abstract: According to the character condition, a person is morally responsible for an action *A* only if a character trait of hers non-accidentally motivates her performing *A*. But that condition is untenable according to the out of character objection because people can be morally responsible for acting out of character. We reassess this common objection. Of the seven accounts of acting out of character that we outline, only one is even a prima facie counterexample to the character condition. And it is not obvious that people act out of character in that sense. We argue that whether the out of character objection succeeds ultimately depends on the unnoticed methodological commitment that cases that may not resemble human life provide good data for theorizing about moral responsibility. But even if such cases provide good data, the forcefulness of the objection is at least deflated given that its persuasive power is supposed to come from clear real-life cases.

Key words: moral responsibility, character, acting out of character, reasons, whims

According to the character condition, a person is morally responsible for an action *A* only if a character trait of hers non-accidentally motivates her performing *A* (cf. Hume 1978: 411, 575). An often-cited challenge to this necessary condition is that an agent can be morally responsible for acting out of character (see, e.g., Arpaly and Schroeder 1999: 182; Doris 2006: 475; Kauppinen 2016: 54; Sher 2002: 385; Sripada 2016: 1227–1228). Indeed, many philosophers find it obvious that the character condition is a non-starter:

A characterological conception [of moral responsibility] promises to be underinclusive about whom to blame or punish. . . . For instance, it will exculpate Richard Herrin for his deliberate murder of Bonnie Garland, on the ground that this was out of character, despite the fact that there was no evidence that he suffered from significant cognitive or volitional impairment. (Brink 2013: 138)

We do not always withhold praise and blame when people *appear* to act out of character. We probably all have stories like this one: a person I would describe as wonderfully conscientious misses an appointment without a very good excuse. She is responsible for missing the appointment, even blameworthy, despite the fact that in some sense she acted “out of character.” Now we can always re-evaluate her character on the basis of this new information, but reflection on the case suggests that acting from traditional character traits is not a requirement for being responsible. (Nelkin 2005: 191; italics added)

Robert J. Hartman, Department of Philosophy, Ohio Northern University, Ada, OH, USA; <https://orcid.org/0000-0002-5115-7636>; roberthartman122@gmail.com. Benjamin Matheson, María Zambrano Fellow, University of Valencia, Spain; <https://orcid.org/0000-0001-5047-5803>; matheson.philosophy@gmail.com.

But a case in which a person is morally responsible for an action that is contrary to our expectations does not itself generate a counterexample to the character condition. A counterexample requires acting out of character in some deeper sense. Although the out of character objection is often mentioned, few elaborate on what they mean by acting out of character, and those who do say what they mean are not always in agreement. A compelling counterexample requires clarity and agreement about the nature of acting out of character. The widespread failure to meet this requirement provides a reason to re-examine the received wisdom that the character condition is untenable for this reason.

We clarify the objection by considering seven conceptions of acting out of character. In §1, we argue that six of those conceptions do not themselves provide even a *prima facie* counterexample to the character condition. Either the person in the alleged counterexample isn't morally responsible for the action, or the person doesn't act out of character in a sense that itself suffices to refute the character condition. In the latter case, the view of character is in some way problematic.

As we point out those problematic views of character, we outline a suitable account of character for the character condition: a character trait is a mental disposition to notice, think, feel, and be motivated to act in trait-relevant ways in a broad enough range of trait-relevant circumstances, and the mental disposition is stable and evaluable in the sense that it contributes to its possessor being a morally good or bad person (cf. Arpaly and Schroeder 1999: 171–173; Sher 2006: 19–22). This view of character accounts for the general motivation of the character condition: we hold persons, not actions, morally responsible; so, there must be a stable feature of the person that is presently evaluable after the action recedes into the past (Hume 1978: 411; Sher 2006: 12). Our independently motivated account of character fits that role precisely because character traits are morally assessable and stable (cf. Miller 2014: 3–36).

Nevertheless, the seventh conception, which we consider in §2, is a *prima facie* counterexample to the character condition. While we are in full agreement that actual human persons sometimes do act out of character in some sense, we contend that it is not obvious that actual human persons do act out of character in this seventh sense. As a result, there is no obvious counterexample to the character condition from out of character actions unless it is legitimate to point to merely possible cases of acting out of character as good data for theorizing about the nature of moral responsibility. Even if such cases are good data, the forcefulness of the out of character objection is significantly reduced given that its persuasive power is supposed to come from obvious cases in real life.

To be clear, we do not argue that an action's manifesting a character trait is necessary for being morally responsible for the action. We argue merely that the out of character objection does not, as it stands, refute that necessary condition.

1. SIX CONCEPTIONS OF ACTING OUT OF CHARACTER

1.1. The Nonvoluntary Conception

Suppose we understand “acting out of character” to mean that a person acts against her character in a way that is ignorant or coerced (Elliot 2017). For example, Oedipus flees the people that he falsely believes to be his parents, because he aims to avoid fulfilling the prophecy that he is to marry his mother. His character makes it unthinkable that he intentionally marries his mother under that description. Eventually, he ignorantly marries someone who happens to be his mother; thus, Oedipus acts out of character.

But acting out of character on the nonvoluntary conception cannot be a counterexample to the character condition, because such actions are fully excused. Oedipus is not morally responsible for ignorantly marrying his mother, because this kind of non-culpable ignorance is intuitively exculpatory. A counterexample must feature an agent who is clearly morally responsible for acting out of character.

1.2. The First-Time Action-Type Conception

Suppose we understand “acting out of character” to mean that a person performs an action-type for the first time (cf. Doris 2006: 478; Duff 1993: 375). This conception of acting out of character presupposes the *summary view* of character. According to the summary view, a character trait just is a pattern of past thinking and acting (Buss and Craik 1983: 106; Johnson 1997: 43). When a person performs a particular kind of action for the first time, there is no pattern of past actions of that kind, and so the action is out of character. There are, however, two reasons why proponents of the character condition are committed to rejecting the summary view in favor of the dispositional view of character.

First, only the dispositional view can account for the part of the character condition that character must motivate the action. According to the summary view, character is merely a pattern of past actions; it includes no tendencies or motivations to act. The dispositional view, however, describes character traits as stable mental tendencies to notice, think, feel, and act in various ways in trait-relevant circumstances. For example, a compassionate person has a heightened tendency to notice the suffering of others, to believe that their suffering is bad, to feel pity for them, and to be motivated to act in ways that alleviate their suffering. Thus, the character condition presupposes a dispositional view of character.

Second, only the dispositional view can capture the diachronic motivation for the character condition. The dispositional view describes character traits as enduring mental dispositions, which can explain why it makes sense to hold a person morally responsible through time. But the summary view describes character merely as a pattern of past actions. If a person’s having performed a past action does not suffice as an enduring ground for moral responsibility, neither does a pattern of past actions. Thus, the character condition presupposes a dispositional view of character.

On the dispositional view, however, it is possible for a person to act in character when they perform an action for the first time. For example, a person raised among thieves can have character traits that incline her to steal before she steals for the first time. Therefore, being morally responsible for performing an action-type for the first time is insufficient to be a counterexample to the character condition.

1.3. The Infrequent Action-Type Conception

Suppose we understand “acting out of character” to mean that a person acts in a certain kind of way only infrequently (Doris 2006: 479). Due to its statistical nature, this conception also presupposes the summary view. It thereby inherits the problems raised for the previous conception. On a dispositional view, it is possible that a person’s character trait disposes her to *A* and yet she *As* only infrequently. As a result, being morally responsible for an action of a type that is infrequently performed is also insufficient to be a counterexample.

1.4. The Integration/Probability Conception

Suppose we understand a person’s “acting out of character” to mean that she acts from weaker or less integrated mental dispositions (cf. Altshuler 2013: 187). On this view, a mental disposition is a character trait only if it is part of her stronger or more integrated mental dispositions. For example, Sally is overall tight-fisted with weak mental dispositions to be generous. In this case, Sally acts out of character when she pays for everyone at dinner.

There are two problems with appealing to this conception to provide a compelling counterexample to the character condition. First, it is intuitive that weaker or less integrated mental dispositions can be part of a person’s character; they are merely the weaker or less integrated part. That is, Sally can be divided against herself with stronger and weaker character traits opposed to each other. Second, and more importantly, the character condition’s general motivation extends to weaker or less integrated mental dispositions. Consider a plausible definition of character that fits the general

motivation for the character condition: a mental disposition is a character trait if and only if (i) it is enduring, (ii) it would be activated in a broad enough range of similar circumstances, and (iii) a person with it is thereby an appropriate object of normative assessment—in particular, it reflects on her being a good, bad, or intermediate person (cf. Miller 2014: 15). This definition explains the character condition's role in holding persons morally responsible: character traits are enduring features of persons that are normatively assessable. But notice that weaker or less integrated mental dispositions can be enduring, thick, and normatively assessable. For example, Sally's meager generosity is stable, and it reflects well on her to some degree, at least in contrast to her lacking any generous dispositions. As a result, weaker or less integrated mental dispositions should be character traits for proponents of the character condition. Thus, being morally responsible for acting from weaker or less integrated mental dispositions itself is insufficient to provide a counterexample to the character condition.

1.5. The Identification Conception

Suppose we understand “acting out of character” to mean that the action stems from mental dispositions with which she doesn't identify. On this view, a mental disposition counts as a character trait only if the agent identifies with it (Williams 1981: 14).

The identification conception has problems similar to those raised for the last conception. First, it seems plausible that a mental disposition with which a person doesn't identify can be part of her character. For example, a person's racist tendencies can be part of her character even if she doesn't identify with them (Hursthouse 1999: 116n7). Second, the general motivation for the character condition covers mental dispositions with which a person isn't identified, because those dispositions can be stable, thick, and normatively assessable. So, being morally responsible for acting from mental dispositions with which one doesn't identify is itself insufficient to be a counterexample to the character condition.

1.6. The Global Trait Conception

Suppose we understand “acting out of character” to mean that the action is at odds with unmixed global character traits (Sher 2006: 22). Global character traits would require at least consistent behavior across a very wide range of circumstances relevant to the particular trait, and stable behavior in such circumstances over time (Doris 2002: 22). A globally temperate person is strongly disposed to drink the right amount of alcohol at home, at happy hour, with strangers, on a date, with family, while tired, while happy, and so on. Global temperance is an *unmixed* trait, because it disposes a person only to act temperately—and not both temperately and intemperately. Thus, a globally temperate person acts out of her unmixed global character if she drinks too much on a particular occasion, because having global temperance rules out possessing global intemperance.

But proponents of the character condition shouldn't hold that there are only unmixed global character traits, because unmixed global character traits don't adequately describe actual human character. People have many conflicting sides to their character. A common criticism of film characters is that they are one- or two-dimensional due to their being insufficiently like actual people who have many traits in tension with each other.¹

This commonsense observation that human character is multifaceted is supported by social psychological studies that suggest that mundane features of circumstances have radical and disparate effects on moral behavior (see Doris 2002; Miller 2013).² This empirical literature has led some philosophers to believe that unmixed global traits such as Aristotelian moral virtues and vices aren't instantiated, or are rarely instantiated, in human life, because the research suggests that human behavior isn't cross-situationally consistent in the way that unmixed global character traits require (Miller 2014: 195).

There are two leading metaphysics of character traits that explain this fragmentation.³ First, there is John Doris's (2002: 64) local traits view. On this view, human beings have many local

character traits that dispose them to certain kinds of behavior in finely specified circumstances. A public-temperate person would be inclined to drink the right amount in public, but perhaps not on his own, or with his family, etc. Importantly, if a person with a global character trait of generosity acts in a stingy way that manifests local stinginess, the stingy action is in character. Thus, acting out of unmixed global character itself doesn't suffice for acting out of character. Second, there is Christian Miller's (2013) mixed trait view. On this view, global character traits needn't be unmixed, and a mixed global helping trait issues in a broad range of helpful and unhelpful actions. Miller's view implies that all behavior in a help-relevant situation can be explained as being in character (Doris and Spino 2015), because any surprising helping and non-helping action can be explained as revealing a new side of the mixed trait. Miller (2013: 20n38) himself agrees that it is unclear what helping and non-helping behavior would be inconsistent with the relevant mixed global trait.

What, then, are the prospects for the global trait conception providing a counterexample to the character condition? If all global character traits are mixed, then people don't act out of character, because this conception of acting out of character requires having unmixed global character traits. But if some global character traits are unmixed and if a person acts out of such character, the action needn't be out of character altogether, because it can manifest local character. Thus, acting out of unmixed global character is insufficient to be a counterexample to the character condition.

None of these six conceptions *themselves* suffice to provide a counterexample to the character condition. So, if someone acts out of character in a way that is aptly described by one of these six conceptions and it is a genuine counterexample to the character condition, it is a counterexample *because* she acts out of character also according to the seventh conception: the pure reasons conception.

2. PURE REASONS CONCEPTION

Suppose we understand "acting out of character" to mean that an agent acts for a reason that is wholly disconnected from her character. Whatever reasons are exactly, they are less "thick" and temporally stable than character traits (cf. Miller 2014: 24–32). It does seem conceptually possible for an agent to have a reason that isn't generated by, influenced by, or connected to her local or global character. If an agent can act on such a reason *and* be morally responsible for so acting, the out of character objection would refute the character condition on moral responsibility.

It might seem straightforward to come up with a case in which a person acts out of character in this way. For example, a professor is having a bad day: his paper was rejected. After getting the news, he harshly reproaches a teaching assistant. The other teaching assistants all comment that this is so unlike the professor; he's normally so mild. Even so, they blame him for his "out of character" action.⁴ There are various clear senses in which the professor is both acting out of character *and* morally responsible for it. The professor may be acting on a subtle part of his character such as a weak, local, peripheral, or unendorsed character trait. The professor's action can plausibly be explained in at least one of these ways precisely because his subtler character traits that govern anger have had their stimulus conditions satisfied by the rejection of his paper. In our view, real life cases of acting out of character can always plausibly be explained in one of these ways (see Hume 1975: 88). Importantly, however, none of these conceptions of "acting out of character" themselves generates a counterexample to the character condition on moral responsibility. So far, then, the professor case is not a counterexample to the character condition.

But suppose it is just *stipulated* that the professor is acting from a pure reason—that is, one that is entirely disconnected from his character. This means that the professor's subtle anger-governing character traits, including those that are local, weak, peripheral, or unendorsed, do not influence or shape his reason to reproach his teaching assistant, and none of his other character traits influences or shapes it either. Isn't this case now enough to undermine the character condition?

To answer this question, note first that it isn't obvious that the stipulation respects actual human psychology. Until now we have considered a range of cases of acting out of character that clearly occur in real life. The stipulation, however, appears to override the fact that our actions are always explicable in terms of overt or subtle features of character. Although it is conceivable that a person acts on a pure reason and so we do not deny its conceptual possibility (cf. Quinn 1993: 236), we do deny that acting on a pure reason is something human beings obviously do.⁵ We also deny that anyone can point to a real life case as obviously exemplifying the pure reasons conception of acting out of character. If actual human persons do not in fact act out of character according to the pure reasons conception, the stipulated professor case is *merely* conceptually possible. And if the out of character objection is to refute the character condition on moral responsibility, then merely conceptually possible cases, the only cases in which people obviously act out of character according to the pure reasons conception, must be good data on which to build an account of moral responsibility.

Proponents of the out of character objection now face an unnoticed methodological burden. While it is common to appeal to merely conceptually possible cases to learn about the nature of moral responsibility (e.g., Frankfurt 1969; Mele 2006: 188–189; Pereboom 2001: 110–117), it is also common enough to reject them as good data for theorizing (e.g., Arpaly 2006: 109–116; Cowley 2014: 16–17). For the latter group of philosophers, the out of character objection fails to show that the character condition is false. Some philosophers in the former group also believe that we should be less confident in our intuitions about merely conceptually possible cases (McKenna 2008: 157; Mele 2019: 31–32). As a result, the intuition that the professor acting on a pure reason is morally responsible for harshly reproaching the teaching assistant should get less traction in our theorizing about moral responsibility, and so we should be less confident that there is a counterexample that refutes the character condition.

The out of character objection's success therefore depends on the outcome of this debate. There is a deep methodological disagreement upon which the fate of the out of character objection rests, and further methodological investigation is required to illuminate whether the out of character objection is successful.

3. CONCLUSION

People act out of character: they do things they have never done before, they act on traits with which they do not identify, they act on local traits, and so on. They often are morally responsible for acting out of character in these ways, which is why the out of character objection seems so compelling. But we have argued that those actions themselves do not pose a problem for the character condition on moral responsibility, because those actions may be in character on the account of character to which proponents of the character condition should be committed. Subsequently, we agreed that there may be a clear counterexample when a person is morally responsible for acting out of character according to the pure reasons conception. But it is not obvious that people act out of character in that way in real life. To get a clear counterexample, then, proponents of the out of character objection must rely on merely possible cases. If merely possible cases are not good data for theorizing about the nature of moral responsibility, there is no clear counterexample to the character condition. Even if such cases provide data suitable for theorizing, the persuasive power of the out of character objection would still be mitigated, because we should be less confident in our intuitions about cases that are remote and do not obviously resemble real life. So, although the out of character objection is often taken to provide a compelling counterexample to the character condition based on real life cases, either there is no clear counterexample or we should be less confident in our intuitions about the counterexample. Either way, the out of character objection loses its power.

NOTES

We are grateful for comments, suggestions, and objections from higher seminar participants at Lund University, Stockholm University, and the University of Gothenburg as well as from the discussion group participants at the Beacon Project at Wake Forest University. We thank also several anonymous referees from this journal for comments and objections that led to great improvements in this paper.

1. This criticism has its roots in Lajos Egri's (1942) *The Art of Dramatic Writing*.
2. We merely buttress our commonsense claim by appeal to social psychology. Even if some of these experiments do not replicate, social psychology would still provide some support for our claim.
3. Most social psychologists and philosophers are skeptical about the claim that we lack character altogether (see Miller 2014: 199–200).
4. Thanks to an anonymous referee for suggesting this case.
5. If all intentional actions are in character, one might worry that this would generate dialectical problems: there would be no character-based excuses and the character condition would be trivial. But such worries can be quelled. First, there can still be character-based excuses in a *full* account of moral responsibility because, recall, the character condition is a mere necessary condition. On Arpaly and Schroeder's (1999: 172–173) whole-self view, for example, acting on a local character trait can generate a partial excuse if the trait is unrepresentative of the whole self. Second, the character condition isn't trivial; as noted earlier, it is explicitly about explaining why a person continues to be morally responsible after the action recedes into the past.

REFERENCES

- Altshuler, Roman. 2013. "Practical Necessity and the Constitution of Character," in *The Moral Philosophy of Bernard Williams*, ed. Alexandra Perry and Chris Herrera. Newcastle: Cambridge Scholars Publishing.
- Arpaly, Nomy. 2006. *Merit, Meaning, and Free Will*. Princeton, NJ: Princeton University Press.
- Arpaly, Nomy, and Timothy Schroeder. 1999. "Praise, Blame, and the Whole Self," *Philosophical Studies* 93(2): 161–188. <https://doi.org/10.1023/A:1004222928272>
- Brink, David. 2013. "Situationism, Responsibility, and Fair Opportunity," *Social Philosophy and Policy* 30(1–2): 121–149. <https://doi.org/10.1017/S026505251300006X>
- Buss, David M., and Kenneth H. Craik. 1983. "The Act Frequency Approach to Personality," *Psychological Review* 90(2): 105–126. <https://doi.org/10.1037/0033-295X.90.2.105>
- Cowley, Christopher. 2014. *Moral Responsibility*. New York: Routledge. <https://doi.org/10.4324/9781315729831>
- Doris, John. 2002. *Lack of Character*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9781139878364>
- Doris, John. 2006. "Out of Character: On the Psychology of Excuses in Criminal Law," in *Ethics in Practice*, 3rd ed., ed. Hugh LaFollette, 474–484. Oxford: Blackwell Publishing.
- Doris, John, and Joseph Spino. 2015. "Review of *Character and Moral Psychology*, by Christian Miller," *Notre Dame Philosophical Reviews*. <https://ndpr.nd.edu/reviews/character-and-moral-psychology/>.
- Duff, R. A. 1993. "Choice, Character, and Criminal Liability," *Law and Philosophy* 12(4): 345–383. <https://doi.org/10.1007/BF01000637>
- Egri, Lajos. 1942. *The Art of Dramatic Writing*. New York: Simon & Schuster.
- Elliot, Jay R. 2017. *Character*. London: Bloomsbury.
- Frankfurt, Harry G. 1969. "Alternate Possibilities and Moral Responsibility," *The Journal of Philosophy* 66(23): 829–839. <https://doi.org/10.2307/2023833>
- Hume, David. 1975. *Enquiries Concerning Human Understanding and Concerning the Principles of Morals*, 3rd ed., ed. L. A. Selby-Bigge. Oxford: Clarendon Press. <https://doi.org/10.1093/actrade/9780198245353.book.1>
- Hume, David. 1978. *A Treatise on Human Nature*, 2nd ed., ed. L. A. Selby-Bigge. Oxford: Clarendon Press.
- Hursthouse, Rosalind. 1999. *On Virtue Ethics*. Oxford: Oxford University Press.

- Johnson, John A. 1997. "Units of Analysis for the Description and Explanation of Personality," in *The Handbook of Personality Psychology*, ed. Robert Hogan, John Johnson, and Stephen Briggs, 73–93. San Diego: Academic Press. <https://doi.org/10.1016/B978-012134645-4/50004-4>
- Kauppinen, Antti. 2016. "Blame and Character in Hume and Beyond," in *Questions of Character*, ed. Iskra Fileva, 46–62. Oxford: Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199357703.003.0003>
- McKenna, Michael. 2008. "A Hard-Line Reply to Pereboom's Four-Case Manipulation Argument," *Philosophy and Phenomenological Research* 77(1): 142–159. <https://doi.org/10.1111/j.1933-1592.2008.00179.x>
- Mele, Alfred R. 2006. *Free Will and Luck*. Oxford: Oxford University Press. <https://doi.org/10.1093/0195305043.001.0001>
- Mele, Alfred R. 2019. *Manipulated Agents: A Window into Moral Responsibility*. Oxford: Oxford University Press. <https://doi.org/10.1093/oso/9780190927967.001.0001>
- Miller, Christian. 2013. *Moral Character: An Empirical Theory*. Oxford: Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199674350.001.0001>
- Miller, Christian. 2014. *Character and Moral Psychology*. Oxford: Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199674367.001.0001>
- Nelkin, Dana. 2005. "Freedom, Responsibility, and the Challenge of Situationism," *Midwest Studies in Philosophy* 29(1): 181–206. <https://doi.org/10.1111/j.1475-4975.2005.00112.x>
- Pereboom, Derk. 2001. *Living without Free Will*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511498824>
- Quinn, Warren. 1993. *Morality and Action*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9781139172677>
- Sher, George. 2002. "Blameworthy Action and Character," *Philosophy and Phenomenological Research* 64(2): 381–392. <https://doi.org/10.1111/j.1933-1592.2002.tb00007.x>
- Sher, George. 2006. *In Praise of Blame*. New York: Oxford University Press. <https://doi.org/10.1093/0195187423.001.0001>
- Sripada, Chandra. 2016. "Self-Expression: A Deep Self Theory of Moral Responsibility," *Philosophical Studies* 173(5): 1203–1232. <https://doi.org/10.1007/s11098-015-0527-9>
- Williams, Bernard. 1981. "Persons, Character, and Morality," in *Moral Luck: Philosophical Papers (1973–1980)*, 1–19. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9781139165860.002>