

1. Background

Moral responsibility: I'm concerned with the sort of moral responsibility that makes one *deserving* of whatever is bad about blame, or of other consequences associated with being blameworthy.

Strawsonian account of MR: One that understands MR in terms of quality of will, or due regard. (E.g. Strawson, Arpaly, McKenna, Hieronymi)

Plan: Bring together a number of features that I think that such an account should have, begin to answer some outstanding questions.

2. Previously defended general account of blame (Björnsson 2017a,b; 2022)

Blame has its home in domains of skill. Build on abilities to

- (i) Identify how it goes with values constituting what is to be promoted the domain.
- (ii) Identify when outcomes for those values are due to agents' responsiveness, which can thus be reinforced or adjusted accordingly (feedback learning). (Involves distinguishing when outcomes are "normal", or in line with systematic tendencies, and when they "just happen".)
- (iii) Normatively expect a certain level of skill: to be disposed to identify and react to—be alarmed by—instances falling (or threatening to fall) below such levels so as to promote correction.

In jointly exercising these, we see something as someone's fault:

FAULT: For Y to be X's fault is for Y to be bad and explained in a normal way by X's responsiveness falling short of applicable standards.

Seeing A as F: Quasi-perceptual state distinct from the explicit personal-level judgment that A is F. I might judge that something isn't my fault but continue to see it as my fault.

Core aspect of blaming. Specific general function of blame, given function of the three abilities:

FAULT FUNCTION: The specific general function of seeing Y as X's fault is to prompt satisfaction of the relevant standard of responsiveness. The normal way for this to happen is (i) for the "demand" mechanisms to direct negative attention to X's substandard responsiveness, (ii) for the mechanisms for feedback learning to direct negative attention to the role of that responsiveness in bringing about Y, and (iii) for the negative attention to these features to prompt improvement of the responsiveness.

BLAME: To blame X for Y is for one to see Y as X's fault and for the mechanisms normally performing the function of this state to be operative, or to communicate such blame.

BLAMEWORTHINESS: X is *to blame* for Y, X is *blameworthy* for Y, and blaming X for Y is *fitting* or *correct*, if, and only if, Y is bad relative to a justifiable value and explained in a normal way by X's responsiveness falling short of applicable justifiable standards.

Explains: (i) Ubiquity of blame. (ii) Normality condition on relation to object of blame. (iii) Blame's structure: unlike most attitudes, it has *object* and *target*. (iv) Unified account of excuses, exemptions, justifications. (v) The abilities required to be subject to blame.

3. Previously defended account of moral blame specifically

Strawsonian account: The relevant responsiveness is a responsiveness of the will to moral values.

MORAL BLAMEWORTHINESS: X is *morally blameworthy* for Y if, and only if, Y is morally bad and explained in a normal way by X's responsiveness to moral values falling short of applicable moral standards.

Why accept an account like this?

- (i) Gets right what the general account got right.
- (ii) Unifies standard conditions on blameworthiness (control, knowledge, qow, indirect resp)
- (iii) Gets motivating features of paradigm cases of moral blameworthiness clearly right.
- (iv) Explains why someone who is as responsive to available reasons as can be demanded of them isn't to blame for failure to achieve good.
- (v) Natural account of degrees of blameworthiness:

DEGREE FROM DEVIATION: The degree of blame X deserves for Y depends on the value of the outcome and on how much of a deviation of the agential aspect from what can be properly demanded is required in the (normal) explanation of the object of blame.

- (vi) Degrees of blameworthiness in relation to degrees of ignorance or lack of awareness (inspired by Arpaly 2003: 160–62):

- (1) X knows that not helping is morally wrong because she is uniquely placed to help and can do so at very small personal cost, but she nevertheless refrains because she cares little about the plight of others or about morality.
- (2) X knows that she is uniquely placed to help but refrains from helping because she cares little about the plight of others and is committed to a form of egoistic morality according to which her small personal benefit justifies not helping.
- (3) X knows that she is uniquely placed to help but cares too little about morality generally to conclude that the action is morally wrong, and too little about the plight of others to help.
- (4) X has all the evidence available to conclude that she is uniquely placed to help and that it would be wrong not to, but cares too little about the plight of others or about morality generally to draw those conclusions.
- (5) X had signs available that would have led anyone caring about the plight of others and about moral matters to look up, take in the situation, reflect, realize that she is uniquely placed to help at little personal cost, conclude that she ought to do so, and jump to action.
- (6) X had no such signs available.

Intuitively, line is sharp between (5) and (6). Requirements of clear-eyed akrasia: between (1) and (2). Requirements of awareness of wrong-making reasons: between (3) and (4).

- (vii) Explanatory component of responsibility and blame attributions lets us explain away various forms of skeptical worries, including ultimacy worries.

4. Responsiveness

Various ways of understanding quality of will:

Maxim from which one acts? Too narrow; failures to intend, or even consider intending.

Normative judgment? Needs to be embodied. Better talk about responsiveness.

Emotional disposition (“caring”, in a thick sense)? Too strong. Blame avoided without such disposition.

Lack of sympathy (slights; as these prompt reactive attitudes (Shoemaker))? No, victims with adaptive preferences; victimless crimes.

Desire/indifference? Arpaly & Schroeder: To intrinsically desire that P / that not-P is to construe P as a reward / not-P as a punishment. (Reward/punishment-systems for learning.) One is to blame for an action when it is rationalized by lack of sufficient desire for the good.

- (Caveat: not concerned with desert-entailing blameworthiness.)
- (Blame for outcomes? *Normal* explanation more generally apt than *rationalizing* explanation.)
- Not clear that what morality demands of us is merely that we construe the good as reward and the bad as punishment.
 - o Being alarmed by my failure to give someone’s interests or point of view too little weight, I reveal a strong desire to have done better by this value. What I blame myself for is not lacking that desire, but being unresponsive to the reasons at hand.
 - o *Reply*: Trace back to opportunity to learn responsiveness, and lack of desire then?
 - o *But*: Still seems to misconstrue mechanism by which we learn from what it is demanded that we learn!

5. Problem: Blameworthiness for out of character wrongdoing

Forgotten Children: Caring parent sometimes just forget to pick up their child from daycare, or forget them in the back seat of a hot car.

Forgotten Milk: Normally reliable Randy forgets to buy milk on the way home as he had promised, as his thoughts drifted.

Problem with these and many similar cases: it might seem that the agent does care, perhaps more than others, but are nevertheless morally to blame for failures to act on that caring (e.g. Sher 2009).

First response (Björnsson 2017b:143): Nonmoral blame?

Fatal Failure: To disarm the ticking bomb in a densely populated building, you had to quickly solve a complex mathematical puzzle. Unfortunately, in spite of a highly focused effort, you failed to prevent the lethal outcome.

You can blame yourself for the failure. Arguably, you are not morally to blame for your failure, or for the fatal outcome. You don’t deserve to suffer from it (though agent-regret would be fitting).

- *But*: Not clear that this response gets cases right, as they seem to display failures to care enough, or be suitably responsive *on the occasion*.

Second response: Agents are not to blame in those situations. Moral agents are capable of reliability, not infallibility, and it is unreasonable to blame people for failures that we cannot avoid.

- *But:* Distinguish different objects of blame. To say that we cannot but fail sometimes is not to say that we couldn't avoid failing on the occasion.

6. Solution: Externally defined demands on responsiveness

What is the measure of *enough* responsiveness?

- No careometers!
- Responsive in *enough* possible worlds? But what worlds? And doesn't this just reinforce the *out of character* problem?

Generally, how should we think of the relevant requirements on caring?

Requirements based on (relative) importance

- Can't be the whole story. Less important things might be easier to respond to depending on circumstances.

Requirements based on abilities and opportunities

- But how do we measure these abilities and the quality of the opportunities?

Requirements based on participation

- If you participate in a practice, you are required to have enough skill to do action A under circumstances C, for various As and Cs.

Suggestion: Out-of-character cases are cases where one fails to be responsive enough to relevant values to A under C.

Remaining question: How far short does one fall when being generally reliable but failing on occasion?

7. Moral blameworthiness and desert? (Björnsson 2022; Björnsson & Eskens draft)

Strawsonian / Wittgensteinian answer: *This is how we roll!*

- *But:* Explanatorily and normatively incomplete.

New proposal: *Balancing norms and moral harms.*

Balancing norms: Norms requiring that we give people's interests, points of view and other values a certain comparative weight over time. If we give someone less weight, it becomes increasingly more important that we give them more weight.

Giving weight: Responding to values, interests or points of view so as to promote or act on them.

Moral harms: We have interests in being given the weight that morality requires.

Culpable wrongdoing: Involves perpetrator (and thus moral community) giving less weight than morality requires. Calls for rebalancing of interests (including moral interest) and points of view.