The Subtleties of Fit:
Reassessing the Fit–Value Biconditionals

Rachel Achs and Oded Na'aman

Forthcoming in Philosophical Studies. Please cite the published version.

ABSTRACT

A joke is amusing if and only if it's fitting to be amused by it; an act is regrettable if and only if it's fitting to regret it. Many philosophers accept these biconditionals and hold that analogous ones obtain between a wide range of additional evaluative properties and the fittingness of corresponding responses. Call these the *fit–value biconditionals*. The biconditionals give us a systematic way of recognizing the role of fit in our ethical practices; they also serve as the bedrock of various metaethical projects, such as fitting-attitude analyses of value and the 'fittingness first' approach. Yet despite the importance of the biconditionals, there is very little discussion of their proper interpretation. This paper argues that any plausible interpretation of the fit–value biconditionals must disarm several kinds of apparent counterexample. For instance, that an achievement is pride-worthy doesn't imply it is fitting for me to take pride in it because the achievement might not be mine or that of anyone close to me; that a joke is amusing doesn't imply it is fitting for me to be amused by it for six straight months; and that a person is loveable doesn't imply it is fitting for me to love him romantically because that person might be my sibling. We consider possible responses to such counterexamples and develop what we consider the most promising interpretation of the biconditionals. The upshot is that certain widespread assumptions about fit and its relation to value and reasons should be reconsidered.

Keywords: fittingness; value; evaluative properties; reasons; biconditionals

I. Introduction: Magnolia Is Not Perfect

Consider a person, call her Magnolia, who loves only what is loveable, desires only what is desirable, is proud only of what is pride-worthy, and admires only what is admirable. Magnolia regrets only what is regrettable, feels contempt only for what is contemptible, laments only what is lamentable, and blames only what is blameworthy; she prefers only what is preferable, envies only what is enviable, and trusts only what is trustworthy. Everything Magnolia feels, every attitude she adopts, and every reaction she has can be said to match a certain type of value in its object. Can we infer from this description of Magnolia that all of her responses are fitting? That is, can we infer that all of Magnolia's responses are merited by, or called for by, their objects?

Common view would seem to have it the answer to this question is "yes." Since each of Magnolia's responses matches the type of value in its object, each of those responses is fitting.[1] To be sure,

[1] In the introduction to a new volume on fittingness, for example, Chris Howard and R. A. Rowland gloss fit as

> the relation in which each of our responses stand when I admire an admirable effort, you laud a laudable performance, Beri believes a credible proposition, and Dhitri desires a desirable outcome. Likewise, it is the relation that fear stands in when its object is fearsome, that love stands in when

Magnolia's responses might not be beyond criticism. Perhaps, although Magnolia has fitting responses, she does not have them for the reasons they are fitting. If so, then Magnolia's responses may be fitting, and yet not *held fittingly*. Alternatively, Magnolia's responses might be based on the reasons they are fitting, and yet her evidence might (misleadingly) recommend different responses than the ones she has. If so, then her responses would be *subjectively unfitting* – that is to say: not fitting in light of her evidence. And even if Magnolia's responses are based on the reasons they are merited *and* accord precisely with her evidence, she may have *moral* or *prudential* reasons against envying the enviable or regretting the regrettable. Certainly, there are times when it is best, *all-things-considered*, not to respond to an object according to its value – such as when the consequences would be dire. Thus, Magnolia might be failing to respond on the proper bases, or properly given her evidence, or as she has moral, prudential, or overall reason to respond. Still, it may appear that, in terms of the objective fittingness of her responses, Magnolia reaches perfection.[2]

The main thesis of this paper is that this common view needs clarification because when taken at face value it is false: It does not follow from our description of Magnolia that all her responses are fitting. Magnolia might unfittingly love someone loveable, for example, if she loves a parent in the way it is fitting to love a sibling or if she loves a sibling in the way it is fitting to love a romantic partner. Magnolia might feel unfitting shame about what is shameful, if the shame she feels concerns a perfect stranger's nudity rather than her own. Furthermore, Magnolia might unfittingly regret something regrettable or be unfittingly amused by something amusing, if she regrets a minor mistake or is amused by a mediocre joke for too long. There are many more examples. In short: The fact that a certain type of response matches a type of value in its object does not entail that it is fitting by everyone, in every situation, permanently. So it is a mistake to infer from the evaluative properties of the objects Magnolia responds to that her specific responses are fitting.

Our thesis concerns the 'fit–value biconditionals' – biconditionals of the form "*x* is admirable if and only if it is fitting to admire *x*," "*x* is blameworthy if and only if it is fitting to blame *x*" – which are taken to hold between objects which bear certain evaluative properties and the fittingness

---

> its object is lovable, and that blame stands in when its object is blameworthy. (Howard and Rowland 2022, 1)

[2] The preceding paragraph distinguishes between whether a response is fitting in the sense of being normatively supported by fit-related considerations (the analogue of a belief being 'propositionally justified') and whether, in addition to that, it is also *held fittingly*, on the basis of reasons that make it fitting (the analogue of a belief being 'doxastically justified'). We are focused here on the former notion. The common view (with which we take issue) is that admiration is always the *normatively* fitting response to the admirable.

The preceding paragraph also distinguishes between 'objective' and 'subjective' fit (Chappell 2012, 689 n10), or as others put it, between 'fit' and 'warrant' (D'Arms and Jacobson 2000a, 78; Scarantino and De Sousa 2018, sec. 10.1). An attitude is subjectively fitting if the agent is justified, given her evidence, in taking it to be objectively fitting. We, like most, focus on objective fit, since we take it to be the more basic notion. But see (Fritz 2021, 8559 n10) for the prioritization of subjective fit. In summary, what we're concerned with here is normative, objective fit – henceforth, 'fit.'

of etymologically connected responses.[3] To say that the fit–value biconditionals are accepted widely is, as Chris Howard has recently put it, "understating things" (2022, 3).[4] Indeed, the biconditionals are important to contemporary theorizing about fit in several ways. For instance, they provide a template for analyzing value-properties in terms of fitting responses – a crucial item in the toolkit of anyone who holds fit to be the most explanatorily fundamental normative notion (Chappell 2012; Howard 2019b; McHugh and Way 2016). The biconditionals are also standardly wheeled out to help new initiates latch on to the notion of fit. 'Fit,' we are told, *is* the relation that admiration bears to its object when and only when its object is admirable, and that fear bears to its object when and only when its object is fearsome.[5] Thus, both the theoretical promise of the notion of fit, and, indeed, our very grasp of the relation, seem to depend on the biconditionals.

However, as our doubts about Magnolia's responses suggest, when left unqualified, the fit–value biconditionals are vulnerable to counterexamples. Amusement at the amusing is not *always* fitting; nor is shame at the shameful. Perhaps this means that some who subscribe to the biconditionals are implicitly relying on a qualified interpretation of them – one that avoids such apparent counterexamples.[6] But, if so, then we should explicitly assess possible interpretations of the biconditionals, to examine their plausibility and implications. For we shouldn't be drawing inferences from the biconditionals without being clear on what they say. Moreover, if it turns out

[3] Throughout our discussion, we'll use the terms 'evaluative property' and 'value-property' interchangeably.

[4] Philosophers who endorse the biconditionals include Brandt 1946; D'Arms and Jacobson 2000a, 2000b; Howard 2018, 2022; Schroeder 2010; Tappolet 2016; and Wiggins 1987. Berker 2022 does too, although he prefers to think of the properties on the left-hand side not as *value* properties, but rather 'thick-fittingness properties' (45–47). Our primary interest here is in properties such as the amusing, the blameworthy, the desirable, etc., regardless of whether these are value-properties, and our conclusions will bear on any claim about the relation between these properties and fitting responses. Whether and how our conclusions will also bear on attempts to analyze goodness in terms of fitting responses will depend on what one takes the relation between these properties and goodness to be.

[5] In addition to the citation in note 1, see also Howard 2018, 2.

[6] For instance, in a well-known article, D'Arms and Jacobson at one point qualify the claim that something is fearsome if and only if it is fitting to fear, suggesting that perhaps they think of the biconditionals more as loose rules of thumb: "To think the tiger is fearsome is to think fear at it appropriate [in the sense of fitting]," they write, "but only when the tiger is nearby and on the loose – not, for instance, while you sit reading this article" (2000b, 729). But others think of the biconditionals as "equivalences" and hold that there cannot be a change in the fitting response to an object without a corresponding change in its evaluative property (Howard 2018, 2; 2022, 3) – a principle that could only be true if the covariances between an object's value-properties and the fittingness of corresponding response types were extremely strict. And even those who hint that the biconditionals need some qualification then slip into speaking interchangeably of the considerations that are relevant to a response's fittingness and those that are relevant to whether its object possesses the corresponding evaluative property, as if nothing (say, about a particular subject's relation to an object, or about the quality of her response to it) could undermine the fittingness of one of her responses without also affecting whether its object possesses the corresponding type of value-property. In the very same article, D'Arms and Jacobson go on to refer to fittingness as "the particular species of appropriateness of response that is relevant to property ascription" (2000b, 730) – as if, contrary to their previous comment, nothing could bear on whether fear of the tiger is fitting *without* also bearing on whether it's fearsome.

that there is no compelling way to qualify the biconditionals, then we are in danger of losing not only fit's explanatory promise but also our main method of latching onto the relation. In other words, without a proper interpretation of the biconditionals, the notion of fit, as a distinctive normative category, disintegrates.

In this paper, we confront the apparent counterexamples to the fit–value biconditionals so as to clarify what the relation between evaluative properties and the fittingness of etymologically connected responses actually is. We start by describing a series of counterexamples to the (unqualified) biconditionals, and then we mine existing discussions of similar cases to develop two potential lines of response. The first – *the wrong-kind-reason response* – holds that the biconditionals don't actually need qualification, because our counterexamples can be explained away by considerations unrelated to fit. The other – *the specification strategy* – holds that our counterexamples are disarmed by giving a description of the biconditionals which more carefully specifies the evaluative properties and valuing-responses at stake. We argue that neither of these responses to our cases is successful and, indeed, that a disjunctive strategy (which tries to combine them) also fails.

We instead endorse a third line of response, which holds that the biconditionals must be qualified in yet a different way. The fact that an object possesses a value-property entails that a certain type of response is fitting, but it does not entail by whom, when, and how that response is fitting; conversely, a fitting response entails that its object possesses a certain type of evaluative property, but it does not entail for whom, when, and how it possesses that property. Thus, the shameful merits shame and the regrettable merits regret, but objects that possess these properties merit these responses only from some people, in some ways, sometimes. We draw from this a lesson that's standardly overlooked: Considerations bearing on a response's fittingness *surpass* those that bear on the value of its object. Thus, we cannot do justice to the subtleties of fit by attending to value alone.

II. Counterexamples

We now present a series of apparent counterexamples to the fit–value biconditionals, grouping them into three kinds, for ease of exposition. The groupings are somewhat artificial. One might argue, for instance, that our second group is a subclass of our first, or, alternatively, a subclass of our third. Nevertheless, it helps to organize our cases. *Positionality*, *temporality*, and *quality*, correspond to failures of fit with respect to *who* has the response, *when* they do, and *how*.

Positionality Counterexamples

As standardly presented, the biconditionals don't delineate *for whom* the relevant valuing response is fitting. But sometimes an object possesses a certain evaluative property and yet the associated response is not fitting because of the agent's relation to the object. The following examples exhibit apparent failures of the fit–value biconditionals due to *positionality*:

> STANDING
> Poppy is blameworthy for performing some wrongful act. Violet lacks standing to blame Poppy because she has performed the same act repeatedly, without compunction. Violet's blame of Poppy seems unfitting.

PARTIALITY
Iris and Basil are competing for a job. Basil is more qualified, so it is preferable that Basil get the job. But Lily is Iris' good friend and has no special relationship with Basil. We propose it would not be fitting for Lily to prefer that Basil be hired.

ENABLING RELATION
Rose reads in the newspaper that Ivy won an award for her contributions to everything. Rose is filled with pride. Ivy's achievement is indeed pride-worthy. But since Rose bears no significant relation to Ivy, we believe her pride is not fitting.

DISABLING RELATION
Laurel's team is competing against Orchid's team in a race to find the cure for COVID-19. Laurel's team finds a cure first. Orchid's failure is disappointing, but it does not seem fitting for Laurel to be disappointed by Orchid's failure.

In all four cases, the instantiation of an evaluative property does not imply that its etymological mate is fitting. So the biconditionals apparently don't hold. Moreover, each of the four cases stands for a different subtype of positionality failure. There are various standing cases, various cases of partiality, and many responses that are fitting only given – or in the absence of – certain background conditions.[7] Therefore, there are many more examples of each subtype.

Temporality Counterexamples

As standardly presented, the biconditionals include no limitations on *when* the relevant valuing response is fitting. But sometimes an object has a certain evaluative property and yet the associated response is not fitting due to one's temporal relation to the object. The following examples exhibit apparent failures of the fit–value biconditionals due to *temporality*:

DIMINISHMENT
At lunch, Jasmine is amused by Marigold's amusing quip. A month has gone by and Jasmine still hasn't ceased chortling heartily at Marigold's quip. Jasmine's amusement no longer seems fitting.

TEMPORAL ASYMMETRY (FORWARD)
Having stopped following the news in January 2022, Dahlia continues, in March 2022, to fear Russia's invasion of Ukraine. Russia's invasion is fearsome, but it occurred in February 2022. Dahlia's later fear strikes us as unfitting.

TEMPORAL ASYMMETRY (BACKWARD)
Ginger grieves the death of Mark Twain. Twain's death is grievous. But then Mark Twain announces: "Reports of my death are greatly exaggerated." So Ginger's grief is unfitting.[8]

[7] For discussion of normative background conditions and modifiers, see Dancy 2004 and Bader 2016.

[8] We've described our two TEMPORAL ASYMMETRY cases as involving agents who are mistaken about what has occurred, because we presume that this sort of ignorance is what would most naturally lead a person to unfittingly fear a past event, or to grieve a future one. Strictly speaking, however, ignorance is

In all three cases, the instantiation of an evaluative property does not imply that its etymological mate is fitting. So the biconditionals apparently don't hold. Each one of the cases exemplifies a different subtype of temporality failure. In DIMINISHMENT, the fact that Jasmine has been amused for a while makes amusement no longer fitting. This is an example of a rationally self-consuming emotion (Na'aman 2021). In the two TEMPORAL ASYMMETRY cases, the relevant response is fitting only if its object is in the future or in the past.[9] Fear, anxiety, and hope are examples of forward-looking temporally asymmetric responses; regret, anger, and relief are examples of backward-looking temporally asymmetric responses. So there are many more examples of each subtype of temporality case.

Quality Counterexamples

As standardly presented, the biconditionals contain no limitations on *how* a valuing response is fitting. But sometimes an object has a certain evaluative property and yet the associated valuing response is not fitting due to its *quality*. The following examples exhibit such apparent failures of the fit–value biconditionals:

> SIZE
> Holly meets her colleague and they are both wearing the exact same wardrobe. This is surprising, but Holly is so surprised that she falls out of her seat and cannot recover for several minutes. Her surprise doesn't seem fitting.
>
> TONE
> Posy loves her loveable romantic partner. But her love for him does not seem fitting because she loves him as if he were her son.

In both cases, the instantiation of an evaluative property does not imply that its etymological mate is fitting. So the biconditionals apparently don't hold. In SIZE, the degree of surprise is not fitting to the degree of surprisingness of its object; in TONE, the love for a loveable object is off-key and therefore not fitting. Each case is an instance of a subtype of quality case. Failures of size and tone are possible in a wide variety of responses. Think of disproportional regret and joy about regrettable and joyous objects, or offkey shame and amusement about shameful and amusing objects.[10]

---

not necessary for generating TEMPORAL ASYMMETRY counterexamples to the unqualified biconditionals. For instance, one might imagine an agent who can't help but unfittingly and (irrationally) grieve a death she *knows* hasn't happened yet – in the same way that an agent might unfittingly (and irrationally) fear a non-venomous spider, in spite of knowing that it poses no threat.

[9] These cases have been discussed in debates about the metaphysics of time. See Prior 1959. For more recent discussion, see Yehezkel 2013, Pearson 2018, and Miller forthcoming.

[10] Additionally, Daniel Telech and Leora Dahan Katz argue that while the fitting tone of blame is often one of anger, in certain cases fitting blame takes the tone of disappointment (Telech & Dahan Katz 2022).

We take ourselves to have shown that there are many counterexamples to the fit–value biconditionals, left unqualified. And yet, these purported covariances play a central role in theorizing about fit. Can they be saved? Two potential strategies for disarming our counterexamples can be gleaned from the contemporary literature on fit. In the next sections we describe these strategies and, in turn, explain their limitations.

III. The Wrong-Kind-Reason Response

As D'Arms and Jacobson (2000a) famously observe, it is all-too easy to commit the "moralistic fallacy" of mistaking moral – or other 'wrong-kind' – reasons for considerations that bear on fit. The *wrong-kind-reason response* to our cases takes this point to heart and insists that they aren't really counterexamples to the biconditionals. It is true, one might hold, that the subjects in our cases all have valuing responses that are inappropriate. But our subjects' responses are not inappropriate in the sense of being *unfitting*. Rather, there exist mere value-related – so-called 'wrong-kind' – reasons for our subjects not to have fitting responses, which make their responses inappropriate in some *other* sense. For example, one who takes this line may propose that it is fitting to blame without standing, but nevertheless *morally* inappropriate to do so. Or one might hold that it is fitting to resent a minor mistake for many years, but nevertheless *prudentially* unwise.

Of course, 'wrong-kind' reasons *can* make fitting attitudes inappropriate (or, at the very least, inappropriate to *want* or *bring about*).[11] So if cases in which they did so were claimed to be counterexamples to the biconditionals, we would employ a version of this strategy to defend them. If it were claimed, for example, that desiring the desirable is unfitting when an evil demon has warned you not to, then we would respond that this case demonstrates nothing against the biconditionals at all. And we certainly agree with the sentiment behind the wrong-kind-reason response that we should be wary of mistaking (mere) value-related reasons for fit-related ones. Nevertheless, we contend that the wrong-kind-reason response is not a plausible reply to any case we have raised. Indeed, we find it evident that the valuing responses of our subjects are unfitting. If Ivy and Rose are perfect strangers, then Ivy's achievement doesn't *merit* Rose's pride. And a case of unplanned wardrobe coordination just doesn't *call for* over-the-top surprise.

While we suspect many will agree that 'wrong-kind' reasons can't explain away most of our cases, we think some will be attracted to employing this strategy at least in response to STANDING and DIMINISHMENT. After all, it is common to hold that hypocrisy does not affect the fittingness of blame at all. For example, R. Jay Wallace thinks that, if there's an issue with blaming sans standing, it is not that such blame is unfitting, but rather that it is "morally problematic" (2019, 545; see also Bell 2013, 267, and King 2019, 269). Similarly, philosophers often appeal to wrong-kind reasons to explain why emotions appropriately diminish over time. For example, Dan Moller argues that a grievous loss merits enduring grief even if there are "welfare enhancing" or "instrumental reasons" for the diminution of grief (fc, 7).

[11] Even wrong-kind-reasons skeptics (e.g., Hieronymi 2005; Way 2012) could employ a version of this strategy to argue that the attitudes in our counterexamples look suspect not because they are unfitting, but rather for other reasons (e.g., they're unfitting to want or bring about).

Yet we find it compelling that what is problematic about our subjects' attitudes is that they are unmerited by their objects – even in the most (supposedly) controversial of our cases. Our conviction is due to the presence of several indicators of fit in *all* the cases.[12]

For example, one indicator of fit is that it sounds natural to use fit-talk (e.g. 'worthy' or 'merited') to describe what is off about our subjects' responses. It sounds natural to use the language of fit in saying (of STANDING, for example) that "while Poppy may merit blame from *someone*, she doesn't merit *Violet's* blame" or to say that "while Poppy may be blameworthy, she isn't worthy of *Violet's* blame, since Violet lacks the standing to blame her." Similarly, it sounds natural to use fit-cognates to say (for instance, of Jasmine's amusement in DIMINISHMENT) that "a mild quip merits only short-lived amusement."

A second indicator of fit is the type of rational impact on our attitudes that recognition of fit-affecting considerations can have. Recognizing facts that bear on fit can rationally impact our attitudes directly. Recognizing facts that bear merely on the value of our attitudes, on the other hand, can rationally impact those attitudes only indirectly, by motivating us to manage them (Schroeder 2010; 2012). The considerations that bear on the attitudes of the subjects in our cases can have the type of rational impact characteristic of fit-affecting facts – particularly the type of impact characteristic of the absence of an enabler or presence of a disabler on an attitude's fittingness.[13] For example, the (newly recognized) fact that she's being a hypocrite can be a consideration directly in light of which Violet ceases to blame Poppy, just as, say, the (newly learned) fact that she was mistaken about being descended from some great general can be a consideration directly in light of which she ceases to feel pride in his accomplishments. This contrasts with the type of rational impact that mere value-related reasons can have. Realizing that her blame is making Poppy get defensive might motivate Violet to try and manage her blame, but it can't be a consideration that rationally influences Violet's blame directly.

A third indicator of fit is that unfitting attitudes seem to misrepresent. In saying this, we don't claim, as some do, that fit can simply be thought of *as* accurate representation (e.g., Tappolet 2016, 87). Our claim is rather that there is a deep connection between fittingness and accurate representation (which is precisely why thinking of fittingness in terms of representational correctness has seemed attractive to many philosophers). *When* a response is fitting, it represents correctly, and *when* a response is unfitting, it represents incorrectly – such that part of what seems off about an unfitting response is always that it seems to misrepresent. Again, this indicator of fit is present even in STANDING and DIMINISHMENT. Part of what seems off about the blame of

[12] For brevity, the following arguments focus on cases to which the wrong-kind-reason response seems most tempting. But cases analogous to a couple of our others (PARTIALITY and TONE) have also been dubbed "wrong-kind-reason" cases by some philosophers. Jonas Olson holds, for example, that considerations of partiality are 'wrong-kind' (2009, 374). So we take it Olson would claim that it is not unfitting for Lily to prefer that better-qualified Basil get the job, but rather morally wrong, given her friendship with Iris. Similarly, Chris Howard has claimed that considerations of love's quality bear not on its fit, but rather on assessments of other kinds (Howard 2019a, 128). Although our examples concern STANDING and DIMINISHMENT, our claim is that the indicators of fit that we mention in the main text are present in *all* of our counterexamples. So we think Olson and Howard are also wrong to treat such cases as they do.

[13] For discussion of fit-affecting enabling and disabling conditions, see Na'aman 2021 and Howard 2022.

a hypocrite, like Violet, is that such blame seems to falsely cast its target as having violated a norm which the subject, *herself*, respects.[14] And part of what seems off about Jasmine's month-long amusement is that it seems to misrepresent Marigold's quip as much funnier than it possibly could be.

Thus, although 'wrong-kind' reasons do sometimes make it inappropriate to fittingly value valuable objects, this is not what is going on in our counterexamples to the fit–value biconditionals. All of the cases we raise are ones in which subjects are responding *unfittingly*, and we hope that pointing out the indicators of fit present in our cases has helped to elicit this intuition.

Moreover, as we've already mentioned, the wrong-kind-reason response isn't even a tempting reply to some of our cases. For instance, it is baldly implausible to claim that, while Holly's outsized response to her wardrobe coordination with her colleague is inappropriate for 'wrong-kind' reasons, it is nevertheless a fitting response. After all, it is standardly allowed that failures of proportionality are failures of fit (e.g., D'Arms and Jacobson 2000a). But if it is uncontroversial to acknowledge SIZE as a case in which surprise is unfitting to the surprising, then philosophers must already be (at least sometimes, implicitly) thinking of the biconditionals as somehow qualified. In the next section, we consider what philosophers might say to explain why SIZE is compatible with a commitment to the fit–value biconditionals. Accordingly, we develop a second course of response to our counterexamples, which we call the *specification strategy*.

IV. The Specification Strategy

The idea that the size of a response bears on its fittingness is commonplace. For example, D'Arms and Jacobson note that one can "urge that an emotional response is unfitting because it is an overreaction. Thus your envy might be too large for the circumstances, if what you have is almost as good as your rival's" (2000a, 74). *Extreme* envy, they suggest, is not a fitting response to something that isn't *extremely* enviable. Similarly, Chris Howard writes: "The degree to which something is regrettable corresponds to the intensity of regret that it's fitting to feel in response to that thing" (Howard 2022, 3). If something is only a *little* regrettable, Howard suggests here, then the fitting response is only a *little* regret – regret that's little insofar as it's not that intense.

This intuitive idea about the relevance of proportionality to fit suggests what philosophers would say to explain why SIZE, as well as other cases of the same subtype, are not really counterexamples to the fit–value biconditionals. They would say that the *true* biconditionals don't hold merely between an evaluative property and valuing response of the same type, but rather between an evaluative property *of a certain type and degree* and a valuing response of *the same type and degree*:

> An object has an evaluative property of type P to degree D if and only if a valuing response of type P to degree D is fitting.

In SIZE, Holly has an *extreme* surprise response – falling out of her seat – to an event that is only *moderately* surprising – unplanned wardrobe coordination. Holly's surprise at this surprising event is unfitting, to be sure. But this case is no counterexample to the *true* biconditionals, because the

[14] For discussion of how hypocritical blame misrepresents, see Achs, manuscript.

true biconditionals hold that what is surprising to a particular degree covaries with fitting surprise to the *same* degree.

This solution to the case of SIZE suggests a general strategy for avoiding our counterexamples. It is to hold that the biconditionals are subject to counterexample only insofar as they are *underspecified*. The true covariances are not between, for instance, the properties of being loveable *in any way* and being fitting to love *in any way*, etc. Rather, the true covariances are between the properties of being loveable *in a specific way* and being fitting to love *in a correspondingly specific way*, etc. Once the biconditionals are thus qualified, moreover, our so-called "counterexamples" are disarmed, because the evaluative property/response pairs that feature in our cases are not the ones that feature in the *true* biconditionals with their more specific types of value and correspondingly more specific types of response. This is the *specification strategy* for replying to our counterexamples.

We have seen how the specification strategist handles cases like SIZE. She appears able to similarly handle cases like TONE: by holding that evaluative properties and valuing responses don't just come in a variety of degrees, but also in a variety of *tones*. For instance, it seems plausible that there are different tones of loveability (e.g., romantically loveable; loveable-as-to-a-mother) and, correspondingly, different tones of love (e.g., romantic love; maternal love). Thus, the specification strategist can hold that the *true* biconditionals are more specific not just with respect to degree but also with respect to tone:

> An object has an evaluative property of type P, to degree D, in tone T if and only if a valuing response of type P, to degree D, in tone T is fitting.

If this is the true content of the biconditionals, then Posy's maternal love of her romantically loveable partner may, indeed, be unfitting. But this poses no counterexample to the *true* biconditionals, which say that a valuing response *of a particular tone* is fitting to an object that possesses a corresponding evaluative property *of the same tone*.

That seems to take care of our quality cases. But the specification strategist looks able to handle positionality cases too, by holding that the true biconditionals are more specific not just with respect to their *degree* and *tone*, but also with respect to the *agents* whose responses are invoked and in relation to whom the values obtain. For instance, it seems plausible that some things might be disappointing only relative to some people and not others. So perhaps the *true* biconditionals say that objects which have evaluative properties *relative to particular agents* are fitting objects of corresponding valuing responses by and only by *those same agents*:

> An object has an evaluative property of type P, to degree D, in tone T, relative to A if and only if a valuing response of type P, to degree D, in tone T, by A is fitting.

Thus the specification strategist can hold that DISABLING RELATION is not a counterexample to the true biconditionals either, because Orchid's team's failure is not *disappointing relative to Laurel*.

And indeed, perhaps the true biconditionals feature value-properties and responses relativized not merely to *agents* but rather to *agents at times*. Then Dahlia's unfitting fear in TEMPORAL

ASYMMETRY (FORWARD) also needn't be a counterexample, since Russia's invasion of Ukraine *isn't* fearsome *relative to Dahlia when she fears it* – i.e. after it has already happened.

Consider, then, the specification strategist's template for the proper presentation of the fit/value biconditionals:

> SPECIFICATION
> An object has an evaluative property of type P, to degree D, in tone T, relative to A, at t_n if and only if a valuing response of type P, to degree D, in tone T, by A, at t_n is fitting.

The specification strategy appears to explain away cases where an object has an evaluative property of a certain type but is not a fitting object of a valuing response of the same type. However, this is not all. The specification strategy appears also to avoid a set of counterexamples that mirror the ones we presented in Section II: cases in which a response of a certain type seems fitting to an object, but the object seems to *lack* an evaluative property of the same type.

Consider, for example, PARTIALITY. In this case, it is preferable that the more suitable candidate, Basil, get the job, but not fitting for Lily to prefer that Basil get the job over her good friend, Iris. Now notice that, in this case, it also seems true that it would be fitting for Lily to prefer that Iris get the job although, ex-hypothesis, it is not preferable for Iris to get it. So it seems that Lily's preference is fitting but its object is not preferable: we have a fitting valuing response without a corresponding evaluative property.[15] Now consider TEMPORAL ASYMMETRY. We said that it is not fitting for Dahlia to fear Russia's fearsome invasion of Ukraine, since what she fears has already happened. This is a case where an evaluative property seems instantiated and yet the corresponding type of response seems unfitting. And here's a complementary case: Azalea has a severe peanut allergy, so it is fitting for her to fear peanuts. However, peanuts are not fearsome.[16] Again, a valuing response is fitting to an object that doesn't possess the corresponding type of evaluative property.

What these cases suggest is that the character we described at the outset, Magnolia, has a counterpart. While Magnolia has responses that correspond to their objects' evaluative properties and yet are possibly unfitting, her counterpart has fitting responses to objects that possibly lack the corresponding evaluative properties.

Qualifying the biconditionals according to SPECIFICATION seems to avoid *both* types of counterexample. It is not fitting for Lily to prefer that Basil get the job because Basil getting the job is not preferable *to her*; but it is fitting for Lily to prefer that Iris get the job because Iris getting the job *is* preferable *to her*. Similarly, it is not fitting for Dahlia to fear Russia's invasion because the invasion is not fearsome *to her*; but it is fitting for Azalea to fear the peanut because the peanut, while not fearsome to most, *is* fearsome *to her*. In this way, the specification strategy seems to

[15] This interpretation of the case coheres with the account of value offered by McHugh & Way 2022, according to which, roughly, an object is valuable if and only if it is fitting for a *neutral* agent to value it.

[16] We thank Knut Olav Skarsaune for this example.

disarm in one stroke both the *inflation of fitting responses* and the *inflation of evaluative properties*.

V. Failures of Specification

Let's take stock. Surprise can be unfitting to the surprising if it is disproportionate to its object's surprisingness. The specification strategist acknowledges as much. So she suggests that the *true* biconditionals don't just link value-property *types* to mirroring fitting response *types*, but rather link value-property *subtypes* to mirroring fitting response *subtypes*.

Note that this suggestion maintains a very intimate connection between value-properties and the fittingness of corresponding response types. In particular, making this suggestion allows the specification strategist to hold on to the idea that an object's value-properties, on their own, *entail* exactly what the fitting response to it is, and *vice versa*. Knowing merely that an object is surprising isn't yet enough to know how to respond fittingly to it, since knowing merely that something is surprising does not yet tell us what *type* of surprise it's fitting to have towards that object. But if we also know how surprising it is, and in what tone of surprisingness, and for whom, and when, *then* we know enough to draw an inference about the fitting response. Conversely, if we just learn *enough* about the type of response that's fitting to an object, then we can draw inferences – just given that – about its value properties.

However, holding that the true biconditionals follow the template of SPECIFICATION does not rescue the biconditionals from counterexample. For value-property subtypes and fitting response subtypes *don't* always mirror one another, as SPECIFICATION claims. For example, according to SPECIFICATION, if a joke is amusing relative to A at t_n, then it's fitting for A at t_n to be amused by that joke. But a joke can remain amusing for a person at a time, even if it is no longer fitting for her to be amused by the joke.

To see the point, first consider what it means for an object to possess an evaluative property relative to a person at a time. What it means, we submit, is for that object to possess that property from the perspective of that person at that time – for it to be true, from the point of view of that person at that time, that the object possesses the relevant property. If we're in New York, then Boston is north relative to us because, from our perspective, Boston is north. Analogously, an object is amusing relative to us at a time if and because, from our perspective at that time, it is true that the object is amusing.

Now consider the question of whether, even once she's been amused by Marigold's quip for a long while, that quip can still remain *amusing relative to Jasmine*. We submit that it can. After all, the facts about the quip that make it amusing from Jasmine's point of view might remain unchanged. The quip may still demonstrate excellent wit, or contain a boundary-pushing turn of phrase. Jasmine may continue to judge the quip amusing, or repeat it to others given its amusingness for her. All these factors indicate that the quip is still amusing from Jasmine's perspective. And yet, if she's already been amused by the quip for a long time, then Jasmine's amusement will cease to be fitting. Thus, Marigold's quip can be amusing relative to Jasmine at t_n even if it isn't fitting for Jasmine to be amused by that quip at t_n.

Indeed, there are many responses that can become unfitting for a person to have, even though the corresponding value relative to that person doesn't change. Continuing to regret a mistake often ceases to be fitting, even when that mistake remains regrettable to me. The facts that make it regrettable don't change, and I continue to guard against repeating that sort of mistake, warn others against making it, etc. In all these ways, the mistake remains *regrettable* from my perspective; it is simply that my mistake doesn't merit endless regret. Or consider the grief a person may feel at the loss of her brother. Perhaps that grief won't ever disappear completely, but it may fittingly diminish as she finds other ways to live with the loss and commemorate her brother. And yet, it is not as if her brother's death becomes any less *grievous to her*. The loss still matters to her, she still reflects on her brother often, and she finds different ways to express his significance to her. But even a permanently grievous loss such as this one need not merit perpetual, paralyzing grief.

There are also non-temporal positional cases where the agent-relativity of a value property and the agent-relativity of the fitting response don't track. According to SPECIFICATION, if someone is blameworthy relative to A, then it is fitting for A to blame that person. But someone can be blameworthy relative to A even if it *isn't* fitting for A to blame her. If Violet lacks the standing to blame Poppy, for instance, then it isn't fitting for her to blame Poppy. Yet Violet's lack of standing doesn't make Poppy any less *blameworthy relative to Violet*. It is, after all, still true from Violet's perspective that Poppy committed a morally horrible act without excuse. And it will still be appropriate for Violet to judge that Poppy is blameworthy, to try and avoid, or learn from, what Poppy has done, etc. It just isn't fitting for Violet to *blame* Poppy.[17]

The above cases show – *contra* SPECIFICATION – that it isn't *sufficient* for the subtype of a response to mirror the subtype of its object's value-property, in order for that response to be fitting. Even when there are various familiar ways in which a joke remains amusing to us, a mistake remains regrettable to us, and a person remains blameworthy to us, being amused by that joke, regretting that mistake, or blaming that person may not be fitting.

In addition, we doubt SPECIFICATION's implication that mirroring of value-property/response subtypes is *necessary* for a response's fittingness. According to SPECIFICATION, the *tone* of a response needs to be the same as the *tone* of value-property possessed by its object, in order for that response to be fitting. If Posy's partner is *romantically loveable*, for instance, then what he merits from her is *romantic love*; if he's *admirable-for-kindness* then he merits *admiration-for-kindness*. We should say that we're not sure it's possible to make sense of the idea that for each and every value-property/response pair, the value-property and response come in tones that are comparable for sameness. Or at least, we're not sure it is possible to make sense of this idea without understanding the tone of a value property *in terms* of the tone of the response that's fitting to it – which would be question-begging.[18] But, insofar as value-property/response pairs can be compared

[17] In speaking, as we do here, of "blameworthiness relative to A," we do not mean to deny that blameworthiness is an "agent-neutral property," in the sense that *who is asking* has no bearing on whether or not a person is blameworthy. One can speak of "blameworthiness relative to A" even if it's true that *if* S is blameworthy relative to A, then S is blameworthy relative to everyone else as well. While blameworthiness may well be agent-neutral, many other evaluative properties are likely not; it seems clear that x can be disappointing relative to us, for instance, *without* being disappointing relative to another person. In general, we think that some evaluative properties are agent-relative, and some are agent-neutral, and take no stand on which are which.

[18] We'll return, in the next section, to the thought that understanding specified value-properties in terms of specified fitting responses (or *vice versa*) would be question-begging.

for tones, it seems to us that responses of varying tones can each be fitting to an object that's valuable in a single tone.

Consider, for instance, the ways in which one's regret may fittingly change during a process of repair, from the moment one realizes one's mistake until after one has taken steps to avoid repeating it. Insofar as we try our best to think of regret and regrettability as coming in comparable tones, it seems to us that the tone of one's regret may fittingly change, even as the tone of one's mistake's regrettability remains the same. The mistake *itself*, after all, doesn't change in any way, so it's unclear why the *tone of its regrettability* would. And yet the type of regret that's fittingly felt about that mistake surely changes at different stages of the process, such that the *tone* of that fitting response may be said to change. So it seems false to us that, in order for regret of a mistake to be fitting, it must remain regret of a single tone – "regret of tone T" – in order to match the single-toned – "tone T" – regrettability of that mistake.

Similarly, perfect mirroring of the *size* of an object's value-property with the *size* of response to that object doesn't seem necessary for fittingness either. According to SPECIFICATION, in order for a response to be fitting, the *degree* of response must be the same as the *degree* to which its object is valuable; both response and value-property must be of "degree D." But it seems to us – again, insofar as we try our best to non-question-beggingly compare the size of a value-property and response – that valuing responses of *different* degrees can both be fitting to an object that's valuable to a *single* degree.

Consider, for instance, the difference in the levels of delight with which a person from the United States and a person from England might respond to eating a delicious meal. The meal needn't be any *more delightful* relative to the American as opposed to the Englishman. Indeed, we can imagine that these are men of identical gustatory tastes, eating the exact same meal, and that they both think it is the best meal they've ever had. Yet, we find it plausible that, given cross-Atlantic cultural differences, the size of *delight* with which it would be fitting for the American to respond – consisting in that delight's intensity, its duration, the types of expression he's inclined to give to it, etc. – is greater than the size of delight with which it would be fitting for the Englishman to respond. So greater and lesser degrees of delight can both be fitting responses to something delightful to a single degree. And the same seems true of surprise, amusement, regret, and most if not all other valuing responses with respect to their corresponding value-properties.[19] But if responses of *varying* degrees can all be fitting to an object that's valuable to a *single* degree, then it isn't true that fittingness requires a response's degree and the degree of its object's value to be the *same*, as SPECIFICATION claims.

Thus, in spite of its initial promise, we hold that SPECIFICATION isn't a good template for the biconditionals. The biconditionals can't be rescued by holding that they *really* concern the covariance of fitting response *subtypes* with the value-property *subtypes*, because the subtype of a fitting response need not mirror the subtype of its object's value-property.

[19] Jamie Fritz has also recently embraced the idea that responses fitting to an object valuable to a single degree can range in size (fc). And Barry Maguire writes: "Sometimes what is fitting is a range of attitudes or strengths of an attitude. For example, the fact that rush-hour traffic in Los Angeles is light today makes it fitting to feel more than mildly pleased but less than ecstatic" (2018, 792).

At this point, one might be tempted to try a disjunctive strategy. That is, one might insist that SPECIFICATION *does* give the true form of the biconditionals and try to hold that our (apparent) counterexamples to SPECIFICATION are wrong-kind reason cases. Yet this strategy won't work either. For one thing, our worries about SPECIFICATION with respect to *degree* and *tone* are worries about whether the matching of response subtype to value-property subtype is even *necessary* for fittingness. The examples we've adduced to make this worry vivid don't involve the intuition that a response is inappropriate to its object, and so there is nothing here for an appeal to wrong-kind reasons to explain away. Our examples are simply cases which suggest that the specification strategist's preferred rendering of the biconditionals is false. And for another, our temporal and positional counterexamples to SPECIFICATION were just reiterations of two of our original counterexamples: DIMINISHMENT and STANDING. But we've already argued (in Section III) that the wrong-kind-reason response doesn't work for these cases.

We draw the following lesson. The specification strategy fails because value-property subtypes and fitting response subtypes don't perfectly mirror one another. So, in order to properly describe the relation that holds between evaluative properties and etymologically connected responses, we must *let go* of the idea that an object's value properties, on their own, entail the precise response that's fitting to that object, and *vice versa*. The fitting response to an object *can't* simply be read off of its value-properties (nor *vice versa*). Below is what we take to be the correct template for the fit–value biconditionals.

> GENERALIZATION
> An object has an evaluative property of type P, to *some* degree, in *some* tone, relative to *some* agent, at *some* time if and only if a valuing response of type P, to *some* degree, in *some* tone, by *some* agent, at *some* time, is fitting.

We call this template "GENERALIZATION" because, instead of entailing that an object's value-properties, on their own, determine what particular responses are fitting to it, it entails merely that an object's value-type determines that there is *generally some* response of a matching type that's fitting to it (and *vice versa*) – an existential generalization. From the fact that an object is fearsome, for instance, we can infer that *some* type of fear is fitting to it. And from the fact that fear fits an object, we can infer that it's fearsome in *some* way. But in order to know *which* type of fear is fitting to it (or in *which* ways it's fearsome), we may have to look beyond the object's value-properties (or beyond the responses that are fitting to it). We can't, in other words, know *whose* fear is fitting, or *when*, or in what *tone*, or *degree*, from attending to an object's fearsomeness alone.

VI. A Rejoinder?

We hold that the specification strategy fails because value-property subtypes and fitting response subtypes don't always mirror one another. However, some readers may find our counterexamples to SPECIFICATION unconvincing. We say that someone can be blameworthy relative to a person but not the fitting object of that person's blame and that a joke can remain amusing for a person even if that person's amusement is no longer fitting. To motivate such claims, we appeal to substantive notions about what facts make someone blameworthy or something amusing for a person and then note that these facts do not always entail that person's fitting blame or amusement.

Yet a possible rejoinder is that our argument betrays a misunderstanding of what evaluative properties are.

No matter whether a wrongdoer is guilty or a joke is witty, one might protest, what it *means* for someone to be blameworthy or for something to be amusing is for that person to be fittingly blamed or for that thing to be a fitting object of amusement. Accordingly, what it means for something to be *amusing relative to A at t* is for that thing to be the *fitting object of amusement by A at t.* Amusingness and the fitting object of amusement pick out one and the same property because they are one and the same *concept*. So surely, the objector submits, the only natural way to understand what it is to be amusing *for* a person – amusing from that person's point of view – is as being the fitting object of amusement *by that person*. Moreover, if this is the way we conceive of agent-relative amusingness, then it immediately follows that a joke cannot remain amusing to a person at a time when it is no longer fitting for that person to be amused by that joke. The same is true of blameworthiness, according to the objector. If a person is blameworthy for me, then it follows from what blameworthiness means that that person is, for me, the fitting object of blame – i.e. the fitting object of *my* blame.

In fact, all of our counterexamples to SPECIFICATION can be done away with if we just conceive of each specified evaluative property *in terms of* the specific valuing response that's fitting to it: if we understand whether an object has an evaluative property of type P, to degree D, in tone T, relative to A, at t_n *as* whether a valuing response of type P, to degree D, in tone T, by A, at t_n fits that object. Moreover, it seems that understanding specified evaluative properties in terms of the fittingness of specific responses *guarantees* no counterexamples to SPECIFICATION. For, on this way of thinking, if a response is not fitting by a certain agent, at a certain time, in a certain tone, or to a certain degree, then the corresponding evaluative property cannot be instantiated.

The view we are envisaging has recently been expressed by Chris Howard, who suggests that, just as being a bachelor is equivalent to being an unmarried male, being fitting to regret is equivalent to being regrettable. If a bachelor can be married or if there is an unmarried male that is not a bachelor, Howard says, then he loses any sense of what bachelors and unmarried males are. Similarly, if something can be fitting to regret without being regrettable, or regrettable without being fitting to regret, Howard confesses, then he loses any sense of what being regrettable and being fitting to regret are (Howard 2022, 12).

We call this way of understanding evaluative properties *formalism* because, instead of beginning with a substantive conception of the regrettable, the amusing, etc., and then thinking about specific evaluative properties in terms of these substantive conceptions of value, the formalist understands the blameworthy, the regrettable, and the amusing as *conceptually identical* with what is fittingly blamed, regretted, or is the fitting object of amusement, and then understands specified evaluative properties *in terms of* whether specified valuing responses are fitting.

Formalism has attractions. It may seem a natural view of the relation between evaluative properties and valuing responses, and it effectively preempts our counterexamples to SPECIFICATION. However, formalism should be rejected. First, a dialectical argument: formalism cannot be appealed to in order to disarm our counterexamples. Though our apparent counterexamples lose all force if formalism is true, if our apparent counterexamples have force they cast doubt on the truth of formalism. So which is it: *modus ponens* or *modus tollens*?

The answer is clear when we recall that what we are doing is investigating how evaluative properties and valuing responses are related. How do the blameworthy relate to blame, the regrettable to regret, and the surprising to surprise? It is not enough, it is often said, that an object is *usually* blamed, regretted, or is the object of surprise for it to bear the corresponding type of evaluative property. An object can be blameworthy, regrettable, or surprising even if it is rarely blamed, regretted, or the object of surprise. It is also not enough, it is said, that an object is something we're *disposed* to blame, regret, or be surprised by. For an object can be blameworthy, regrettable, or surprising even if no one is disposed to blame it, regret it, or be surprised by it. Thus, it is suggested, an object is blameworthy, regrettable, or surprising when blame, regret, and surprise are *fitting* to it.[20] Note that, in each of these steps, it is assumed that the relation between evaluative properties and valuing responses is to be revealed by a careful study of examples. Being blameworthy is not the same as being the usual object of blame because there are cases where the two come apart; the same is true of being blameworthy and being what we are disposed to blame. In raising our counterexamples, we continue the very same inquiry. We describe cases that suggest that an object can be blameworthy while blame is not fitting to it, regrettable while regret is not fitting to it, surprising while surprise is not fitting to it, etc.

At this stage, formalism cannot be asserted in order to avoid our counterexamples, because the counterexamples have been provided in the course of an assessment of what the relation between evaluative properties and corresponding response types is. Whereas formalism just *assumes* a way in which value-properties and responses are related. Similarly, the dispositionalist cannot reject cases that suggest that a dispositionalist analysis of evaluative properties is false by asserting the truth of dispositionalism. Our question is how evaluative properties and corresponding responses are related. To assume formalism – to think about an object's possession of an evaluative property *in terms* of whether a corresponding response is fitting to it – simply begs that question. However, if there were an independent argument in favor of formalism, the formalist might legitimately disarm our counterexamples after all.

This brings us to our second objection to formalism: it is false. Consider the following argument. If formalism is true, the relation between evaluative properties and valuing responses is like the relation between being a bachelor and being an unmarried male. To determine whether *all* bachelors are unmarried males, we need only to properly understand these concepts. Once we understand being a bachelor *as* being an unmarried male, there is no need to consider exceptions because these concepts are identical. So their coextension is trivial. If formalism is true, the relation between, say, being regrettable and being the fitting object of regret is similarly trivial. This, we take it, is Howard's point in saying that if being regrettable and being the fitting object of regret come apart, then he loses his very grip on these notions.

However, the analogy between the two pairs of concepts – bachelors and unmarried males, and evaluative properties and fitting valuing responses – breaks down. The very fact that we can engage in a serious philosophical discussion of the relation between evaluative properties and fitting valuing responses indicates that the relation is not obvious to those who possess these concepts. It makes little sense to ask: which unmarried males are bachelors, where, and when? The answer is all of them, everywhere, always – simply because being one *means* being the other. By contrast,

[20] For examples of such reasoning, see D'Arms and Jacobson 2000a; Howard 2018; Shoemaker 2017; and Tappolet 2016, 83.

we can sensibly ask: when, how, and by whom is it fitting to regret a regrettable object? From the fact that an object is regrettable it does not follow trivially that it is a fitting object of regret by everyone, in every way, always. Even if it were, indeed, true that it is fitting for everyone, in every way, always, to regret a regrettable object, this fact would be a substantive one, not an upshot of the conceptual identity of being regrettable and being fitting to regret. The same is true of other fit–value pairs. We can ask when, how, and by whom is it fitting to desire a desirable object, praise a praiseworthy object, condemn a condemnable object, etc. This shows that none of these pairs are pairs of identical concepts. In short, the fact that an object bears an evaluative property of a certain type leaves *conceptually open questions* about the fitting response to it. So formalism is false.

The objector might concede that there is a substantive question about the relation between evaluative properties and fitting valuing responses about which competent users of these concepts can disagree. And yet the objector might insist that the answer to this substantive question is that facts about fitting responses mirror facts about evaluative properties. If something is regrettable *simpliciter*, then it's fitting for everyone to regret, in any way, always. And if something is regrettable to someone in a certain way, to a certain degree, at a certain time, then it is fitting for that person to regret it in that way, to that degree, at that time. So even if formalism is false, SPECIFICATION is true. However, we've produced counterexamples to this claim. A mistake that's singularly regrettable can merit different degrees and tones of regret. A joke can be amusing *to an agent* without meriting amusement *from him*. Moreover, having rejected the formalist conception of evaluative properties, there seems to be no reason to *assume* that mirroring of specified properties and fitting responses is true across the board. Indeed, we suspect that the only reason such an assumption is tempting is that the biconditionals have become entrenched in the philosophical literature without being properly qualified.

GENERALIZATION is the template for the biconditionals that actually coheres with cases. According to GENERALIZATION, the fact an object bears an evaluative property of a certain type, in relation to certain agents, at a certain time, of a certain quality, only determines that *some* specific response of the same general type is fitting; it does not determine the full specification of the fitting response. Therefore, the fact that an object bears an evaluative property of a certain type leaves *metaphysically open questions* about the fitting response to it.

VII. Upshots

The fit–value biconditionals correctly maintain that fitting responses are systematically related to their objects' value-properties. Yet, fitting responses and evaluative properties do not stand and fall together. In particular, an object's value-properties do not, on their own, determine the fitting response to it. Instead, various facts – such as, *who* is having a valuing response, *when*, and *how* they're having it – can bear on whether a response is fitting *without* bearing on whether its object possesses the matching type of evaluative property. This result has important implications for attempts to analyze value-properties in terms of other normative notions. In conclusion, we consider these upshots.

First, not all fit-related reasons are of the "right kind" to figure in an analysis of value in terms of reasons.[21] In other words, there are wrong-kind, fit-related reasons. That Violet lacks standing to blame Poppy, for instance, is a fit-related reason for Violet not to blame her. Yet Violet's lack of standing has no bearing on Poppy's blameworthiness, so it is the wrong kind of reason to figure in an analysis of blameworthiness in terms of sufficient reason to blame. Poppy may be blameworthy, and yet there may *not* be sufficient reason to blame her, given a lack of standing. So fit-related reasons and right-kind reasons come apart – although they're often treated as interchangeable.[22, 23]

Moreover, since some facts bear on fit but not on value, fitting-attitude analyses of value face their own version of the wrong-kind-reason problem. Fitting-attitude analyses hold, roughly, that *x* is valuable if and only if, and because *x* is fitting to value. But since there are facts that bear on an attitude's fit without bearing on the value of its object, fitting-attitude analyses must answer the question: Of the facts that determine fit, which are of the right kind to factor into our analysis of value? Marigold's quip may be amusing even if it's *not* fitting for Jasmine to be amused, given that she's already been amused for a while. So the fact of Jasmine's previous amusement must be *disregarded* in order for a fitting-attitude analysis of the amusing to come out true. It is a wrong-kind, fit-affecting fact.

Despite these newfound difficulties, our discussion also sheds light on which fit-affecting facts are relevant to a plausible account of value. An interesting feature of our counterexamples to the unqualified biconditionals is that they are all arguably interpreted as cases in which *background conditions* on the fittingness of our subjects' attitudes aren't met. For instance, it seems plausible that Violet's having standing to blame Poppy wouldn't itself *make* her blame of Poppy fitting, but would rather *enable* Poppy's wrongdoing to make her blame fitting (Achs 2022). Similarly, that Russia's invasion hasn't yet happened plausibly doesn't *make* Dahlia's fear of it fitting, but rather enables, say, the harm it causes to innocents to make fearing it fitting. So it seems that when a response fails to be fitting despite a matching value-property in its object, the explanation is found in a failure of background conditions. This fact might be used to systematically distinguish those fit-affecting facts which bear on value from those that don't. We might say that the "right-kind" of fit-affecting facts are the fit- (or unfit-) *makers*; whereas the wrong kind of fit-affecting facts are facts that constitute fit-affecting background conditions. When the fit-making facts of

[21] Scanlon's (1998, 95-100) "buck-passing" analysis of value in terms of reasons faces the well-known "wrong-kind-reason" problem: the problem that (mere) value-related reasons seem to be the wrong kind of reason to plug in, if we want the statement that "What it is for *x* to be valuable is for there to be sufficient reason to value x" to come out as true. For an overview of this problem, see Gertken and Kiesewetter 2017.

[22] For an example of treating these notions interchangeably, see Schroder 2010, 27–28.

[23] An objection: that Violet lacks the standing to blame is not a *reason* for her not to blame Poppy; rather, it's the absence of an *enabler* of a reason to blame Poppy. So standing does not pose a counterexample to the claim that all fit-related *reasons* are right-kind *reasons*. Our reply: (1) we think an enabler, or the absence of enabler, can function as a reason, at least in the sense that it can be the fact in light of which one entertains or relinquishes an attitude, and (2) even if enablers are not, technically, reasons, our main point holds – namely, that there are fit-*affecting* facts which render buck-passing analyses of value false. We think it is generally assumed that the only facts which render these sorts of analyses false are mere value-related ones. That said, we'll return to the idea that distinguishing between fit-*makers* and fit-*enablers* might be helpful for analyses of value, in just a moment.

amusement obtain, the object of amusement is amusing even if amusement is not fitting due to background conditions.

Saying as much, we hasten to add, does not, itself, provide any aid in distinguishing the fit-*making* facts from those which merely constitute the relevant background conditions. And while, in some cases, it seems easy enough to judge, in other cases it's hard to tell. Moreover, proponents of the fitting-attitude analysis must offer a way to draw this distinction if they want to call upon it to rescue their analysis. So the thesis of this paper raises a serious problem for fitting-attitude analyses of value, but it also points in the general direction of a solution.

One might wonder whether the idea that fit-*makers* are the fit-affecting facts which bear on value couldn't, itself, be incorporated into a template for the biconditionals. Perhaps, for instance:

> CETERIS PARIBUS
> An object has an evaluative property of type P if and only if a valuing response of type P is fitting, *ceteris paribus*.[24]

For the purposes of this schema, the *ceteris paribus* clause must be given a certain interpretation, one on which all else being equal essentially means that all the background conditions required for the fittingness of the relevant valuing response have been satisfied. The fearsome is fitting to fear *ceteris paribus*, in the sense that it's fitting to fear when the relevant background conditions have been met – e.g., having relevant concerns, being in relevant proximity to the object. Similarly, pride at the pride-worthy is fitting *ceteris paribus*, in the sense that it's fitting to take pride in when certain background conditions have been met, such as bearing a close relationship to the potential object of pride. The idea here is that the satisfaction of relevant background conditions is something we take for granted during an episode of valuing (or in a situation which seems apt for one) – unlike the properties of our attitude's object on which we actively focus. We take it for granted, for instance, during an episode of fear, that we have relevant concerns and are in relevant proximity to the object, even insofar as what we focus on are the object's threatening properties – those features that make it fitting to fear. And so relevant background conditions being satisfied counts as "all else being equal," because the satisfaction of those conditions means that assumptions we are making, at relevant moments, are true.

Albeit less precise, we suspect that CETERIS PARIBUS, like GENERALIZATION, is a true description of the relation that holds between evaluative properties and the fittingness of etymologically corresponding response types. In some ways, it is less informative than GENERALIZATION. GENERALIZATION captures the fact that, if a certain type of response is fitting by *some* person (at *some* time, in *some* way) then it is entailed that its object possesses the corresponding type of evaluative property relative to *someone* (at *some* time, in *some* way). For instance, it captures the fact that if, because of her allergy, it is fitting for Azalia to fear peanuts, then peanuts are fearsome relative to *someone* – even if not fearsome *simplicter*. CETERIS PARIBUS doesn't capture this sort of entailment.

[24] Lucy Campbell and Keshav Singh both suggested this template to us. Lucy Campbell also suggested replacing the material biconditional with a subjunctive one ("x is valuable iff it *would be* fitting to value") – an idea we also would have loved to explore, given more space.

But CETERIS PARIBUS is also in some ways more informative than GENERALIZATION. For – provided one keeps in mind the intended interpretation of the *ceteris paribus* clause – CETERIS PARIBUS captures the idea that, *by default*, blame is fitting to the blameworthy, fear is fitting to the fearsome, etc. It's fitting to fear the fearsome by default, not because something fearsome need be something it is fitting for *most* people to fear. It needn't. Rather, it's fitting to fear the fearsome by default because, as long as the background conditions of fitting fear are in place, fearing the fearsome is fitting. Similarly, it is fitting, by default, to blame the blameworthy, because, as long as the background conditions of fitting blame are in place, blaming the blameworthy is fitting. Moreover, that it is fitting to blame the blameworthy, fear the fearsome, etc. *by default*, explains why there is an etymological connection between words for value-properties and words for response types – and, indeed, why the unqualified biconditionals at first seem right. This is a potential attraction of CETERIS PARIBUS.

We ourselves are happy with using either GENERALIZATION or CETERIS PARIBUS to describe the fit–value biconditionals. They are compatible and we suspect both are true. Nevertheless, we would urge a final reminder on any who prefer CETERIS PARIBUS. Please remember that all else *isn't* always equal. There are many factors that can affect whether a valuing response is fitting and do so without making any difference to its object's value. So we cannot replace questions about fitting responses with questions about value-properties without glossing over the many nuances of fit.[25]

[25] We are enormously grateful for the excellent feedback and suggestions we received on earlier drafts of this article. For written comments, discussion, or both, we thank Abdul Ansari, Selim Berker, Lucy Campbell, Justin D'Arms, Sandy, Diehl, Sam Dishaw, Lidal Dror, Alexander Greenberg, Chris Howard, Roberto Keller, Claire Kirwin, Indrik Reiland, Keshav Singh, Knut Olav Skarsaune, Jonathan Way and members of John Hyman's *Roots of Responsibility* work-in-progress group. Thanks also to audiences at the Attitudinal Normativity Workshop in Geneva, the Chapel Hill Normativity Workshop, the S. Louis Annual Conference on Reasons and Rationality, and the Slippery Slope Normativity Summit. Rachel Achs' research on this work was also supported by an ERC advanced grant project, *Roots of Responsibility: Metaphysics, Humanity, and Society*. This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement no. 789270).

REFERENCES

Achs, R. (2022). Blame's commitment to its own fittingness. In C. Howard and R. A Rowland (eds.) *Fittingness: Essays in the Philosophy of Normativity*. Oxford: Oxford University Press.

Achs, R. (Manuscript). Hypocritical Blame Is Unfitting.

Bader, R. (2016). Conditions, modifiers, and holism. In B. Maguire and E. Lord (eds.), *Weighing reasons* (pp. 27–55). Oxford: Oxford University Press.

Bell, M. (2013). The standing to blame: A critique. In J. Coates and N. Tognazzini (eds.), *Blame: Its Nature and Norms* (pp. 263–281). Oxford: Oxford University Press.

Berker, S. (2022). The deontic, the evaluative, and the fitting. In C. Howard and R. A. Rowland (eds.), *Fittingness: Essays in the Philosophy of Normativity.* Oxford: Oxford University Press.

Brandt, R. (1946). Moral valuation. *Ethics* 56: 106–21.

Chappell, R. Y. (2012). Fittingness: The sole normative primitive. *The Philosophical Quarterly* 62 (159): 684–704.

Dancy, J. (2004). *Ethics without Principles*. Oxford: Oxford University Press.

D'Arms, J. & Jacobson, D. (2000a). The moralistic fallacy: On the 'appropriateness' of emotions. *Philosophy and Phenomenological Research* 61: 65–90.

D'Arms, J. & Jacobson, D. (2000b). Sentiment and value. *Ethics* 11: 722–748.

Fritz, J. (2021). Fitting anxiety and prudent anxiety. *Synthese* 199: 8555–8578. https://doi.org/10.1007/s11229-021-03175-8

Fritz, J. (fc). Unfitting absent emotion. *Oxford Studies in Metaethics*.

Gertken, J., & Kiesewetter, B. (2017). The right and the wrong kind of reasons. *Philosophy Compass* 12(5): e12412

Hieronymi, P. (2005) The wrong kind of reason. *Journal of Philosophy* 102: 436–457.

Howard, C. (2018) Fittingness. *Philosophy Compass* 13: e12542. https://onlinelibrary.wiley.com/doi/10.1111/phc3.12542

Howard, C. (2019a). Fitting love and reasons for loving. In M. Timmons (ed.), *Oxford Studies in Normative Ethics Volume 9* (pp. 116–137). Oxford: Oxford University Press.

Howard, C. (2019b). The fundamentality of fit. In R. Shafer-Landau (ed.) *Oxford Studies in Metaethics,* vol. 14 (pp. 217–236). Oxford: Oxford University Press.

Howard, C. (2022). Forever fitting feelings. *Philosophy and Phenomenological Research*: 1–19.

Howard, C. & Rowland, R. A. (2022). Fittingness: A user's guide. In C. Howard and R. A. Rowland (eds.) *Fittingness*. Oxford University Press.

King, M. (2019). Skepticism about the standing to blame. In D. Shoemaker (ed.) *Oxford Studies in Agency and Responsibility* Vol. 6 (pp. 265–288). Oxford: Oxford University Press.

Maguire, B. (2018). There are no reasons for affective attitudes. *Mind* 127(507): 779–805.

McHugh, C. & Way, J. (2016). Fittingness first. *Ethics* 126(3): 575–606.

McHugh, C. & Way, J. (2022). Value and idiosyncratic fitting attitudes. In C. Howard & R. A. Rowland (eds.), *Fittingness: Essays in the Philosophy of Normativity*. Oxford: Oxford University Press.

Miller, K. (fc) Tensed facts and the fittingness of our attitudes. *Philosophical Perspectives*.

Moller, D. (2017). Love and the rationality of grief. In C. Grau and A. Smuts (eds.), *The Oxford Handbook of Philosophy of Love*. Oxford: Oxford University Press.

Na'aman, O. (2021). The rationality of emotional change: toward a process view. *Noûs* 55(2): 245–269.

Olson, J. (2009). Fitting attitude analyses of value and the partiality challenge. *Ethical Theory and Moral Practice* Vol. 12: 365–378.

Pearson, O. (2018). Appropriate emotions and the metaphysics of time. *Philosophical Studies* 175: 1945–1961.

Prior, A.N. (1959). Thank goodness that's over. *Philosophy* 34: 12–17.

Scanlon, T. M. (1998). *What We Owe to Each Other*. Cambridge, MA: Harvard University Press.

Scarantino, A. & de Sousa, R. (2018). Emotion. *The Stanford Encyclopedia of Philosophy*, Winter 2018 Edition, E. N. Zalta (ed.), URL = https://plato.stanford.edu/archives/win2018/entries/emotion/

Schroeder, M. (2010). Value and the right kind of reason. In Shafer-Landau, R. (ed.) *Oxford Studies in Metaethics*, vol. 5 (pp. 25–55). Oxford: Oxford University Press.

Schroeder, M. (2012). The ubiquity of state-given reasons. *Ethics* 122(3): 457–488.

Shoemaker, D. (2017). Response dependent responsibility; or, a funny thing happened on the way to blame. *Philosophical Review* 126: 481–527.

Tappolet, C. (2016). *Emotions, Values, and Agency*. Oxford: Oxford University Press.

Telech, D. & Dahan Katz, L. (2022). Condemnatory disappointment. *Ethics* vol. 132: 851–880.

Wallace, R. J. (2019). Trust, anger, resentment, forgiveness: On blame and its reasons. *European Journal of Philosophy* Vol. 27: 537–551.

Way, J. (2012) Transmission and the wrong kind of reason. *Ethics* 122: 489–515.

Wiggins, D. (1987). A sensible subjectivism? In his *Needs, Values, Truth: Essays in the Philosophy of Value*. Oxford: Blackwell.

Yehezkel, G. (2014). Theories of time and the asymmetry in human attitudes. *Ratio* 27: 68–83.