Chapter 8.

Metaethics: Goodness

In Chap. 5, we saw how Brentano develops a metaonotlogy based on his account of judgment. In this chapter, we will see that he develops a structurally analogous metaethic on the basis of his account of interest phenomena (will, emotion, and pleasure/pain). The chapter presents, develops, and up to a point defends Brentano's account. In §1, I lay out a logical geography within which Brentano's theory can be usefully placed. In §2, I present Brentano's theory. §3 is dedicated to Brentano's *case* for his theory. In §4, I discuss a number of objections, both to the theory and to the argument for it – among them one which I do find devastating to Brentano. This will lead us to discuss Brentano's theory of *beauty* in §5.

1. Introduction: Theories of Value

The organizing question of metaethics, recall from Chap. 5, is 'What does it mean to say that something is good?' Answers to this question can be classified in any number of ways, but there are two deep divisions among such answers that run throughout the history of philosophy. The first is the division between naturalist and nonnaturalist answers; the second is between what I will call 'object-end' and 'subject-end' answers.[1] This section introduces these two distinctions.

At the broadest level, naturalism and nonnaturalism are distinguished by their take on the relation between fact and value. Nonnaturalists hold that there is a *dichotomy* or *categorical difference* between the two realms: matters of value are *irreducible* to matters of fact, and nothing in the latter can *account for* the former.

Naturalists deny this, trying to reductively explain the domain of value in terms of the domain of facts.

It is far from obvious how to formulate such impressionistic characterizations as precise theses. To a first approximation, we may think of nonnaturalism as the following thesis: There are normative truths whose truthmakers have constituents which cannot be accounted for naturalistically. Thus, 'Genocide is wrong' is strictly true, and has as its truthmaker the state of affairs of genocide having the property of being wrong, where the property of being wrong cannot be accounted for by anything within the 'natural' or 'empirical' realm. This is still quite vague, but note that it allows for two distinct ways of being a naturalist: either by denying that there are normative truths, or by allowing such truths but insisting that their truthmakers are fully explicable in naturalistic terms. Both routes have been taken: 'expressivists' (Ayer 1936 Chap. 6, Gibbard 1990), for example, deny the existence of normative truths, while 'Cornell realists' (Boyd 1988, Brink 1989) deny that normative truthmakers have naturalistically recalcitrant constituents.

Two questions still press: (i) what makes a truth 'normative'?; (ii) what makes an entity 'natural(istic)'? To bypass the first question, I propose that we simply focus on putative truths about goodness, which would be normative if anything is. As for the second question, three main strands can be discerned in discussions of (non)naturalism. All three are problematic as accounts of the *nature* of the natural, but make for reasonably reliable *symptoms* of naturalness.

The first and most dominant approach is to construe the natural as that which is invoked in the natural sciences – today's natural sciences in one version, the natural sciences at the 'end of inquiry' in another (Moore 1903a: 92). This leads to the following version of nonnaturalism:

(NN$_1$) There is a statement S of the form '*x* is good,' such that (i) S is true and (ii) at least one constituent of S's truthmaker is neither itself invoked in, nor

reductively explicable in terms of entities invoked in, (end-of-inquiry) natural science.

The idea is that (end-of-inquiry) natural science will not mention goodness, and what it will mention will not be sufficient to reductively explain goodness.

The main problem with NN_1, as a formulation of nonnaturalism, is the following dilemma. Appealing to *today's* natural sciences probably rules out very natural phenomena that are yet to be recognized in our (presumably still incomplete) science. If we appeal instead to end-of-inquiry natural science, the ensuing thesis is somewhat vacuous, since we have no idea what might show up therein and how continuous it might be with our current picture of nature. Signing up on 'naturalism' without any clear grasp of what type of thing 'natural' might refer to only gives the *illusion* of having asserted something. One has no real sense of what one has committed to.[2]

A second strand in discussions of naturalism construes the natural as the spatiotemporal, or the causally-efficacious spatiotemporal (Armstrong 1997, Kim 2003). This suggests the following formulation of nonnaturalism:

(NN_2) There is a statement S of the form '*x* is good,' such that (i) S is true and (ii) at least one constituent of S's truthmaker is neither (causally-efficaciously) spatiotemporal nor reductively explicable in terms of (causally efficacious) spatiotemporal entities.[3]

Here the idea is that nature is a spatiotemporal system of causally interacting entities, and something is 'natural' just when it is suitably embedded in that system.

One problem here is that, intuitively, we take such things as ghosts and witches to be naturalistically unacceptable, yet *if* they existed they would be very much spatiotemporal and causally efficacious. A ghost might live in the castle attic (spatial) for two centuries (temporal) and make candlelights flicker (causal). A witch would live in Salem for a hundred years and would incite adultery. Yet part of

what we want in wielding naturalism is to produce an interdiction on appeal to ghosts and witches in explaining flickers and betrayals.

A third approach construes the natural as that which can be empirically observed (Copp 2003). The resulting nonnaturalist thesis is:

(NN₃) There is a statement S of the form '*x* is good,' such that (i) S is true and (ii) at least one constituent of S's truthmaker is neither empirically observable nor reductively explicable in terms of empirically observable phenomena.

The goodness of generosity, on this view, is neither observable itself nor reductively explicable in terms of what is.

The fundamental problem here is that NN₃ provides an epistemological characterization of an ontological notion. Furthermore, the characterization seems too narrow: leptons too are neither observable (because too small) nor reductively explicable in terms of what is observable (because fundamental), yet are clearly natural phenomena.

All three characterizations face real difficulties, then.[4] Nonetheless, all three cite features that tend to be symptomatic of natural entities. Certainly an entity which were invoked both in current and end-of-inquiry science, were spatiotemporal and efficacious, and were empirically observable would qualify as natural. Thus we may take these three characterizations collectively as useful *guides* to 'the natural.'

<center>❧❧</center>

We say that genocide is bad, but also that approval of genocide is bad; and that generosity is good, but also that approval of generosity is good. Such evaluations seem to go hand in hand: generosity is good iff generosity-approval is; genocide-approval is bad iff genocide is. (Also: generosity is good iff generosity-disapproval is bad, and genocide is bad iff genocide-disapproval is good.) This raises a Euthyphro-

<center>4</center>

style question: is generosity-approval good because generosity is, or is generosity good, ultimately, precisely because approval of it is? Some ethical theories ground the value of generosity-approval in the value of generosity, others ground the value of generosity in that of its approval. Let us say that the former offer an *object-end* answer to our Euthyphro question, the latter a *subject-end* answer.[5]

The distinction between object-end and subject-end views applies both to naturalist and to nonnaturalist theories of value. The aforementioned Cornell realists, for example, are object-end naturalists: they ground the goodness of generosity-approval in the goodness of generosity, then the goodness of generosity in functional properties that acts of generosity exhibit. Meanwhile, some (though not all) response-dependent theorists of value are subject-end naturalists: they ground the goodness of generosity in the goodness of generosity-approval, then ground the goodness of generosity-approval in some purely psychological properties of the approval (Lewis 1889, Gert 2009).

Twentieth-century *non*naturalism has been dominated by the object-end approach expounded by Moore in *Principia Ethica* (Moore 1903a). This tradition has continued to be influential in twenty-first-century nonnaturalist thinking (Shafer-Landau 2003, Enoch 2011). But twenty-first-century metaethics has also seen an impressive rise in *fitting attitude* accounts of value (D'Arms & Jacobson 2000, Zimmerman 2001, Danielsson & Olson 2007, Way 2012, Chappell 2012, Kauppinen 2014, McHugh & Way 2016), which are paradigmatically subject-end.[6] We may formulate the fitting-attitude account of goodness as follows:

(FA$_1$)   For any good g, (i) it is fitting to take a pro attitude toward g, and (ii) g is good *because* (i).[7]

Some (though not all) fitting-attitude theorists are subject-end *non*naturalists: they ground the goodness of generosity in the fittingness of generosity-approval, but then deny that the fittingness of generosity-approval can be accounted for in any nonnormative terms.

The distinctions between naturalist and nonnaturalist theories of value, on the one hand, and object-and subject-end theories, on the other, produce a matrix of four positions on the nature of value: object-end naturalism, subject-end naturalism, object-end nonnaturalism, and subject-end nonnaturalism. Where does Brentano's theory of value fit in this matrix?

2. Brentano's Subject-end A-naturalism

Brentano does not comment on whether he is a naturalist or nonnaturalist. Furthermore, whether he is best interpreted as a naturalist or as a nonnaturalist depends on what symptom of naturalness we focus on. It is beyond question, however, that Brentano took a subject-end rather than object-end approach to value. Accordingly, I start with a presentation of the subject-end character of Brentano's account (§2.1), which I then fill in with some details (§2.2); I then take up the issue of naturalism (§2.3).

2.1. Brentano as Fitting-Attitude Theorist

Brentano states his preference for the subject-end approach in many places. Here is one clear statement from an 1889 lecture before the Vienna Legal Society, a lecture which then became the basis for his main metaethical work, *The Origin of Our Knowledge of Right and Wrong*:[8]

> We call something 'good' when the love concerned with it is correct/fitting (*richtig*). In the broadest sense of the word, the good is that which can be loved with a correct/fitting love – that which is love-worthy (*Liebwerte*). (Brentano 1889: 19 [18])

Here is another, from a separate 1889 lecture:

> … everything thinkable divides into two classes: one containing everything for which love is fitting (*passend*), the other everything for which hate is. Whatever belongs in the first class we call good, whatever is included in the other we call bad. (Brentano 1930: 25 [21-2])

And a third, from the 1876-94 lecture notes on practical philosophy:

> … when we call certain objects good, and others bad, we are saying thereby nothing more than that whoever loves the former and hates the latter has taken the right stand (*verhalte sich richtig*). (Brentano 1952: 144 [90])

Two terminological reminders are in order. First, as we saw in Chap. 7, Brentano uses the terms 'love' and 'hate' to cover all pro and con attitudes (respectively). Secondly, Brentano's *richtig* is not supposed to mean anything like 'true' or 'veridical.' As noted in Chap. 5, in one footnote he offers *konvenient*, *passend*, and *entsprechend* as interchangeable with *richtig* (Brentano 1889: 60 [74]). This makes clear that what he has in mind is something like 'appropriate,' 'fitting,' or 'suitable.' When we take these two terminological clarifications into account, we see that Brentano is a straightforward fitting-attitude theorist: he grounds the goodness of any *x* in the fittingness of pro attitudes toward *x*.

Indeed, Brentano is plausibly the *first* fitting-attitude (FA) theorist. It is sometimes said that Sidgwick floated FA before Brentano's *Origin*, when he wrote that ultimate or final good is 'that of which we should desire the existence if our desire were in harmony with reason' (Sidgwick 1884: 108). But first, it is not clear that the 'should' in Sidgwick's formulation is normative; it appears to be rather the past subjunctive of 'shall.' And in any case, Brentano actually asserts his FA already in his 1866 habilitation (aged 28!):

> … we call something good insofar as it is desire-worthy (*begehrenswert*). (Brentano 1866: 141; see also 1874: II, 100 [247])

Thus Brentano has a good claim to the status of first fitting-attitude theorist.

It might be objected that Brentano could not possibly be a *subject*-end theorist, since in some places he stresses the independence of goodness from our mental life. Consider this striking passage:

> The objective condition for any [pro attitude] being potentially correct/fitting lies in the goodness of the object at which it is directed. This says nothing, however, as to whether there is anyone who could harbor such a [pro attitude]. The analogy with the true in the

sense of the existent will make this immediately clear. When we say that for something [to exist] means that it can be correctly/fittingly [believed in], we do not thereby say that there must be such [a belief]. What is would continue to exist even if nobody were to [believe in] it. (Brentano 1952: 208 [131]; see also 1889: 24 [24])

This passage appears, if anything, to give priority to the goodness of objects over the fittingness of attitudes toward those objects. At the very least, it may motivate attributing to Brentano a 'no-priority view' of the sort McDowell (1985) seems to defend.

However, the suggestion that Brentano takes mind-independent goodness to be prior to fittingness runs against the veritable difficulty that Brentano does not seem to think there *is* such a thing as worldly goodness inhering in things themselves. In a 1909 letter to Oskar Kraus, he writes:

It is incomprehensible to me what you seek to gain here with your belief in the existence (*Bestehen*) of goodness with which the emotions are found to correspond (*in einer adäquatio gefunden*). (Brentano 1966: 207; see also Brentano 1930: 81 [71])

It is hard to see how to square such a statement even with a no-priority view. It is much more plausible that, in stressing the independence of the good from our attitudes toward goods, Brentano is merely stressing the independence of what it is *fitting* to have a pro attitude toward from the attitudes we *actually* have. To say that it is fitting to have a pro attitude toward generosity is to assert a potentially-counterfactual conditional: *if* anyone should have a pro attitude toward *generosity*, that attitude would be fitting. This does not guarantee that anyone *will* have such a pro attitude. It is only this cleft between attitudes we have or will have and the attitudes it would be fitting for us to have, I suggest, that Brentano shines a light on in the passage that stresses the attitude-independence of goodness. There is no intimation here of any further cleft, between what it would be fitting for us to have a pro attitude toward and what is good.

8

Brentano is clearly an FAist, then. Recall that for him, the essential characteristic of pro attitudes is the attitudinal feature of presenting-as-good. So *his* FA comes to the following:

(FA$_2$) For any good g, (i) it is fitting to have a mental state that presents-as-good g, and (ii) g is good because (i).

As in Chap. 5, any circularity here is merely apparent: 'presents-as-good' should be understood as a merely suggestive but syntactically unstructured expression ('good' is no more a syntactic part of 'presents-as-good' than 'cat' is a syntactic part of 'catastrophe'). What FA$_2$ asserts, perfectly coherently, is that fitting presenting-as-good is prior to goodness: we can understand what it is for something to be good only by understanding what it is to fittingly favor something.

More precisely, Brentano applies his fitting-attitude account to *intrinsic*, *noninstrumental* goodness – goodness 'in the proper sense,' as he puts it. We call a salad healthy, a jog healthy, and a body healthy. But only the body can be healthy in a proper, independent sense; the salad and the jog are healthy in a derivative sense, insofar as they are instrumental to the body's health. In a similar way, claims Brentano, some things are good only in a derivative sense, namely, the sense that they are instrumental in bringing about things which are good in and of themselves. The real question is what makes anything intrinsically good in this way, and it is to answer this question that he offers the fitting-attitude account. In other words, Brentano's overall account of value is better formulated as follows:

(FA$_3$) For any good g, either (a) (i) it is fitting to have a mental state that presents-as-good g, and (ii) g is good because (i), or (b) there is a g^*, such that (iii) g is instrumental in bringing about g^*, (iv) it is fitting to have a mental state that presents-as-good g^*, and (v) g^* is good because (iv).

Here Clause (a) accounts for intrinsic or proper goodness and Clause (b) for instrumental or derivative goodness. Any thing is good that exhibits one or the other.

2.2. What is Fittingness?

In its rather short modern career, FA has generated a sizable critical literature (Baldwin 1999, Rabinowicz & Rønnow-Rasmussen 2004, Heathwood 2008, Alfonso 2009, Bykvist 2009, Reisner 2009 *inter alia*). The ensuing debates have focused mostly on the advantages and disadvantages of FA. But whatever betides these debates, surely the deepest challenge facing FA is to explain what it means for an attitude to be *fitting* (or correct, or appropriate).[9] Thus FA's first order of business is to provide a compelling account of *the nature of fittingness*.

Brentano's account of attitude fittingness parallels closely his account of belief fittingness. Indeed, he develops the analogy explicitly virtually wherever he presents his FA. Thus, the essentials of the account are presented in the comparatively lengthy §27 of *Origin*, but Brentano prepares the way in §26 with a recapitulation of his account of belief fittingness (see also Brentano 1952: 134-47 [84-91]). The account of belief fittingness, recall from Chap. 5, proceeded in two steps: (i) an account of fittingness in terms of self-evidence and (ii) a primitivist account of self-evidence as graspable only through 'assisted revelation.' In like fashion, Brentano's account of pro/con-attitude fittingness involves a primitive phenomenon analogous to self-evidence, graspable only through inner perception against suitable contrasts; the fittingness of the attitudes is then understood in terms of this primitive phenomenon.[10]

To see what Brentano has in mind here, consider first an approximative example. Suppose you pay a cashier for your beers with a 10€ bill, but he gives you back change for 20€. You immediately register his mistake, whereupon two palpable forces instantly start swirling in your consciousness: (a) a desire to tell the cashier of his mistake, and (b) a desire to pocket the extra money. After the fact, you may construct various reasons for and against (a) and (b). But during the two-second interval in which this is all going down, there is something very simple that you just *experience*: you experience that (a) is right or fitting whereas (b) is wrong

and unfitting. One could maintain that fittingness in the practical domain just is the feature that this experiential contrast highlights – the feature that (a) exhibits and (b) lacks.

The resulting moral theory would be inadmissibly naïve about the sources of our experiences of fittingness or rightness. Accordingly, Brentano's theory takes a similar but more nuanced form. (It is a separate question, to which we will revert in due course, whether his nuancing is sufficient to overcome this initial naïveté.) Recall that for Brentano, it is not belief fittingness itself which can only be grasped directly; it is self-evidence, a particularly acute kind of fittingness. Compare the judgment that I have hands and the judgment that *it seems to me* that I have hands (or the judgment that I have a visual experience *as of* having hands). Both judgments are fitting. Moreover, in both cases we feel an irresistible compulsion to endorse the judgment. Nonetheless, there is an important difference between the two, claims Brentano: the judgment that it seems to me that I have hands is self-evident, whereas the judgment that I actually do have hands is not. We have discussed Brentano's defense of this claim in Chap. 5. The claim that concerns us *here* is that some interest states exhibit an analogous feature, one that goes beyond their fittingness and their irresistibility; it is 'the analogue of self-evidence in the domain of judgment' (Brentano 1889: 22 [22]). We can say perhaps that in the same sense in which some judgments are self-evident, some pro or con attitudes are *self-imposing* (my label).[11] It is this feature, this self-imposingness, that Brentano takes to be primitive and unanalyzable, graspable only directly in inner perception.[12] And it is what serves as normative bedrock in his moral philosophy – all value talk is ultimately to be analyzed in terms of this notion.

<center>✍➢</center>

As in the case of Brentano's 'assisted revelation' account of self-evidence, contrast cases can help inner perception home in on, and grasp with greater clarity, self-imposition. Brentano proffers a contrast of the following form. Imagine creatures similar to us in every respect but two: (i) they find chocolate disgusting and mud

<center>11</center>

delicious and (ii) they disapprove of kindness and approve of meanness.[13] These creatures have pro attitudes toward mud and meanness and con attitudes toward chocolate and kindness. Consider now *three* contrasts: (a) the contrast between our and their attitudes toward chocolate and mud; (b) the contrast between our and their attitudes toward kindness and meanness; (c) the contrast between (a) and (b). The feature which interests Brentano is that which comes through clearly in (c). In both (a) and (b), we feel irresistibly compelled toward our own attitude distribution. Nonetheless:

> [In (a),] the compulsion (*Drang*) is an instinctive urge; [in (b),] the natural positive emotion (*Gefallen*) is a higher love [pro attitude] with the character of correctness (*als richtig charakterisierte*). When we encounter in ourselves (*in uns finden*) such a love, we notice not only that its object is loved and loveable, and that its absence and contrary are hated and hateable, but also that the one is love-worthy (*liebenswert*) and the other hate-worthy… (Brentano 1889: 23 [22]; see also Brentano 1952: 145-6 [91])[14]

Clearly, we cannot shake an intense feeling of commitment to our own attitude distribution toward mud and chocolate; we feel strongly compelled to prefer chocolate. The same feeling of unshakable commitment to our own attitude distribution attends the contrast between our and their attitudes toward kindness and meanness; again we feel strongly compelled to prefer our own attitude distribution. But in this case, there is an *additional* difference we appreciate when we entertain and compare the two sets of attitude: ours feel *incontrovertibly justified*; they *impose themselves* on us all-overridingly.

Careful: such descriptors as 'all-overridingly self-imposing' and 'incontrovertibly justified' are supposed to be *helpful* in intimating the practical analogue of self-evidence. But the only way to *really* grasp that which they attempt to intimate is to hold before the mind clearly the feature which contrasts the two attitude distributions in (b) but not in (a). I labeled that feature 'self-imposition,' but we should not mistake a label for a description. Just like self-evidence, self-imposition *cannot* actually be described – it is primitive and graspable only in inner

perception. In a way, the descriptive characterizations offered here tell us *where to look*, not *what we see* when we look where we should.

Once we grasp in inner perception the nature of self-imposition, fittingness is defined in terms of it in the same way belief fittingness was defined in terms of self-evidence. That is: a pro attitude is fitting iff either it is self-imposing or it is such that had the subject's attitude been self-imposing, it would be a pro rather than con attitude.[15] So, Brentano's FA could be formulated more fully as follows:

(FA$_4$)   For any good g, either (a) (i) any self-imposing mental state directed at g would present-as-good g, and (ii) g is good because (i), or (b) there is a g^*, such that (iii) g is instrumental in bringing about g^*, (iv) any self-imposing mental state directed at g^* would present-as-good g^*, and (v) g^* is good because (iv).

In Chap. 5 we encountered some difficulties for the parallel account of existence. There I recommended that Brentano treat self-evident beliefs as just specially acute or manifest cases of fitting belief, so that fitting beliefs are understood simply as beliefs which resemble manifestly fitting ones in relevant respects. A similar move is possible for attitude fittingness and self-imposingness: a self-imposing pro attitude could be seen as an attitude whose fittingness is manifest from the first-person perspective. It is with reference to such manifestly fitting pro attitudes that we can grasp directly the nature of fittingness, but once we grasp it we can imagine a fitting attitude whose fittingness is not manifest in the same way.[16]

2.3. Naturalism or Nonnaturalism?

Brentano's theory of value reductively accounts for goodness in terms of fitting pro attitudes, then, and for attitude fittingness in terms of (actual or counterfactual) self-imposition. Is this a form or moral naturalism or nonnaturalism?

In discussing the formulation of moral (non)naturalism, we have distinguished three conceptions of the natural. The first identified the natural with

what is (a) invoked by (current or future) natural science, or else (b) is reductively explicable in terms of what is. More specifically, it construed moral naturalism as the claim that there are no moral truths whose truthmakers have non-natural constituents by the lights of this 'natural science' criterion. On the face of it, this understanding of moral naturalism casts Brentano's account as rather *non*naturalist. Brentano does accept that there are moral truths. For him, the truthmaker of a moral truth such as 'Generosity is good' is the fact that it is fitting to have a pro attitude toward acts of generosity, and ultimately, the fact that anyone who adopted a *self-imposing* pro or con attitude toward acts of generosity would adopt a *pro* attitude.[17] In this ultimate truthmaker, such notions as 'pro attitude' and 'acts of generosity' are plausibly invoked in some natural sciences (notably psychology and sociology). The problem is self-imposition. As a normative notion, self-imposition does not seem like the kind to be invoked in any natural science. Certainly it is not invoked in any *current* natural science, but the problem seems principled, so that future natural science is unlikely to invoke self-imposition as well. It is furthermore clear that for Brentano, self-imposition cannot be reductively explained in terms of some notions that are or will be invoked by natural science, since it cannot be reductively explained at all – it is a primitive notion. So, by the lights of the 'natural science' criterion, Brentano's FA seems nonnaturalist.

On the other hand, perhaps Brentano would insist that there is at least one natural science that does invoke self-imposition, namely, descriptive psychology. After all, the contrastive procedure by which self-imposition comes into inner-perceptual relief is a bit of Brentanian descriptive psychology. On these grounds, some scholars have argued that Brentano is a naturalist after all (Olson 2017).

This is connected also to the third conception of the natural we have considered, that of 'empirical observability.' Recall that self-imposition is revealed by inner perception, and inner perception is the cornerstone of 'psychology from an *empirical* standpoint' (that is, *empirical* psychology). We may therefore say that self-imposition is empirically perceptible, and to that extent can be seen as a natural phenomenon. Indeed, its empirical perceptibility is what grounds its appearance in

descriptive psychology, a natural science. So by these lights, Brentano's FA seems naturalist.

On the other hand, for Brentano empirical perceptibility is not quite empirical *observability*. The latter requires the capacity to *attend* to that which one perceives. The natural scientist does not just *see* things, she *looks at* them and *observes* them – this is part of what it means to *study* them. Now, typically what is perceptible is also observable. But as we saw in Chap. 1, the move from perception to observation is problematic in the mental domain, the domain of phenomena studied in descriptive psychology. What is outer-perceptible is also outer-observable at the time of its occurrence, but what is inner-perceptible is *not* also inner-observable at the time of its occurrence. This is because outer observation does not alter the character of the thing observed, whereas inner observation does. I am not defending any of these claims here, but am pointing them out as potential grounds for denying that, all said and done, Brentano takes self-imposition to be *observable* – observable the way it really is independently of its being observed, that is.

So far we have reached no decisive result regarding the naturalistic status of self-imposingness. Might a tiebreaker be provided by the conception of the natural as the spatiotemporal? Unclear. At one level, this conception does favor the attribution of nonnaturalism to Brentano. Recall, first of all, that Brentano rejects properties in his ontology. So *his* theory of value is not a theory about the status of the *property* of self-imposition. Instead, it concerns the status of certain special concrete particulars, of the sort discussed in Chap. 6. In addition to the substance Beyoncé and the 'accident' famous-Beyoncé, there are also the accidents generosity-approving-Beyoncé, fittingly-generosity-approving-Beyoncé, and self-imposingly-generosity-approving-Beyoncé. Our question is whether self-imposingly-generosity-approving-Beyoncé is a spatiotemporal particular or not. And the answer, it seems, must be that it is temporal but *not* spatial. Because of his substance dualism, Brentano holds that Beyoncé herself is an *immaterial* substance (a soul). This seems to imply that such accidents as generosity-approving-Beyoncé and self-imposingly-

15

generosity-approving-Beyoncé are immaterial accidents. For they coincide with Beyoncé, and it is hard to make sense of an immaterial concrete particular coinciding with a material one. So self-imposingly-generosity-approving-Beyoncé is probably an immaterial concrete particular in Brentano's ontology. This would appear to cast it as non-spatial:

> All physical phenomena, it is said, have extension and spatial location… The opposite applies, however, to mental phenomena; thinking, willing, etc. appear extensionless and without spatial location. (Brentano 1874: I, 120 [85])[18]

It would seem, then, that self-imposingly attitudinizing persons are *not*, in Brentano's picture, spatial entities neatly embedded in the spatiotemporal manifold that current-day naturalists take to constitute the natural world. To that extent, Brentano seems like a nonnaturalist by the lights of the spatiotemporal criterion.

At the same time, part of the reason current-day naturalists take nature to be a spatiotemporal system is that they do not think of mental phenomena as aspatial. If they did, they might very well expand their conception of nature to include aspatial phenomena, perhaps insisting only that all natural phenomena must be (i) temporal and (ii) causally efficacious. (Self-imposition appears to be very much causally efficacious in Brentano's picture!) Thus to the extent that Brentano comes across as a nonnaturalist by the lights of the spatiotemporal criterion, this is not because of any special feature of his *metaethical* theory; it is only due to an ancillary position on the mind-body problem. (A reist of a materialist bent who accepted every detail of Brentano's FA could qualify as a naturalist about goodness.)

I conclude that none of the standard conceptions of the natural renders an unequivocal verdict on whether Brentano is a moral naturalist or nonnaturalist. At bottom, however, I think the reason it is hard to settle this question is Brentano's own fault. There is an internal tension in *his* conception of self-imposition that pulls us in different directions. On the one hand, he wants it to be an *inner-perceptible* feature and to that extent a *psychological* feature, a feature that shows up in a complete descriptive psychology. On the other hand, he wants it to be a genuinely

normative feature, something that goes beyond what takes place and pertains to what ought to take place. Importantly, however, this is not part of a reductive project of accounting for normativity in nonnormative terms. As we will see in the next section, Brentano stresses the difference between the following two claims:

(C₁) x is good because x elicits in us a pro attitude.

(C₂) x is good because it is fitting for x to elicit in us a pro attitude.

Brentano argues, as we will see, that we must move beyond C_1 and adopt C_2. Accordingly, the facts that it is fitting to disapprove of generosity and that it is unfitting to approve of genocide are fully independent of what actually *happens* in the world. If it just so happened that nobody had ever actually enjoyed a self-imposing attitude, then a complete empirical description of the world would leave out the fact that it is fitting to disapprove of genocide. And yet, this would not make it any less the case that it is fitting to disapprove of genocide and unfitting to approve of genocide. So: there is not reductive project here, so that self-imposition is an irreducibly normative property – and yet it is an inner-perceptible property, hence a psychological property.

In §4.3, I will discuss more critically this inner tension in Brentano's conception of self-imposition. The point I am making now is just that this Janus-faced character of self-imposition creates conflicting pressures to classify Brentano both as a naturalist and as a nonnaturalist. On the one hand, he seems like a naturalist insofar as he takes a certain *psychological* feature to be normative bedrock. But ultimately, we cannot seriously attribute to Brentano moral naturalism, given that he allows a complete description of the natural world to leave out the fact that it is fitting to disapprove of genocide (hence that genocide is bad) – and not because of any expressivism about what is thereby said. If this is naturalism, it is unlike any naturalism we are familiar with. So we cannot comfortably classify Brentano either as a naturalist or as a nonnaturalist. Moreover, this does not seem like a case of a perfectly stable view that simply defies traditional categories; it looks

more like an unstable view that tries to eat a cake and have it too – in this case, to maintain irreducible normativity but have it submit to perceptual detection.

For want of a better option, I hereby resolve to consider Brentano's a form of moral *a-naturalism*. What does 'a-naturalism' mean? Nothing more than a metaethical theory that cannot be comfortably classified as naturalist.

3. The Case for Subject-end A-naturalism

In *Origin*, Brentano offers no sustained argument for his a-naturalist fitting-attitude theory of value. There is a gesture toward a very patchy argument from elimination in §§8-10 and another gesture toward an argument from analogy (with the case of truth) in §§26-7. In his lecture notes, however, Brentano develops a more sustained version of both arguments. The argument from analogy probably captures Brentano's most basic motivation for his fitting-attitude theory, but it presupposes much Brentanian orthodoxy; accordingly, I leave discussion of it to the end of the section (§3.4). The bulk of this section is dedicated to Brentano's argument by elimination, which enjoys a more neutral starting point. Strikingly, the argument anticipates substantially Moore's well-known *open-question argument* (Moore 1903a §27). I open with a presentation of Brentano's version of the open-question argument (§3.1), then raise and try to address two lacunas in Brentano's argument (§§3.2-3.3).

3.1. Brentano's Open-question Argument

Brentano's argument (1952: 114-33 [71-83]) proceeds in two steps. First, Brentano sets forth two constraints on the adequacy of any account of 'the right end,' that is, of what goals we ought to pursue. Second, he shows how these constraints cannot be satisfied by various competitors of his FA. The first constraint is:

> If something is recognized as the right end (*Zweck*), no place remains for the question: 'Am I doing good (acting reasonably) if I pursue it?'... Thus we already have one [adequacy] criterion/standard (*Maßstab*) in the requirement that the concept of the right end should not be so formulated that such a question could still arise once something has been established to be that end. (Brentano 1952: 114 [71])

The other is:

> Another requirement is equally justified: the definition (*Bestimmung*) of the right end may not presuppose anything that does not exist. (Ibid.)

The first adequacy constraint is that an account of value not be susceptible to an open-question argument. The second is that it be ontologically adequate, so to speak, availing itself only of what is real (thermodynamics, with its ideal gas laws, would not pass this test!).

Not unlike Moore, Brentano uses his first constraint to brush aside various broadly utilitarian accounts of the nature of the good, notably Mill's. He construes Mill's view as the claim that we should maximize desire-satisfaction, and objects to it as follows:

> ... the concept of 'most-desirable' (*Begehrbarste*) does not coincide with the concept of the right end. One must ask *why* ought one to desire the most desirable state of affairs. (Brentano 1952: 132 [82]; emphasis original)

Here 'desirable' should be understood nonnormatively, as designating simply what satisfies our desires (Brentano speaks of *begehrbar*, not *begehrwert*). Brentano's objection is that such an account leaves open the question of why we should maximize desire-satisfaction. Various other forms of object-end naturalism are targeted by Brentano as well; he wields his ontological constraint against some of them, but the open-question constraint against all.[19]

The ontological constraint is more operative in Brentano's critique of subject-end naturalism, notably naturalist response-dependence theories. According to these, something is good just if it tends to elicit certain responses in normal subjects under normal conditions. Brentano divides such theories into two

groups: those that identify the relevant responses as *perceptual* states, and those that identify them as *emotional* states.

The first hold that for something to be good is for it to elicit the right moral-perceptual experiences in the right subjects under the right conditions. Brentano objects:

> There is just no such thing as a sensation/perception (*Empfindung*) having for an object a quality called moral goodness; it is an *ad hoc* invention. But even if there were such a sense directed at a moral quality, it would be questionable (as it is with colors and sounds) whether the objects of that sensory perception are reality or mere appearance. In other words, the goodness of a given pursuit could be a mere delusion/mirage (*vorgespiegelt sein*), [so] the question would be justified *why* we ought to pursue it. (Brentano 1952: 119-20 [74]; emphasis original)

Brentano denies here the *existence* of moral perception, on the grounds that moral properties are not sensible (see McBrayer 2010). In addition, he points out that the resulting response-dependence theory leaves open the why-question: since perceptual experience is never guaranteed to be veridical, moral-perceptual experience may be nonveridical as well; it would then be very reasonable to ask why we should pursue what such experience presents as good. If so, what-we-should-pursue cannot be identified with what-we-perceive-as-good.

The second response-dependence account Brentano considers is the more familiar one that *x* is good just if it elicits a positive *emotion* in normal subjects under normal conditions. This is more ontologically acceptable, since there clearly *are* moral emotions. However, it does not close the why-question:

> Suppose other, perhaps higher, kinds of beings did not share these human feelings, but on the contrary disapproved of that which we like, having reason on their side? Or is it [i.e., reason] actually on ours? The question is well-placed, and that suffices for us to reject the definition. (1952: 120 [75])

Since emotional reactions can vary, it will always be perfectly intelligible to ask which emotional reactions we should trust, and in particular whether it is our own

20

that we should. Thus subject-end naturalism succumbs to the open-question argument as well.

<center>❧❧</center>

It is noteworthy that Brentano does not offer, in the relevant lecture notes, any argument against object-end *non*naturalism such as Moore's. I suspect he thinks object-end nonnaturalism blatantly violates his ontological constraint: as we already saw, Brentano rejects the existence of worldly goodness (recall the 1909 letter to Kraus). Thus between the two of them, his constraints appear to rule out all major competitors. One wishes, of course, that Brentano had provided *an argument* against eliminativism about nonnatural worldly goodness; but to my knowledge he had not.

There are two kinds of competitor Brentano does not address, however. Probably they simply did not occur to him. One competitor accepts the entire argument, but points out that nothing about the argument forces us to account for fittingness in terms of self-imposingness. Alternatives include accounting for it in terms of *reasons* or in terms of *ideal observers*; we take up this lacuna in §3.2. Another competitor rejects the open-question constraint, arguing that goodness can be reductively accounted for in a way that does not make the why-question disappear. Many modern naturalists have taken this route, offering a Kripkean necessary but *a posteriori* identification of goodness with natural properties and insisting that this is consistent with the persistence of an open question (see Ball 1988); we take up this competitor in §3.3.

In addition, it would presumably strengthen Brentano's argument if he showed that *his* account, as captured in FA$_4$, satisfies his two constraints. Presumably he would not appeal to something he does not think actually exists in his account. But he could still try to show that the account closes the why-question. To my knowledge, he does not do so. Fortunately, in responding to Moore's version of the argument, A.C. Ewing (1939) fills in this gap:

> What Moore is attacking is any attempt to define 'good' wholly in non-ethical terms, and in this I agree with him... What I shall suggest is a definition of 'good' partly in ethical and

<center>21</center>

partly in psychological terms… [Specifically,] 'the good' means 'what it is fitting to bring into existence for its own sake,' or 'what ought to be brought into existence, other things being equal'; but the important point I think is that 'good' has been defined in terms of what Ross calls a pro-attitude. When something is good it is fitting that we should welcome it, rejoice in it if it exists, desire and seek it if it does not exist. (Ewing 1939: 6, 8)

Regardless of whether one ultimately accepts it, FA indeed seems unsusceptible to the open-question argument. Suppose I ask you why I should favor an end to world hunger, and you reply that an end to world hunger is something that merits approval, something it would be fitting to approve of. I can then reasonably ask you why it merits approval, but I cannot reasonably ask you why I should favor something that merits approval. To ask why I should favor something it is fitting to favor is to betray a lack of mastery of the words I am using.[20] Contrary to Moore, then, goodness does submit to analysis – just not an analysis that does away with normative terms. What it submits to is a Brentanian analysis in terms of fittingness. This gives a certain initial attraction to Brentano's FA. Nonetheless, there are still the two kinds of competitor that need to be addressed.

3.2. Alternative Accounts of Fittingness

According to the buck-passing account *of goodness*, *x* is good just when there is a sufficient reason to have a pro attitude toward *x* (Scanlon 1998). This need not, but *can*, be factorized into two distinct ideas: a fitting-attitude account of goodness and a reasons-based account *of fittingness*. According to the latter, the fittingness status of an attitude is a resultant of all the reasons for and against adopting it (Schroeder 2010). When the balance of reasons favors adopting the attitude, doing so is fitting. In short: an attitude A is fitting iff there is sufficient reason to adopt A. Thus, it is fitting to approve of generosity because there are sufficient reasons to do so – the balance of reasons for and against approving of generosity favors approval. If one goes on to treat the notion of reason as unanalyzable and fundamental, one obtains a 'reasons first' approach to value.

The main problem for this approach is the much-discussed 'wrong kind of reasons' (WKR) problem (Crisp 2000, Rabinowicz & Rønnow-Rasmussen 2004). Suppose an omnipotent demon threatens to kill everybody unless I adopt a pro attitude toward him, say admire him or wish him well. Intuitively, this (a) gives me a sufficient reason to adopt the pro attitude toward the demon, but (b) does not make the demon good. The problem appears to apply mutatis mutandis to the reasons-based account of fittingness. For it is also intuitive that the demon's demand (a) gives me a sufficient reason to adopt a pro attitude toward the demon, but (b) does not make it fitting, in the pertinent sense, to do so.[21]

The literature is replete with attempts to overcome the WKR problem, and some may well succeed. But it is noteworthy that the problem cannot even arise if we adopt Brentano's primitivist account of fittingness. If the nature of fittingness, or at least manifest fittingness (i.e., self-imposition), is grasped directly through inner perception against appropriate contrasts, then far from being an embarrassment, the evil-demon scenario is a potential *tool* of the account. For it provides further contrast, hence further opportunity for homing in on the right feature in inner perception. In contemplating side by side, so to speak, (a) an admiration for a colleague of striking integrity and (b) an admiration for a person who threatens terrible consequences unless we admire her, we *grasp directly* an evident difference between the two. Something is missing in (b) that is present in (a), and that feature is what we have labeled fittingness.[22]

It might be objected that this is a cheap response to the WKR problem. *Of course* we can see intuitively that there is a difference between (a) and (b); nobody contests that. The problem is to *account* for this difference in a principled and noncircular manner. However, for Brentano there is no account of the difference to be had, principled or otherwise; all we can aspire to is inner-perceptual acquaintance with the right feature. Admittedly, underlying the objection under discussion is an understandable kind of intellectual disappointment: being told 'just look in this direction and pay close attention' does not offer material for reasoned consideration and debate, and in some sense does not advance our theoretical

understanding of the phenomena. This, I suspect, is the substance of the complaint at hand. But although understandable, the relevant intellectual dissatisfaction is (for Brentano) something we must learn to live with. Every theory has its primitives, and the same stock complaint could be lodged against it when it confesses to its own primitives. The 'reasons first' approach will be in the exact same position when it comes to elucidating the notion of a reason.[23] The advantage of Brentano's account is that it involves a sophisticated, thought-out approach to the appreciation of primitives, namely, the assisted-revelation approach discussed in Chap. 5.

<p style="text-align:center">❧❧</p>

'Reasons first' is probably the most popular approach to fittingness. A more recent but perhaps more attractive approach may be called 'ideality first.' The idea is that a pro attitude is fitting just if the an ideal observer would adopt it (Kauppinen 2014).[24] The reason it is fitting to approve of generosity is that an ideal subject would. When this is plugged into FA, we effectively obtain the ideal-observer account of value: x is good iff 'any ideal observer would react to x [with a pro attitude] under such and such conditions' (Firth 1952: 321).

Kauppinen motivates this view by noting that our instinctual, prereflective pro attitudes may lead to various practical problems: 'acting on just any desire may result in ill health, garish wallpaper, a broken marriage, or a distrustful community' (Kauppinen 2014: 580). To avoid these practical problems, we try to step back from our instinctual pro attitudes and form more sophisticated ones, ones that lead to fewer problems. Ideal observers are ones who form those pro attitudes which bring such problems to the minimum. Other things being equal, their reward is good health, tasteful wallpapers, lasting marriages, and widespread trust.[25]

Debates surrounding ideal-observer theories tend to center on the nature of the ideal observer (Brandt 1955), the motivation for idealizing (Enoch 2005), potential indeterminacy problems (Carson 1984), and epistemological concerns about non-ideal humans' access to ideal observers' attitudes (Sayre-McCord 1994).

Here I want to present a *metaphysical* argument against 'ideality first,' arguing that it has implausible consequences for what is ontologically bedrock.

A major attraction of moral naturalism about value is the parsimonious picture of ontological bedrock it offers: it allows one to claim that only the microphysical is fundamental. Nonnaturalist approaches to value, in contrast, require that we add something of a normative character to our ontological bedrock. For Moore, that 'something' is simply the property of goodness. For Brentano, it is rather self-imposition, or more accurately self-imposingly-attitudinizing-subjects. In an 'ideality first' framework, meanwhile, what must be added to our ontological bedrock are *ideal observers*. What I want to argue in the remainder of this section is that ideal observers are not of the right sorts of thing to serve as ontological bedrock.

As noted, there is disagreement about what constitutes an ideal observer. A natural view construes the ideal observer as (i) a perfect reasoner who (ii) knows all the relevant facts. One could then say that it is fitting for S to approve of *x* just if a perfect reasoner who knew all the (nonnormative) *x*-relevant facts would approve of *x*; that is, just if someone with (i) complete knowledge of the nonnormative facts and (ii) the disposition to perform all and only epistemically justified inferences would approve of *x*. The immediate problem with this is that there probably *are* no such subjects in the actual world. (Who among us knows *all* the nonnormative facts pertaining to, say, whether Brazil ought to negotiate a trade agreement with Singapore?) Accordingly, all ideal observers *reside in other possible worlds*. This is a problem for the claim that ideal observers are *fundamental*. For it entails that certain fundamental entities are not actually present in our world. They occur only in nonactual worlds. What is mysterious is how the ontological bedrock of *this world*'s beings can include *other worlds'* entities. How are *actual* facts about fittingness and goodness supposed to be undergirded by 'facts' about ideal observers which are *nonactual*?

The basis of the objection is the thought that while we resort to modal notions routinely in philosophy, it is highly controversial to do so *at the fundamental level*. We expect the ungrounded grounders of reality, so to speak, to be part of the actual world. This is why metaphysicians make enormous efforts to find actual truthmakers for modal truths (see Armstrong 1997 Chap. 10 for discussion). Modulo Lewis' (1986) industrial-strength realism about concrete possible worlds, it is a challenge to identify an element of *reality* (read: the actual world) that makes true 'Possibly, there are seven-foot tall butterflies.' Likewise, it should be a challenge to find an element of reality that makes true 'In possible world W_{17}, Jimmy (an all-informed perfect reasoner) approves of a trade deal between Brazil and Singapore.' But note that if we did identify some such actual-world truthmaker for this claim, then presumably that truthmaker would preempt ideal observers as fundamental.

It would represent only minor progress to trade modality for dispositions here. It might be suggested that we idealize not the observers themselves, but the triggering conditions of their dispositional states. The claim would be that if *ordinary* subjects are disposed to approve of *x* upon (i) learning all the relevant facts and (ii) performing all and only justified inferences of relevance, then it is fitting to approve of *x*.[26] In this picture, it is fitting to approve of generosity because *ordinary* subjects are disposed to do so under *ideal* triggering conditions. The problem, obviously, is that most philosophers reject the notion of dispositions at the fundamental level of reality. It is typically assumed that dispositions must be grounded in some categorical bases.

A theist 'ideality first' theorist might of course claim that God is an actual-world all-knowing perfect reasoner. So it is fitting to approve of generosity because God does. This is not quite a divine *command* theory but a 'divine example' or 'divine guidance' theory (see Jordan 2013) – the sort encapsulated in 'What would Jesus do?' This view brings the ideal observer into the actual world, making appeal to fundamental modality (or dispositions) unnecessary. One obvious problem with it, however, is that it would work only for theists. In any case, Brentano himself would reject this, despite his theism, on the grounds that a divine command/guidance

theory is susceptible to the open-question argument: 'If divine revelation is appealed to, the question of whether God's word is worthy of belief remains open…' (Brentano 1952: 30 [19]; see also 1952: 115 [72]).

One might suggest instead that we simply lower our expectations from so-called ideal observers, lower them sufficiently as to welcome some such into our actual world. Thus, in his critique of Firth's ideal-observer theory, Brandt (1955) argues that the ideal observer must be recognizably human to be ethically relevant, and therefore should meet only two kinds of requirement: being (i) 'a disinterested, dispassionate but otherwise normal person' with (ii) 'vivid awareness of [those facts] which would make a difference to his ethical reaction in [a particular] case' (Brandt 1955: 410). It might be thought quite plausible that *someone* in our world is normal, disinterested, dispassionate, and has vivid awareness of facts that make a difference to whether she approves of a trade agreement between Brazil and Singapore.

This suggestion faces at least three major problems, however. First, it is far from obvious that for *every* ethical issue the actual world happens to contain someone with the right profile. For example, whether lobsters undergo a phenomenally conscious experience is plausibly relevant to whether it is fitting to eat them; but setting aside the fact that nobody *currently* knows whether lobsters are conscious, it might be held that only someone who *has been* a lobster can know whether lobsters are conscious. Secondly, if ideal observers can vary widely among them in their inferential capacities, disagreements among them are likely to be rife: some would conclude that it is fitting to approve of abortion and some would conclude it is unfitting. This produces a can of worms for the ideal-observer theorist.[27] Thirdly, and most embarrassingly, consider a case where (i) a disinterested actual subject with vivid awareness of the facts regarding crack cocaine approves of legalizing it, while (ii) a perfectly informed perfect reasoner would disapprove of legalizing it. It would be folly to follow the suboptimal reasoner's verdict over the optimal one's. But if so, the suboptimal (actual)

reasoner's attitudes should fix attitude-fittingness only when they coincide with the optimal (nonactual) reasoner's.

I have considered four possible assays of ideal observers: as nonactual fully informed perfect reasoners; as ordinary subjects with idealized dispositions; as God; as disinterested actual subjects with vivid awareness of relevant facts. On all four, I have argued, ideal observers are ill suited to serve as ungrounded grounders of actual-world normative facts, including fittingness facts. This casts a long shadow over the 'ideality first' approach to fittingness.

To be sure, the 'ideality first' theorist may choose to offer *no* account of observers' ideality, construing it as primitive and inexplicable. (She is entitled to her own primitives, after all!) But this leaves us with the challenge of how to grasp the nature of observers' ideality without a theoretical explanation of what such ideality consists in. Brentano, we have seen, *has* a story to tell about how we are to grasp the nature of self-imposition in the absence of any theoretical explanation. That story appeals to direct encounter in inner perception, which is self-evident and therefore infallible. For Brentano, there is no way we can have self-evident encounter with the ideality of an observer, since the latter is not an inner-perceptible phenomenon. This is why primitivism about ideality is not as promising, he would claim, as primitivism about self-imposition.

<div align="center">❧❦</div>

The other alternative to Brentano's self-imposition-based account of fittingness is the 'reasons first' approach, but the latter still has to contend with the WKR problem, which incidentally has no force against Brentano's account. The upshot is that while Brentano himself does not pursue his argument from elimination far enough, leaving 'standing' two alternative versions of FA, these alternatives can be ruled out by additional considerations.

3.3. A Posteriori Naturalism

Gilbert Harman objected to Moore's open-question argument as follows:

> … as it stands the open-question argument could be used on someone who was ignorant of the chemical composition of water to 'prove' to him that water is not H_2O… [But, s]ince this argument would not show that water is not H_2O, the open-question argument cannot be used as it stands to show that for an act to be an act that ought to be done is not for it to have some natural characteristic C. (Harman 1977: 19)

The fact that for any natural characteristic C, one can still *meaningfully* ask whether C is good only shows, for Harman, that the thesis that goodness = C cannot be a priori. But it can perfectly well be a posteriori and true, indeed a posteriori and *necessarily* true.[28] This is precisely the tack adopted by Cornell realists (Boyd 1988), who model the reduction of goodness to some natural property on the reduction of water to H_2O.

It will not help here to simply reject the Kripkean approach to necessary a posteriori truths and adopt the more deflationary, 'Australian' approach. According to the latter, every necessary a posteriori truth can be factorized into two claims, neither of which both necessary and a posteriori (Jackson 1998 Chap. 3). For example, 'Necessarily, water = H_2O' results from the superposition of two ideas: (1) Necessarily, water = the actual occupant of the water role; (2) The actual occupant of the water role = H_2O. Here, 'the water role' means (roughly) the total causal role of water in nature. And the notion that 'water' refers to whatever plays or occupies that role is claimed to be an a priori conceptual truth.[29] Meanwhile, the claims that H_2O is what plays or occupies that role is claimed to be a posteriori but contingent. It is only the fortuitous superposition of one a priori necessary truth and one a posteriori contingent truth, according to this more deflationary account, that produces a composite truth both necessary and a posteriori. The point I want to make here is just that even against the background of this more deflationary account, the open-question argument seems fallacious (Jackson 1998: 150-1). For a naturalist could still argue as follows: (1) Necessarily, goodness = the actual occupant of the goodness role; (2) Natural characteristic C = the actual occupant of the goodness role; so, (3) Necessarily, goodness = C.

Obviously, Brentano did not consider the possibility of this kind of moral naturalism. Accordingly, we are consigned to considering whether arguments completely external to the Brentano corpus could be wheeled in to rescue Brentano's open-question argument for FA. Unfortunately, the topic is too vast and too complex to settle here. I will restrict myself to citing two extant objections to the aqueous bit of reasoning just laid out. I should stress that my contribution to this discussion is nil; I am merely reporting other philosophers' work, work that a Brentanian *could* appeal to in supplementing Brentano's argument against moral naturalism.

Horgan & Timmons (1992) have objected to Premise 1 that there is a deep asymmetry between 'Water = the actual occupant of the water role' and 'Goodness = the actual occupant of the goodness role.' On Twin-Earth, where everything is the same as on Earth except that what plays the water role is not H_2O but XYZ, the word 'water' means something different than it does here. If earthlings and twin-earthlings met one day, they might have an apparent disagreement on what water is, until they realized that they were just using the word differently. They could then all agree that, as long as they speak English, there is no water on Twin-Earth (and if they switched to speaking Twin-English, it would be correct to say 'There is no water on Earth'). Contrast now a third planet, Moral Twin-Earth (MTE), where everything is like on Earth except that the goodness role is occupied by C* and not C. If earthlings and MTEists met one day, they would quickly disagree on whether it is C or C* that constitutes goodness. But here it does not seem that the disagreement is merely verbal, so that everybody might at some point realize that the word 'good' means something different in English and in Twin-English. On the contrary, the disagreement will persist and, if C and C* are sufficiently different, Earthlings and MTEists may reasonably war over who is right. The question seems substantive (Horgan & Timmons 1992: 165). This shows that it cannot be in the very *meaning* of 'good' to pick out the *actual* occupant of the goodness role, as Premise 1 claims.[30]

Premise 2 is questionable as well. In the case of water and H_2O, we can point at empirical research designed to expose the molecular composition of water,

leading to the empirical discovery that it is H_2O which plays the water role. In the case of goodness, by contrast, it is hard to imagine what empirical research is supposed to identify the occupant of the goodness role (see Huemer 2005 Chap. 4).

I conclude that although Brentano's argument by elimination contains two important lacunas, it might admit of suitable supplementation to provide real support for FA$_4$.

3.4. The Basic Motivation

We do not fully appreciate the basic motivation for Brentano's FA if we only rely on his open-question argument. The analogy Brentano belabors with his theory of existence is potentially of greater importance.

Recall that according to Brentano, saying that *x* exists amounts to saying that it is fitting to believe in *x* (where the relevant fittingness is ultimately assayed in terms of self-evidence). The basic motivation for this is that there is no property of existence – so that attributions of such a property to objects would result in falsehoods – and yet there is this important difference between 'Ducks exist' and 'Ghosts exist': the former is true while the latter is false. To make sense of the possibility of existential truth, Brentano makes a double move. The first is intended to illuminate the truth of existential *judgments*, the second that of existential *sentences*. The first move is to adopt an attitudinal account of judgments' existence-commitment, so that a judgment could commit to the existence of an object without quite attributing to that object a property of existence. The second move flows directly from this first one. Since the content of an existential judgment is exhausted by the object, and not by the proposition that the object exists, the corresponding existential sentence cannot be understood to simply express the proposition that the object exists. Instead, it must be understood as expressing the idea that the correct judgment to make toward that object is the kind of judgment that commits to the existence, rather than to the nonexistence, of the object.

Consider now the analogy with goodness. As we have seen in that letter to Kraus, Brentano rejects the existence of a property of goodness. It follows that any attribution of goodness to something would result in a falsehood. And yet there is this important difference between 'Genocide is good' and 'Compassion is good': the former is false while the latter is true. How can we make sense of moral truth? The answer proceeds in an analogous two-step fashion. First, we must adopt an attitudinal conception of *mental* goodness-commitment, so that a pro attitude toward an object could commit to the object's goodness without attributing goodness to it. Second, and in consequence, we must understand *linguistic* goodness-commitment as a matter of expressing not the proposition that the object is good, but the more subtle idea that the right attitude to take toward that object is a pro rather than con attitude. This is how we get to a fitting-attitude account of moral (and more generally evaluative) talk.

One weakness here is that it is clear, from a certain perspective, why there could not be a property of existence: things do not divide into those that exist and those that do not! But it is unclear why there should not be a property of *goodness*; things *do* divide, after all, into those that are good and those that are not. Insofar as the basic motivation for FA takes the nonexistence of a property of goodness as its starting point, it seems to beg the question against any form of moral realism, whether naturalist or nonnaturalist. What is missing from Brentano's overall case for his theory of value, then, is an argument against such realism. Here too, we could attempt to supplement Brentano's argumentation with independent considerations, notably the notorious argument from queerness (Mackie 1977). But it is a curious fact that Brentano himself appears to feel no need to provide *reasons* for setting aside moral realism.

4. **Problems for Brentano's Metaethics**

In this section, I consider three central objections to Brentano's theory of value. I start with the Euthyphro-style objection that the subject-end order of explanation

gets things backwards (§4.1). I then consider two objections presented by Moore in a review he wrote of Brentano's *Origin* (Moore 1903b): one objects to Brentano's account of *betterness* (§4.2), the other to the notion of direct grasp of fittingness through inner perception (§4.3).

4.1. A Euthyphro Problem

Perhaps the most fundamental liability of Brentano's theory of value is the counterintuitiveness of the subject-end order of explanation. The notion that genocide is bad because it is fitting to have a con attitude toward it strikes us as getting things backwards: surely the reason we should have a con attitude toward genocide is that genocide is so very bad.

  One approach to this problem, taken by Ehrenfels for instance, is to *celebrate* the view's counterintuitiveness. For Ehrenfels, the fitting attitude account of value is something of a precious *philosophical discovery*:

> It is not that we desire things because we recognize in them this mystical, unfathomable essence 'value,' but rather that we ascribe 'value' to things because we desire them. (Ehrenfels 1897: 219)

In the background here is the intellectual influence of Austrian economics (see Ehrenfels 1893: 76, Smith 1986, Reicher 2017), one of whose major ideas was the market-based account of *economic* value: the economic value of a thing is a function of how much people actually want it (demand), as well as how scarce it is (supply). People do not want gold because it is valuable; gold is valuable because people want it (and there is not a lot of it). When we move from economic value, which is a *descriptive* notion, to *moral* value, which is a *normative* notion, we may need to also move from the desires people *actually* have to the desires they *ought* to have, that is, the desires it would be fitting for them to have. The result is a fitting attitude account of moral value.[31]

The basic idea behind this unapologetic reaction to the Euthyphro objection is that Brentano cannot shy away from this feature of his theory, since, in a way, it is *the whole point* of the theory. That goodness is actually grounded in fittingness, rather than the other way round, as we might naturally (read: pre-philosophically) assume, is not so much an unwelcome consequence of the view as the view itself!

A second and more substantive line of response distinguishes two nearby intuitions that one might wish to accommodate. Although Brentano cannot accommodate the intuition that it is fitting to oppose genocide precisely because genocide is bad, he *can* accommodate the following nearby intuition: it is fitting to oppose genocide because of *what genocide is like* – where 'what genocide is like' is unpacked in terms of genocide's *nonnormative* properties. Thus Brentano can say that it is fitting to disapprove of genocide because genocide involves the death of so many people, their intentional targeting for their ethnic or racial affiliation, and so on. Seen in this light, the following order of explanation does not seem all that unintuitive: genocide is bad because it is fitting to disapprove of it, and it is fitting to disapprove of it because of what it is like, descriptively speaking. From this perspective, FA boils down to the claim that normativity proper comes into the picture only through the fittingness of pro and con attitudes.[32]

4.2. The Good, the Better, and Fitting Attitudes

In the same year Moore published his *Principia*, he also published a lesser-known review of Brentano's *Origin*. The review opens thus: 'This is a far better discussion of the most fundamental principles of Ethics than any others with which I am acquainted' (Moore 1903b: 115). Despite its generous opening, the review proceeds to develop a powerful critique of Brentano. A centerpiece of this critique is an attack on Brentano's account of betterness in the image of his account of goodness.

According to Brentano, *x* is better than *y* iff it is fitting to prefer (*vorziehen*) *x* over *y* (Brentano 1889: 24-6 [25-6]). Now, one might have held that to prefer one thing over another is just to like or love it more; and some have attributed this view

to Brentano (Olson 2017). But it seems to me that Brentano explicitly rejects such a reductive account of preference in terms of like/love:

> … 'the better' seems to be that which is worthy of being loved with a greater love. But is this really so? … Someone will say, perhaps, that the intensity of enjoyment/liking (*Gefallen*) is what we call here the magnitude of love. If so, the better would be that which ought to be enjoyed/liked with more intense enjoyment/liking. But that would be, upon closer inspection, an absurd definition. It would imply that in every specific case where someone is happy over something, only a certain amount of happiness (*Freude*) is permissible; whereas I should think it could never be objectionable to be maximally (wholeheartedly, as they say) happy over what is really good. (Brentano 1889: 24-5 [25])

The argument is this. If preference were a matter of greater liking, then fitting preference would be a matter of fitting greater liking. However, there are cases where it is fitting to prefer *x* over *y* even though there is no upper limit on how much it is fitting to like *y*. Even if it is fitting to prefer art to football, it does not follow that it is unfitting for me to like football with the maximum amount of liking I am capable of. Therefore, preference cannot be a matter of greater liking. Instead, it must be a sui generis species (*besondere Spezies*) of interest state (1889: 25 [26]).

Moore's main complaint against this is psychological – he thinks preference as conceived by Brentano is just not part of our psychological repertoire:

> If, as usual, we mean by 'preference' a *feeling*, it is obviously a feeling *only* towards the thing preferred, *not* towards the thing which is not preferred. When we say that we prefer one thing to another we usually mean either (1) that we *like* the one *more* than we like the other, or (2) that we choose the one and do *not* choose the other: there is no such thing as a single feeling, called 'preference,' directed to *both* the things. (Moore 1903b: 119; italics original)

If Moore is right that there is no evaluative mental state directed at object-pairs, then Brentano's nonreductive take on preference is doomed.

It is hard to see, however, what motivates Moore's proclamations here. Preferring football to basketball seems very much a mental state directed at both football and basketball. Certainly basketball shows up in the content of the

preference; it is hard to see in what sense Moore thinks it is not *directed toward* basketball.

A more perspicacious objection Moore raises connects the nonreductive account of preference with a nonreductive account of betterness:

> [Brentano] does not clearly recognize that to know one thing to be better than another must be to know that it has in a higher degree the very property which we mean by 'good in itself.' (Moore 1903b: 120; see also Ehrenfels 1988)

If betterness is a matter of fitting preference, and (fitting) preference is irreducible to (fitting) relative liking, then betterness is irreducible to greater goodness. But betterness *just is* greater goodness!

This is a much better objection. Suppose art is more valuable than football. Then it would seem to follow that football must have a smaller quantity of goodness than art. But Brentano's theory has no resources to explain or even accommodate this fact.

Brentano's response here would be to bite the bullet. Perhaps he could claim that betterness is not *in general* reducible to greater goodness, though such reduction does work for goods *of the same kind*. If knowledge is intrinsically good (Brentano 1952: 183 [116]), then other things being equal, more knowledge is better than less (knowing both Euclidean and non-Euclidean geometry is better than knowing only Euclidean). In some cases, however, goods belong to categorically different kinds, so that it makes no sense to say that one contains more goodness than the other. One might hold that in such cases no value comparison is possible, but Brentano's view is rather that these are instances of betterness that do not amount to greater goodness:

> There are also cases of preferences with the character of correctness where the better is not the greater sum [of good]; rather, the superiority (*Vorzug*) is grounded in a *qualitative difference*. It is evident, for example, that positive knowledge is more valuable than negative, other things being equal. (Brentano 1952: 213 [134]; my italics)

As we will see in greater detail in the next chapter, Brentano holds that knowing what *is* the case is more valuable than knowing what is *not* the case. Obviously, this is not a case of having more of the same: other things being equal, knowing that the Higgs boson exists is not *more knowledge* than knowing that the ether does not. Moore would say that even though the former bit of knowledge is not *more knowledge* than the latter bit, it does contain *more goodness*. But this might strike one as an arbitrary assertion. Instead, Brentano suggests that the betterness of positive knowledge is irreducible to greater goodness. In a way, for Brentano each kind of knowledge has its own distinct sort of goodness, and one sort is plain better than the other (without being in any sense more plentiful than the other).

Accepting Brentano's nonreductive account of preference (and betterness), one might wonder why Brentano does not consider the opposite reduction: perhaps to have a pro attitude toward *x* is just to prefer *x*'s existence to *x*'s nonexistence. Thus, to admire Shakespeare is to prefer a world with him to one without him. This would provide for a more unified account of value, reducing goodness to betterness. However, this requires construing preference as a *propositional* attitude, directed not at (pairs of) concrete particulars, but at (existential) states of affairs. As we saw in Chap. 4 and 6, this is anathema to Brentano. Brentano is stuck, then, with a disunified account of value in terms of mutually irreducible goodness and betterness. Furthermore, in being thus disunified, the account breaks the symmetry between value and truth that Brentano otherwise stresses so much. This is one instance, among many, where Brentano's penchant for unities and symmetries does not blind him to the peculiarities of the phenomena.

4.3. Can We Really (Inner-)Perceive Normative Properties?

A major disadvantage of Moore's object-end nonnaturalism is its reliance on *intuition* to account for *knowledge* of goodness. The epistemology of intuition is a notorious can of worms, and philosophers aplenty have suspected that direct

intellectual acquaintance with abstracta, such as intrinsic goodness and badness, is simply not part of our psychological repertoire.[33]

Brentano's subject-end theory of value appears to enjoy a definite advantage here, as it avoids appeal to intuitional encounter with goodness and badness. For Brentano, our knowledge of good and bad is not direct in this way, but is rather based on prior knowledge of fitting and unfitting. We know something to be good by knowing that pro attitudes toward it are fitting. Knowledge of fittingness is in turn based on knowledge of self-imposingness. It is knowledge of self-imposition that is direct and foundational in our moral knowledge. And our direct encounter with it is through inner perception, not intuition:

> … when we call certain objects good, and others bad, we are saying thereby nothing more than that whoever loves the former and hates the latter has taken the right stand. The source of these concepts is inner perception, for only in inner perception do we grasp (*erfassen*) ourselves as something which loves or hates. (Brentano 1952: 144 [90])

It is the central thesis of *Origin* that moral knowledge *originates in* inner-perceptual acquaintance with self-imposingness.

The immediate problem here is that inner-perceiving the items that exhibit self-imposition does not seem to guarantee inner-perceiving self-imposition itself. From the fact that an item is perceptible it does not follow that all its *properties* are perceptible. Thus, one may see a rubber ball at the distance without seeing that it is rubber. Likewise, one may inner-perceive an experience of indignation at Trayvon Martin's murder without inner-perceiving that one's indignation occurs on a Monday – occurring on a Monday is simply not the kind of feature one can pick up using *inner perception*. And by the same token, one may very well inner-perceive that same indignation without inner-perceiving its self-imposingness.

But there may be a deeper problem lurking here. Self-imposition is not any old property of our inner-perceived pro attitudes – it is a *normative* property. And it is specially mysterious how we could possibly (inner-)perceive normative properties. Recall that Brentano himself branded outer perception of normative

properties 'an ad hoc invention' (1952: 119-20 [74]). Why should we be more sympathetic to *inner* perception of such properties? Moore belabors this very point in his review:

> Obviously the conception of 'good,' as Brentano defines it, cannot be derived merely from the experience of *loving*, but only from that of '*right* loving' – from the perception of the *rightness* [or fittingness] of a love: its *origin* cannot be merely the perception of a love which *is* right, but in which this quality is not perceived, it can only be a perception in which it is itself *contained*. But ... The quality of 'rightness' [fittingness] is *not* a psychical content and the perception of it is *not* an impression in the ordinary sense of these words. (Moore 1903b: 117; italics original)[34]

The claim that fittingness and self-imposition, qua normative properties, should not be directly perceptible is antecedently very plausible. Unless this difficulty can be overcome, any advantage over Moore's appeal to intuition is illusory.

One option for Brentano is to seek inspiration from proponents of moral *outer* perception. Cuneo (2003, 2006), for example, suggests that we perceive normative properties *by* perceiving the nonnormative, natural properties upon which they supervene. Just as I hear my wife's approach *by* hearing her distinctive footfalls, I may also see the goodness of a situation *by* seeing the visible natural grounds of the situation's goodness. For example, I may see a child's delight at receiving a gift and thereby see the prima facie goodness of the gifting. If this works for external perception, a Brentanian might suggest, it should also work for inner perception: perhaps we can inner-perceive the self-imposingness of a pro attitude *by* inner-perceiving (some of) the attitude's subvenient psychological properties.

An immediate question is what purely psychological property might subvene self-imposition – especially given that the latter is ontologically bedrock. One view is that an attitude's normative property of *being* self-imposing supervenes upon that attitude's psychological property of *feeling* self-imposing, i.e., having the felt character of self-imposition. This latter property is surely inner-perceptible. It might be suggested, then, that we can inner-perceive an attitude's being self-imposing *by*

inner-perceiving its feeling self-imposing – just as we can outer-perceive a flower's *being* purple by outer-perceiving its *looking* purple.[35]

The main problem with this move is that it undermines the notion of direct grasp (assisted by certain contrasts) of self-imposition. If we inner-perceive an attitude's being self-imposing by inner-perceiving its feeling self-imposing, then the inner perception of being self-imposing is *mediated*: the inner perception of the nonnormative, psychological property of *feeling* self-imposing is *direct*, whereas the inner perception of the normative property of *being* self-imposing is *indirect*. Direct grasp of a normative property seems no longer in the cards. This seems to undermine in turn the 'assisted revelation' theory for the normative property of *being* self-imposing; only the nonnormative property of feeling self-imposing may be revealed.

This result undermines Brentano's entire moral epistemology. Brentano homes in on self-imposition primarily as the 'practical analogue' of self-evidence. Self-evidence, recall from Chap. 5, is more basic than, and guarantees, infallibility – an immunity to mistakes. Likewise, self-imposition should have the advantage of guaranteeing a sort of practical infallibility – immunity to mis*steps*, if you will. And it is this immunity to missteps that bestows on self-imposing attitudes the capacity to serve as foundations of practical knowledge. Once we admit that we enjoy no direct grasp of our attitudes' *being* self-imposing, but only their *feeling* self-imposing, daylight opens between that which is directly grasped and that which guarantees immunity to missteps. An attitude's *feeling* self-imposing does not guarantee immunity to missteps, since it may turn out that the attitude is not *actually* self-imposing. We might put the point this way: by Brentano's own lights, attitudes that merely *feel* self-imposing are not suitable to serve as foundational in our moral knowledge, since they do not exhibit the practical analogue of infallibility; but attitudes which *are* self-imposing cannot serve as foundational either if their being self-imposing is not directly graspable.

The point goes back to the Janus-faced character of self-imposition. It is clear how the property of *being* self-imposing can be irreducibly normative, and also how the property of *feeling* self-imposing can be inner-perceptible; it is much less clear how either property could be *both* normative and perceptible. Certainly Brentano does little to clarify how this could be. The fact that he himself dismisses the possibility of *outer* perception of goodness, hence of the normative status of worldly objects, raises the question of why he sees nothing problematic about inner perception of self-imposition, hence of the normative status of mental states.

<p style="text-align:center">✌❦</p>

My own view is that this is the biggest problem in Brentano's theory of value, perhaps the weakest link in his entire philosophical system. We can get at the daylight between what *appears* right and what *is* right from another angle. The prominent Moore scholar Thomas Baldwin has argued that there is an element of 'wishful thinking' in Brentano's reliance on inner perception to reveal fittingness (Baldwin 2006: 244). There are two complementary concerns here: first, inner perception may fail to detect the fittingness (or unfittingness) of a particular pro attitude; secondly, it may misattribute fittingness to an unfitting attitude or unfittingness to a fitting one. George Washington, a man of rare dignity and courage, who insisted on retiring from the US presidency after only two terms to inspire strong democratic practice in American governance, owned slaves his entire life. Presumably, he harbored a pro attitude toward slave-ownership. This pro attitude was unfitting, but Washington's inner perception failed to detect its unfittingness. Many of us today, including this author, engage in certain meat-eating practices which may well seem in three centuries' time as morally abhorrent as slave-ownership. Presumably, most of us fail to detect the unfittingness of our attitudes toward such meat-eating; many positively attribute fittingness to it. And let us not forget Brentano himself, whose attitudes toward women might raise eyebrows in retrospect:

> On the relative positions of the two spouses I would essentially agree with Bentham's judgment. He demands: a) the subordination of the woman. The man should be in charge (*Vormund sein*). It is no good when the legislature and the executive work against each other. In general it may be said that the man's cognitive power (*geistige Kraft*) is greater, not because women are less smart, but rather because most of them have lower endurance for cognitive labor (*geistiger Arbeit*). (Brentano 1952: 392 [239])

Judging by this passage, Brentano harbored a con attitude toward women 'being in charge,' and a pro attitude toward the subordination (*Unterordnung*) of the wife to the husband. I am going to go ahead and assume that at least one of these attitudes is unfitting. Its unfittingness evidently eluded Brentano's inner perception. In general, much of our lives we are altogether unaware of our attitudes' fittingness or unfittingness. Sometimes we do experience our pro attitudes as fitting or unfitting, but as Baldwin observes, 'we do not have a sufficient understanding of the origins of our feelings to be confident about such an experience of their correctness [or fittingness]' (Baldwin 1999: 237; my translation). If so, for most pro and con attitude we can contemplate, we are in no position to establish whether they are

In this respect, Brentano's theory may bear an unfortunate resemblance to the conscience-based ethical systems of certain early British moralists. According to Butler (1726), for instance, conscience is a faculty that monitors one's mental states and actions and approves or disapproves of them.[36] If one held that what makes conscience (dis)approve of an attitude is the latter's (un)fittingness, the view would be extremely close to Brentano's.[37] This resemblance of ethical theory did not escape Brentano, who is happy to regard conscience theory as in some sense a predecessor (1952: 157-8 [98-9]). The problem with such ethical systems is brought out compellingly in Jonathan Bennett's 'The Conscience of Huckleberry Fin' (Bennett 1974). Bennett argues that the dictates of conscience lie downstream of moral theory, so that when one's moral theory is misguided one's conscience will be as well. An example is Huck Finn's pangs of conscience upon failing to turn in his runaway-slave friend Jim; another is SS chief Heinrich Himmler's pride at following his conscience in trying to see through the Final Solution despite persistent gut feelings of moral discomfort. The parallel problem for Brentano's position is clear:

Huck Finn and Himmler seem to inner-misperceive unfitting attitudes as fitting and fitting attitudes as unfitting.

Three possible responses to this problem should be considered. Hill (1998) suggests that Huck Finn may be misidentifying the voice he hears within him – he thinks it is the voice of conscience, but in reality it is just social conditioning. In a similar vein, Brentano might suggest that although Washington judged his pro attitude toward slavery fitting, that judgment's source was not inner perception but something else – again social conditioning, perhaps (see Gubser 2009). If successful, this move would protect the authority of inner perception.

The price, however, would be to sacrifice any first-person insight into when inner perception (or conscience) is operative. We would be justified in relying on our inner perception, but would have no way to tell when what we are relying on is inner perception. This is precisely the problem Baldwin describes in complaining that we have insufficient understanding of the processes that lead us to judge that our attitudes are fitting or unfitting. In many cases, we have no direct, first-person insight into whether we are undergoing a conscience-driven process or a self-serving one.

A second response is that while inner perception of an attitude's *fittingness* may go awry, inner perception of an attitude's *self-imposingness* cannot. Let us continue to indulge Brentano's supposition that knowledge is an intrinsic good (we will adopt a more critical posture in the next chapter). For Brentano, this must be based on certain deliverances of inner perception: say, that preference for knowing whether 2+2=4 over ignorance as to whether 2+2=4 is a self-imposing preference, a manifestly fitting preference. Brentano may suggest that when we reach such starkly fundamental pro attitudes, to do with intrinsic goodness, we simply cannot inner-*mis*perceive self-imposingness. It is only with respect to much more nuanced pro attitudes, to do with instrumental goodness, that matters start getting obscurer and we lose our ironclad grip on fittingness. A Hutu's pro attitude toward the Tutsi genocide may seem to her fitting, when in fact unfitting, but by Brentano's lights,

genocide is probably not *intrinsically bad* (because, as we will see in the next chapter, life is probably not intrinsically, but only instrumentally, good). Accordingly, one must perform various inferences to realize the badness of genocide; the badness of genocide does not impose itself on us the way the badness of (say) simple pain does. That is why the fittingness of such complex attitudes, to do with instrumental goods, must be derived from the more fundamental *manifest* fittingness of much more straightforward pro attitudes to do with intrinsic goods.

The essence of this response is that Huck Finn and Himmler are not counterexamples to a sufficiently restrictive conception of the domain in which we can unquestionably trust our conscience, or our first-person impression of fittingness. However, the original objection had more to it than counterexamples. Even if we restrict ourselves to ostensibly self-imposing attitudes about intrinsic goods, there is still the principled gulf between (a) a psychological impression of self-imposition and (b) real self-imposition in all its normative glory. As we have seen in the first half of this subsection, there are good reasons to think that in truth we can only inner-perceive our attitudes' *feeling* self-imposing. Brentano must argue that when we reflect on what is intrinsically good, the alleged daylight between feeling and being self-imposing starts to evaporate. But this would need to be shown, and I know of nothing in the Brentano corpus that even tries to show it.

A third response to the 'wishful thinking' objection, decidedly non-Brentanian, is to separate the issues of *acquiring* the concept of fittingness (or self-imposition) and *applying* that concept. Consider unlucky Luke, whose life thus far has been subjected to the following double predicament: whenever Luke has encountered a red object, it was through a thin (and very localized) mist that made the object appear pink rather than red; and whenever Luke has encountered a pink object, a strong (and very focal) red light in the vicinity made the object look red rather than pink. For Luke, vision has been an extraordinarily unreliable guide to the distribution of reds and pinks in his environment. Nonetheless, Luke has no other grasp of the *natures* of red and pink but that afforded to him by visual encounter.[38] Moreover, arguably Luke has an *accurate* grasp of the natures of red

44

and pink, despite having a perverse view of their *extensions*. In consequence, Luke is a full possessor of the concepts RED and PINK, misapplying them systematically though he does.[39] Something tragically close to Luke's predicament may well be the human condition when it comes to moral fittingness. We misapply the concepts of fittingness and unfittingness routinely, because of blinding selfishness, social conditioning, wishful thinking, or just what Arendt (1963) called the banality of evil. Nonetheless, we have an accurate grasp of the *natures* of fittingness and unfittingness. Indeed, Himmler and Washington seem to have the very same concept of fittingness that you and I do, despite wildly misapplying it to some of their attitudes. So it is not implausible to hold that we all fully possess the concepts of fittingness and unfittingness, and possess them thanks to inner-perceptual encounter with our (manifestly) fitting and unfitting attitudes – despite inner perception's unreliable *application* of these concepts.

This response protects Brentano's theory from Baldwin's wishful thinking problem more effectively than the previous two, and I endorse it herewith. I note, however, that it sacrifices the role of inner perception of fittingness and self-imposition in providing us knowledge of their *extension*, and therefore its role in guiding our moral conduct. This is a concession Brentano would most emphatically refuse to make: it is clear from the text of *Origin* that his theory is supposed to account not only for knowledge of the *nature* of goodness and badness, but also for knowledge of their *extension* (see, e.g., §25 and §31). All the same, the track record of moral certitude among our conspecifics suggests that dropping the hope for moral guidance would be a wise one for Brentano to make. The upshot would be an account of what value talk comes down to that does not also deliver a moral epistemology: to say that something is good is to say that self-imposing (or manifestly fitting) attitudes toward it would be pro rather than con attitudes; but how we *know* whether a self-imposing attitude toward something would be a pro or con attitude, when we ourselves do not *have* self-imposing attitudes toward the thing, is a separate matter, requiring a separate account.

5. Moral Value, Aesthetic Value, and Brentano's Theory of Beauty

Moore's *main* objection to Brentano is that his fitting attitude account of value is too broad:

> It is certain that many things, e.g., inanimate beautiful objects, possess the quality of being worthy to be loved, in a higher degree than they possess that of 'rightness'; it may even be doubted whether they possess the latter at all. And it is our duty to effect that which is the most 'right' possible, not that which is most worthy to be loved. Though therefore we can agree with Brentano that everything which is good in itself is worthy to be loved, we cannot agree that everything which is worthy to be loved is good. (Moore 1903b: 116)

It is fitting to approve of El Greco's *Saint Martin and the Beggar*, but the painting is not morally good. According to Moore, this is because although it is fitting to approve of the painting, this does not impose on us a duty to bring it into existence or keep it in existence (in short: a duty to effect existence). So fitting approval cannot ground moral goodness.

As it stands, the objection is not probing. Brentano's is an account of value as such, not of *moral* value. It addresses the metaphysics of goodness, not of moral goodness. Clearly, El Greco's *Saint Martin and the Beggar is* good, though aesthetically rather than morally so. This does raise the question, however, of how Brentano intends to deliver two distinct accounts of moral and aesthetic goodness (i.e., beauty).

There seem to be two general options within the fitting-attitude framework: either we distinguish two kinds of fittingness or we distinguish two kinds of pro attitude. The first option is to distinguish fittingness$_1$ and fittingness$_2$, such that (i) x is morally good iff it is fitting$_1$ to have a pro attitude toward x and (ii) x is beautiful iff it is fitting$_2$ to have a pro attitude toward x. The difference between moral and aesthetic goodness is then a matter of the difference between moral and aesthetic fittingness.

Might Brentano take this approach to account for the difference between moral and aesthetic value? In some respects, the approach is congenial: Brentano could suggest that the difference between these two kinds of fittingness, too, is a difference manifest to inner perception. Indeed, one can take the evident contrast between the fittingness of helping an old lady cross the street and the fittingness of admiring *Saint Martin and the Beggar* as further material for inner-perceptual zeroing in on the kind of fittingness relevant to the constitution of *moral* goodness.

There are two problems with this, however, as an interpretation of Brentano. First, to my knowledge Brentano nowhere distinguishes between moral and aesthetic fittingness. This is significant, since he *does* explicitly distinguish between the fittingness characteristic of interest states ('axiological fittingness,' we may call it) and the fittingness characteristic of judgments ('doxastic fittingness'):

> [There is] a fitting loving or hating and an unfitting loving or hating. This may seem like the analogue of fitting acceptance and fitting rejection, but it is essentially different (*wesentlich anderes*). (Brentano 1907a: 148-9 [144])

The fact that no similar statement is offered for moral and aesthetic fittingness suggests that Brentano does not intend to distinguish moral and aesthetic goodness in terms of two kinds of fittingness.[40]

Secondly, the *reason* the difference between axiological and doxastic fittingness is required is that there is a different standard of fittingness for such essentially different states as interest and judgment. The standard of fittingness for interest states is grounded in the fact that they essentially present-as-good/bad; that for judgment states in the fact that they essentially present-as-true/false. So even if we distinguished between moral and aesthetic fittingness, presumably as two types of axiological fittingness, that distinction would not be basic. It would be grounded in a more basic distinction in the kind of pro attitude we can take toward objects: those that present-as-morally-good and those that present-as-aesthetically-good.

Ultimately, then, the distinction between moral goodness and aesthetic goodness, or beauty, must come down to two different kinds of attitude. More precisely, there must be two different classes of mental state, M_1 and M_2, such that (i) *x* is morally good iff it is fitting to harbor M_1-type states toward *x* and (ii) *x* is beautiful iff it is fitting to harbor M_2-style states toward *x*.[41] This is indeed the path Brentano chooses. To appreciate how this works, I now turn to present what I take to be Brentano's theory of beauty.

❧

Existence and goodness are constitutively tied to fitting belief-in and fitting pro attitude in Brentano's system. But Brentano's psychology recognizes a third fundamental type of mental state, presentation or contemplation (*Vorstellung*). So might beauty be constitutively tied to fitting presentation or contemplation? After all, plausibly, a beautiful thing is *worthy of contemplation* in somewhat the same sense as a good thing is worthy of being approved of and an existent merits being believed in. The idea, then, might be that something is beautiful just if it is fitting to contemplate it – to train our awareness on it.

This does have the appearance of a vaguely Brentanian account of beauty, but it is *not* in fact Brentano's account. The basic reason is that the constitutive link between existence and fitting belief-in depends on the special attitudinal character of belief-in, which I have winkingly characterized as presenting-as-existent (and similarly for the link between goodness and fitting pro attitude). If the beautiful were simply that which it is fitting to contemplate, contemplation would have to exhibit the right attitudinal character as well – what we might call presenting-as-beautiful. But as we saw in Chap. 3, *Vorstellung* is special in exhibiting a completely neutral attitude of mere-presenting. It is thus *not* beauty-committal in the way belief-in is existence-committal and pro attitude is goodness-committal.

This problem with a fitting-contemplation account of beauty points the way to a better Brentanian account, which is indeed Brentano's. The right strategy is to identify a mental state attitudinally beauty-committal and assay beauty in terms of

the fittingness of adopting that state. Brentano calls the relevant state 'delight' (*Wohlgefallen*). Brentano's is a 'fitting delight' account of beauty, then:

> The concept of the beautiful has to do with… [that which] elicits in us a delight with the character of fittingness/correctness (*als richtig charakteriesiertes*). (Brentano 1959: 17)

So, to say that something is beautiful is to say that it is fitting to be delighted by it.

The question, however, is *what is delight*? In all of Brentano's taxonomic discussion, we have not encountered yet this mental state. If understanding the notion of beauty depends on understanding the notion of delight and its corresponding standard of fittingness, the first order of business is to elucidate the nature of delight. Brentano's account of delight starts from a simple observation about the experience of encounter with the beautiful: it always entrains a measure of enjoyment (Brentano 1959: 123). If one manages to contemplate El Greco's *Saint Martin and the Beggar* joylessly, one cannot be said to experience it as beautiful. This motivates Brentano to construe delight as a compound mental state involving both contemplation and joy – a kind of joyful contemplation. More specifically, delight at *x* involves (i) a contemplation of *x* plus (ii) an enjoyment *of the contemplation* of *x*. It is this compound state that, according to Brentano, embodies attitudinally encoded commitment to the beauty of its object: in the same sense in which a belief-in presents-as-existent and a pro attitude presents-as-good, delight is the state whose very essence is to present-as-beautiful. We may suppose that its contrary is a kind of dejected or wretched contemplation; we might call this 'dismay.' Dismay is the state whose essence is to present-as-ugly its object.[42]

Note that the joy component of delight is a pro attitude – joy belongs in the category of interest states. So delight is a compound state with a presentation component and a pro attitude component. As noted, the joy component is a *second-order* state: it is an enjoyment *of the presenting*. Thus to be delighted with *x* is to be in a state which is directed contemplation-wise at *x* and directed enjoyment-wise at the contemplation-of-*x*. In the same vein, we might say that to be dismayed with *x* is

49

to be in a compound state wherein one (i) contemplates *x* and (ii) feels dejection or 'disenjoyment' about this token contemplation.

The fitting-delight account of beauty provides Brentano with resources to respond to Moore's objection. The objection, recall, was that a fitting pro attitude account of moral value is too broad, since it is also fitting to have a pro attitude toward *Saint Martin and the Beggar*, which has no (intrinsic) moral value. The response is that in reaction to *Saint Martin and the Beggar*, what it is fitting to have a pro attitude toward is not the painting itself, but only the contemplation of the painting. The principled distinction between moral and aesthetic value can be drawn, then, in terms of two different kinds of pro attitude it is fitting to adopt in the presence of an object: a first-order pro attitude toward the object itself in the case of the morally good, and a second-order pro attitude toward awareness of the object in the case of the aesthetically good.

This account explains an interesting asymmetry between moral and aesthetic value pointed out by Brentano (1959: 136). Consider that peace is good, as is having a pro attitude toward peace, but mere contemplation of peace is in itself neither good nor bad. For contemplation of peace is compatible with both a pro attitude and a con attitude toward peace. By contrast, the mere contemplation of *Saint Martin and the Beggar* is in itself good. More generally, when an object is beautiful, contemplating it is valuable, whereas when the object is morally good, merely contemplating it is 'neutral' or 'indifferent' – only the adoption of a pro attitude toward it is valuable. (This recovers the sense in which the beautiful is worthy of contemplation.) This asymmetry between the morally good and the beautiful is *predicted* by Brentano's account of the difference between them. Given that a beautiful object merits delight, and a pro attitude toward one's contemplation of the object is a constituent of delight, it goes to the nature of beauty that a pro attitude toward contemplating the beautiful is appropriate. And since for something to be good is for a pro attitude toward it to be appropriate, it follows that the contemplation of a beautiful object is itself good. This flows out of the nature of delight, then – and the fact that delight in the beautiful is always fitting.

The account also explains why there is no sui generis kind of fittingness standard for delight. According to the account, delight is a compound state involving a first-order contemplation and a second-order pro attitude. As noted in Chap. 3, there is no standard of fittingness for contemplation, as mere-presenting is never fitting or unfitting. So the first-order component of delight introduces no standard of fittingness. The second-order component does introduce a standard of fittingness, but it is simply the standard of pro attitudes, not some new and sui generis standard. The standard is met when the first-order contemplative state is indeed good:

> Only when a presentation/contemplation is in itself good and joyful (*erfreulich*) do we call its primary object beautiful. (Brentano 1959: 123)

Accordingly, the delight fittingness can be accounted for in terms of self-imposition, as follows: it is fitting to be delighted with *x* iff anyone who contemplated *x* and harbored a self-imposing attitude toward that contemplation would harbor a *pro* (rather than con) attitude toward that contemplation.

The account still requires some refinement, however. For while every instance of aesthetic delight may involve contemplation of an object and enjoyment of that contemplation, the converse does not seem to hold. Consider the following joke: two muffins are baking in the oven; one says to the other, 'Man, it's getting hot in here'; the other says, 'What, you can talk too?!' Consider now your state of faint amusement at this joke. The amusement involves a contemplation of the scenario described in the joke, as well as mild enjoyment of this contemplation. So a combination of first-order contemplation and second-order pro attitude characterizes amusement (and its attitudinal character of presenting-as-funny) just as much as delight (and its attitudinal character of presenting-as-beautiful). Also, did you know that the southernmost point in Canada lies south of the northernmost point in California? If you found this tidbit interesting, you probably just experienced a state of curiosity or interest (in the narrow, non-Brentanian sense). Plausibly, your state of interest involved both contemplation of the fact relayed and

a measure of pleasure in that contemplation (it *felt good*). So something must distinguish aesthetic delight from such structurally similar states as amusement and interest. A fully developed Brentanian account of delight would tell us what.

Conclusion

Brentano's theory of goodness proceeds in two main steps: a reductive analysis of goodness in terms of self-imposing pro attitudes, and an assisted revelation account of self-imposition. I have attempted to present this position in a sympathetic light, as an elegant, well thought out, and reasonably defensible account of value. At the same time, I have argued that Brentano's theory of goodness is ultimately destabilized by the ambiguous status of self-imposition as a normative and the same time brutally psychological characteristic of mental states.

The discussion has led us to Brentano's account of the beautiful, which complements rather nicely his accounts of the true and of the good. We can appreciate an unmistakable systematicity in Brentano's three accounts. Consider these three statements:

> We call something true when the acceptance [belief] concerned with it is correct/fitting (*richtig*) (1889: 19 [18])

> We call something *good* when the love [pro attitude] concerned with it is correct/fitting. (Ibid.)

> The concept of the beautiful has to do with... [that which] elicits in us a delight with the character of fittingness/correctness (*als richtig charakteriesiertes*). (1959: 17)

Thus truth, goodness, and beauty are all accounted for in terms of the fittingness of certain distinctive mental states.[43]

In one sense, Brentano is a perfect realist about the true, the good, and the beautiful – whether something exists, is good, or is beautiful is not up to us. Nonetheless, when we say that something exists, is good, or is beautiful, all we are

saying is that belief in, pro attitude, or delight is the right attitude to take toward that thing. It is just that whether it is fitting or not to take the relevant attitude is an 'objective,' observer-independent matter. This is so even though it is impossible to truly *grasp* the very notions of the true, the good, and the beautiful without prior grasp of belief-in, pro attitude, and delight (and their fittingness). This is why Brentano wrote the following already in his mid-thirties:

> We see that … the triad of ideals – the beautiful, the true, and the good – can well be explicated in terms of the system of mental faculties. Indeed, this is the only way in which it becomes fully intelligible (*erklären*)… (Brentano 1874: II, 122 [263]; my italics)

In this way, in Brentano's thought the philosophy of mind assumes the role of *first philosophy*. And since for Brentano all mental life is conscious, it is the philosophy of consciousness that serves as first philosophy in his system. This is so in the following sense: the understanding of the true, the good, and the beautiful that the system offers is ultimately grounded in the understanding it offers of our conscious life. This is why we opened this book with discussion of Brentano's theory of consciousness.

Despite this *methodological* primacy of the philosophy of consciousness, let us repeat, Brentano's picture of the world is thoroughly realist. In fact Brentano's world contains just so many individual objects – nothing more. When we say, of any of the concrete particulars inhabiting Brentano's world, that it exists, or is good, or is beautiful, we are just saying that it would be fitting to believe in it, have a pro attitude toward it, or delight in it (respectively). It is in this way that the notions of the true/real, the good, and the beautiful enter our worldview. This entry does not entrain, however, a transcendental mind that does the believing, approving, or delighting. Rather, among the individual objects inhabiting this austere world are individual minds, including believing-minds, approving-minds, and delighted-minds, and indeed correctly-believing-minds, rightly-approving-minds, and fittingly-delighted-minds. It is because (and only because) each of us has on occasion *been* a correctly-believing-mind, rightly-approving-mind, or fittingly-delighted-mind, and

has *inner-perceived* himself or herself to be such a mind, that each of us is able to experience the world in terms of truth, goodness, and beauty.[44]

[1] I go for these rather unlovely labels, as opposed to the more straightforward 'objectivism' and 'subjectivism,' because the latter come with too much baggage. Certainly 'objectivism' can mean two very different things – that what is (intrinsically) good or bad is absolute and subject-independent, or that there are worldly facts concerning what is (intrinsically) good or bad – one of which Brentano was a fervent proponent of and the other something he was a clear opponent of. More on this later.

[2] A related problem is that NN_1 appeals to the notion of a 'natural science,' but what makes a science natural is presumably that its subject matter is part of nature – whereas the question we are trying to answer is precisely what makes a phenomenon part of nature.

[3] It may well be that only spatiotemporal entities are reductively explicable in terms of spatiotemporal entities, in which case the reductive explanation clause in NN_2 is redundant. But it is worth keeping the clause in, in case there is a relevantly useful notion of reductive explanation by the lights of which some a-spatiotemporal entities can be reductively explained in terms of spatiotemporal ones.

[4] In addition, they all appeal to the notion of reductive explanation, which has courted considerable controversy. Nonetheless, we can all agree on paradigmatic instances of reductive explanation – either in the form of a priori reductive analysis or in the form of a posteriori empirical reduction, such as of water in terms of H_2O, heat in terms of mean kinetic energy, and so on. We can then use these paradigms for an intuitive grasp of what reductive explanation is.

[5] Note that a subject-end answer can be 'objectivist' in the sense of demanding that approval of generosity be good universally and context-independently, and be not 'up to the subject.'

[6] This is not to deny that fitting-attitude accounts have had a presence in twentieth-century metaethics (see Ewing 1939, Chisholm 1981b and 1986, Lemos 1989 and 1994, Anderson 1993, Mulligan 1998). But it was clearly a recessive strand in metaethical thinking during the twentieth century.

[7] Here 'because' is intended not in a causal sense, but in an 'in-virtue-of' sense. FA thus grounds the goodness of things (e.g., generous acts) in the fittingness of pro attitudes (e.g., approval) toward them.

[8] The book's title received this English translation in both translations: the first translation in 1902 by Cecil Hague and the 1969 translation by Chisholm and Elisabeth Schneewind. There would be some case for the tidier *On the Origin of Moral Knowledge* as translation of *Vom Ursprung sittlicher Erkenntnis.*

[9] Some have argued that fittingness turns out to be whatever makes FA come out true, thus trivializing FA (see Reisner 2009).

[10] In §26, Brentano ends his discussion of the contrast between blind and self-evident judgments thus: 'everyone experiences the difference between these two manners of judging (*Urteilsweise*) …

[and] the ultimate elucidation [of self-evidence] consists only in a reference to this experience' (Brentano 1889: 21 [20]) – that is, the personal experience of *grasping* the contrast in question. But here the point is raised merely to introduce the pro attitude analog in §27. The section titles have sadly disappeared from the English edition, but in the German §26 is entitled 'Blind and Insightful Judgments' and §27 'Analogous Distinction in the Realm of Pleasure and Displeasure.'

[11] Brentano offers no name for this feature that serves as the Archimedean point of his moral philosophy!

[12] Note an important and potentially confusing difference between self-evidence and self-imposition. Self-evidence is revealed in (suitably aided) inner perception, but it is also *exhibited by* inner perception. Thus inner perception self-evidently reveals self-evidence. In contrast, inner perception reveals self-imposition but does not exhibit it – only interest states exhibit it. Accordingly, (suitably aided) inner perception self-evidently reveals self-imposition, but it would be a category mistake to say that it self-imposingly reveals self-imposition.

[13] I incorporate different examples into Brentano's thought experiment than his, because his presuppose his own first-order ethical commitments, some of which are controversial (we will discuss them in Chap. 9).

[14] I translate the key locution *als richtig charakterisierte* as 'with the character of correctness.' This is in contrast with the translation Chisholm has imposed throughout the Brentanian corpus, 'experienced as being correct,' which seems to me excessively interpretive. Hague's 1902 translation remains much closer to the text in offering 'having the character of rightness.'

[15] This second possibility is the analogue of non-self-evident but fitting belief: belief in duck is fitting, recall, when it is such that had the subject judged on the existence of ducks with self-evidence, she would self-evidently *believe* in (rather than *disbelieve* in) ducks.

[16] In this version of the view, the cashier case helps us home in on the nature of fittingness in its own way, although Brentano's case helps *more*. Both provide inner-perceptible contrasts between fitting and unfitting attitudes, though one does so using a finer contrast. The point is that the direct-grasp approach, when applied to fittingness itself rather than self-evidence or self-imposition, need not be based on a privileged type of contrast. Various different contrasts can be used to shed light from various angles on the single primitive, unanalyzable, fundamental feature that is fittingness.

[17] This formulation in terms of *facts* is of course for convenience. As we saw in Chap. 6, Brentano rejects the existence of facts. Ultimately, for Brentano 'If anyone had a self-imposing attitude toward generosity, it would be a pro attitude' must be paraphrased into 'There is no person who takes a self-imposing attitude toward generosity and does not take a pro attitude toward generosity' (or, if we are happy to bask in inelegance, 'There is no non-generosity-pro-attitudinizing self-imposingly-generosity-attitudinizing person'). Since this a negative truth, it does not have a truthmaker, that is, an entity in the world that makes it true. All we can say here is that it is fitting to disbelieve in the kind of person (a person who takes a self-imposing attitude toward generosity and does not take a pro attitude toward generosity).

[18] This passage appears at the beginning of §4 of Chap.1 of Bk. II of the *Psychology* – the chapter concerned with the demarcation of the mental. By the end of the section, Brentano recommends seeking a different criterion than that of spatiality. However, the reasons for this seem to be heuristic rather than substantive. His reservations about the spatiality criterion are two: that it is a purely negative characterization, and that it is controversial (whereas a demarcation of the mental should command consensus). Still, Brentano does seem to think that this criterion is just as extensionally adequate as the one he ultimately favors, intentionality. It is just that he thinks it is problematic for

the community of inquiry to adopt a criterion whose extensional adequacy not everybody appreciates.

[19] It is perhaps noteworthy that, despite his devout Catholicism, Brentano explicitly wields his open-question argument against divine command theory (more on this in §3.2).

[20] Brentano makes a similar point regarding the parallel question of why we should be believe what is self-evident: he claims that such a question would be 'completely laughable,' that is, does not really remain open (Brentano 1889: 21 [20]).

[21] What is the 'pertinent' sense? The sense the FA theorist needs, that is, the sense needed for grounding goodness in fitting attitudes. There may be a sense in which it is fitting to adopt a pro attitude toward someone who will kill everybody unless we adopt a pro attitude toward her – but that is not the sense of fittingness that could be then used in an FA account of goodness.

[22] Perhaps the most promising approach has been to claim that in the demon scenario, we actually do not have a reason to admire the demon; we only have a related but different reason – a reason to *bring it about* that we admire him, say, or a reason to *want* to admire him (Skorupski 2010, Way 2012). It is questionable whether this response can be fully general, however. Part of its attraction is due to the fact that admiration is not entirely in our control: I cannot simply up and decide to be an admirer of so-and-so. All I can decide is to *try* to become an admirer. But for all that, there may be some other pro attitudes which are directly controllable. If it is true that some pro attitudes we can *adopt*, then we could also *decide to adopt* them and they would be under our control. The other big approach to WKR seeks a further specification of the *kind* of reasons we should appeal to in an account of value, a kind the demon does not provide us with. Here the problem is how to do without circularity. One prominent idea is to distinguish between content-based (or 'object-given') and attitude-based (or 'state-given') reasons for holding an attitude (Parfit 2001, Danielsson & Olson 2007). This faces a number of problems, however, including returning the wrong results in certain cases (Rabinowicz & Rønnow-Rasmussen 2004, Schroeder 2012).

[23] According to Scanlon (1998), a reason for φ-ing is a consideration that count in favor of φ-ing. The buck-passer could, of course, appeal to this elucidation of reason talk, but then the same worry will arise for the notion of favoring. We can postpone the point at which we confess to our primitives, but eventually we will reach it!

[24] Actually, Kauppinen's view is that it is fitting for a subject to hold attitude A toward *x* just if an ideal observer would endorse holding A toward *x*. On the assumption that an ideal observer who *endorses* A would actually *adopt* A, however, this leads to the idea that an attitude is fitting when an ideal subject would adopt it.

[25] For Kauppinen, this sort of idealization serves two purposes. First, it undermines Enoch's (2005) accusation that idealization moves are ad hoc and unmotivated. Secondly, it provides for a principled speciation of value: different kinds of idealization avoid different kinds of practical problems, corresponding to moral, aesthetic, and prudential species of value.

[26] The idealized triggering conditions mean that the disposition may not be *manifested* in the actual world; all, the same, it may be routinely *instantiated* in the actual world. Dispositions are often instantiated without being manifested, as when fragility is instantiated by a vase which never breaks.

[27] One option is to hold that an attitude is fitting if at least one ideal observer would adopt it; this is likely to result in more attitudes counting as fitting than we are intuitively inclined to accept. Another option is to hold that attitudes are fitting only if *all* ideal observers would adopt them; this will likely result in fewer attitudes counting as fitting than we would intuitively accept. A third option is to hold that when ideal observers disagree it is *indeterminate* whether the attitude in question is fitting; this

is likely to result in extraordinarily many cases of fittingness indeterminacy – many more than we would intuitively accept.

[28] Pre-Kripke, one might object that if 'Goodness = C' is true, then it is necessarily true, since morality is necessary (there is no possible world where torturing babies for fun is good). Post-Kripke, however, the ethical naturalist can concede this and still maintain that 'Goodness = C' is true – it is a Kripkean a posteriori necessity akin to 'Water = H_2O.'

[29] It is, of course, an empirical truth, discovered by chemists, that the causal role of water is played in the actual world by H_2O; but this empirical truth merely sustains the necessity 'produced' by (1). For the only reason (2) is necessary is that it is explicitly restricted to the actual world (in every possible world it is true *of our world* that H_2O plays the water role in it) – whereas 'Necessarily, Water = H_2O' is not thus restricted.

[30] Retreating to (1*), 'Necessarily, goodness = the occupant of the goodness role,' is problematic. For it would lock up with (2*), C = is the occupant of the goodness role, but (2*) is not necessary, since it is not explicitly restricted to the actual world (it is not 'rigidified'). Accordingly, all that would follow is that goodness is C in the actual world – that is, that C is the realizer of goodness in the actual world. No identity claim could be established.

[31] It is worth noting that the father of Austrian market economics, Carl Menger, was a colleague of Brentano's at Vienna – and the two seem to have read each other (see Smith 1986).

[32] It might be insisted that the intuition that it is fitting to disapprove of genocide because genocide is bad goes beyond the intuition that it is fitting to disapprove of genocide because of what genocide is like. But this is far from obvious, and in any case this extra intuitive content is something that Brentano could flatly deny with much less embarrassment.

[33] Naturally, this is not a universal position. A conception of intuition as experiential or quasi-perceptual encounter with abstracta is defended by Bealer (1998) and Chudnoff (2013) among others. For a nuanced variant developed specifically for the moral context, see Huemer 2005.

[34] Moore's review is really of the 1902 English translation of *Origin*, wherein *richtig* was translated as 'right' rather than as 'correct' (as it was in the 1969 translation).

[35] Note that adopting this move would deprive Brentano of his main reason for rejecting response-dependent accounts of value that appeal to perceptual responses, which was that there is no such thing as perception of value. But the subsidiary, open-question reason probably persists: just as a grand illusion view of color perception is very much on the table, so is a grand illusion view of value perception – which leaves open the question of whether we should endorse what moral perception presents us with.

[36] Butler writes: 'The mind can take a view of what passes within itself, its propensions, aversions, passions, affections, as respecting such objects and in such degrees, and of the several actions consequent thereupon. In this survey it approves of one, disapproves of another, and toward a third is affected in neither of these ways, but is quite indifferent. This principle in man by which he approves or disapproves his heart, temper, and actions, is conscience.' (Butler 1726: 13-4) Interestingly, like Brentano, Butler's elucidation of what he has in mind by conscience adverts centrally to a contrast case: 'Suppose a man to relieve an innocent person in great distress; suppose the same man afterwards, in the fury of anger, to do the greatest mischief to a person who had given no just cause of offence. To aggravate the injury, add the circumstances of former friendship and obligation from the injured person; let the man who is supposed to have done these two different actions coolly reflect upon them afterwards, without regard to their consequences to himself: to assert that any common man would be affected in the same way towards these different actions, that

he would make no distinction between them, but approve or disapprove them equally, is too glaring a falsity to need being confuted. There is therefore this principle of reflection or conscience in mankind.' (Butler 1726: 15)