



A DANGER OF DEFINITION: POLAR PREDICATES IN METAETHICS

BY MARK ALFANO

JOURNAL OF ETHICS & SOCIAL PHILOSOPHY

VOL. 3, NO. 3 | SEPTEMBER 2009

URL: WWW.JESP.ORG

COPYRIGHT © MARK ALFANO 2009

A danger of definition: Polar predicates in metaethics

Mark Alfano

*Odi et amo. Quare id faciam, fortasse requiris?
Nescio, sed fieri sentio et excrucior.*

- Catullus

Introduction

DEFINITIONS ARE USEFUL TOOLS. THEY can be used to introduce new terms, precisify extant terms or analyze intuitive concepts. In this paper I discuss a threat to the analysis of pairs of intuitive contraries like “tall”/“short,” “hot”/“cold” and “good”/“bad.” I use an example from the history of philosophy to show how independently defining each side of such a pair is apt to lead to contradiction, then point out an analogous problem for the response-dependence semantics for “good”/“bad” advanced by Prinz, as well as for the fitting-attitude semantics advocated by Blackburn, Brandt, Brentano, Ewing, Garcia, Gibbard, McDowell and Wiggins.

Explicit definitions have a characteristic form. To define a predicate “ F ,” one formulates a universally quantified biconditional, where “ $F(x)$ ” occurs alone on the left-hand side and a more or less complex proposition “ $\Phi(x)$ ” occurs on the right-hand side:

$$(1) (\forall x)(Fx \equiv \Phi(x))$$

Definitions may be criticized for being uninformative (Φ is not independently understood), though even uninformative definitions may be helpful when contextualized in a broader theory. Worse, definitions may be circular (Φ contains “ F ”). Even worse, though, *prima facie* contraries like “ F ” and “ $\text{un-}F$ ” may be independently defined in a careless way. If they really are contraries, a contradiction lurks just around the corner.

If the definitientia of “ F ” and “ $\text{un-}F$ ” are not themselves contraries, whenever their extensions overlap, a contradiction will crop up, as the following formalization shows:

$$(2) (\forall x)(Fx \equiv \Phi x)$$

$$(3) (\forall x)(\text{un-}Fx \equiv \Psi x)$$

$$(4) \neg(\exists x)(Fx \wedge \text{un-}Fx)$$

$$(5) (\exists x)(\Phi x \wedge \Psi x)$$

$$(6) (\exists x)(Fx \wedge \text{un-}Fx)$$

When one encounters such a contradiction, one has six options beyond outright dialetheism.¹ First, one can give up the assumption that the presumed contraries really are contraries, eliminating 4. Second, one can give up one of the two definitions; that is, one can reject either 2 or 3. Third, one can replace one of the definitions to make the contraries into pure contradictories, i.e., make $\text{un-}F$ equivalent to $\text{not-}F$:

$$(3') (\forall x)(\text{un-}Fx \equiv \neg Fx)$$

Fourth, one could keep the contraries but transform them into trouser words.² That is, one could say either

$$(3'') (\forall x)(\text{un-}Fx \equiv (\Psi x \wedge \neg Fx))$$

or

$$(2') (\forall x)(Fx \equiv (\Phi x \wedge \neg Fx))$$

Fifth, one can try to rejigger the definitions so as to keep both informative without leading to a contradiction. This strategy, however, leaves the theorist vulnerable to empirical refutation if it is discovered that the new definienda are co-instantiated. Finally, one can attack the evidence for co-instantiation.

1. Piety and impiety in the *Euthyphro*

In the *Euthyphro*, Socrates points out that the definition of piety as that which is loved by the gods leads to a contradiction because the gods quarrel (Plato 1941, 6e-8a). We can reconstruct his *reductio* as follows:

- (7) Something is pious just in case there is some god that loves it, i.e.,
 $(\forall x)(Px \equiv (\exists y)(Gy \wedge yLx))$
- (8) Something is impious just in case there is some god that does not love it, i.e.,
 $(\forall x)(Ix \equiv (\exists y)(Gy \wedge \neg yLx))$
- (9) Nothing is both pious and impious, i.e., $\neg(\exists x)(Px \wedge Ix)$
- (10) The gods quarrel, i.e., $(\exists x)(\exists y)(\exists z)(x \neq y \wedge Gx \wedge Gy \wedge xLz \wedge \neg yLz)$
- (11) Hence, there is something (the thing that the gods quarrel about) that is both pious and impious, i.e., $(\exists x)(Px \wedge Ix)$

To rectify this pair of definitions, Euthyphro changes his definition of piety to

¹ See Priest (2005).

² See Austin (1964, p. 15).

- (7') Something is pious just in case all gods love it, i.e.,
 $(\forall x)(Px \equiv (\forall y)(Gy \supset yLx))$

While he does not explicitly redefine impiety, presumably he intends something like

- (8') Something is impious just in case no gods love it, i.e.,
 $(\forall x)(Ix \equiv \neg(\exists y)(Gy \wedge yLx))$

Clearly, the deduction from 7', 8' and 10 to 11 is no longer valid. Euthyphro therefore follows the fifth method canvassed above – rejiggering both definitions.³

2. Good and bad in the response-dependence theory

A similar line of argument can be used to show that the response-dependence semantics for the moral predicates “good” and “bad” leads to a contradiction. According to response-dependence theory, something is good (bad) just in case we are disposed to have a positive (negative) moral sentiment towards it upon careful reflection.⁴ Response-dependence jibes nicely with moral relativism, but the two are in principle independent. To prevent relativistic issues from clouding the discussion, I shall assume a one-agent universe in the balance of this section; this means that saying that x is good for *someone* is equivalent to saying that x is good for *everyone*. The argument can be easily reconstructed in a multi-agent universe. The parallel argument runs as follows:

- (12) Something is good just in case some agent is disposed to have a positive moral sentiment towards it upon careful reflection, i.e.,
 $(\forall x)(Gx \equiv (\exists y)(Ay \wedge yPx))$
- (13) Something is bad just in case some agent is disposed to have a negative moral sentiment towards it upon careful reflection, i.e.,
 $(\forall x)(Bx \equiv (\exists y)(Ay \wedge yNx))$
- (14) Nothing is both good and bad, i.e., $\neg(\exists x)(Gx \wedge Bx)$

³ Curiously, however, the truth of atheism would still commit Euthyphro to a contradiction. If there are no gods, then according to 6' and 7', everything is both pious and impious. In parallel, if there are no gods, then according to 6 and 7, nothing is pious and nothing is impious. Perhaps this helps to explain Ivan's intuition in *The Brothers Karamazov* that if there is no God, all is permitted.

Another curious feature of Euthyphro's revised definitions is that if the gods are maximally quarrelsome (if there is nothing they all love and nothing they all fail to love), then nothing is pious and nothing is impious.

⁴ These analyses are culled almost verbatim from Prinz (2005); see Prinz (2005, 2006, 2007a, 2007b). For response-dependence theory more generally, see Dreier (1990), Johnston (1989), McDowell (1985), McNaughton (1988), Wiggins (1991) and Wright (1992).

- (15) Ambivalence exists; in other words, some things are such that an agent is disposed to have both positive and negative moral sentiments towards them upon careful reflection, i.e., $(\exists x)(\exists y)(Ay \wedge yPx \wedge yNx)$
- (16) Some things are both good and bad, i.e., $(\exists x)(Gx \wedge Bx)$

12 and 13 together constitute the response-dependence theory of moral terms. 14 is based on the intuition that “good” and “bad” are contraries, and so is intended to be uncontroversial. 15 is an empirical psychological claim that I will try to substantiate through examples. 16 clearly follows from 12, 13 and 15, yielding a contradiction with 14.

Response-dependence semantics analyzes moral properties on the analogy of secondary properties. According to the response-dependence account of, for instance, the predicate “red” as applied to surfaces of objects, an object is red just in case we are disposed to have a red percept in its presence in normal circumstances. To be physically red is to be the sort of thing to cause the perception of phenomenal redness in normal circumstances.⁵ By analogy, then, to be good (bad) is to be the sort of thing to cause good feelings (bad feelings) in an agent after careful reflection. According to response-dependence theorists like Prinz (2005, p. 101), the moral terms “good” and “bad” express a speaker’s sentiments, where sentiments are understood as “dispositions to experience different emotions on different occasions.” Hence, “To judge that something is *good* (or *bad*) is to believe that on reflection it would be recognized as an example of something towards which one harbors a [...] moral sentiment.” The careful reflection rider aims to eliminate knee-jerk reactions, which are clearly corrigible. The idea is that judgments of good and bad are best detected in Bishop Butler’s cool hour.⁶

15 seems to be the best place to dig in one’s heels if one wants to resist my argument; that is, the best way to retain both 12 and 13 without endorsing a contradiction is the sixth method canvassed above – attacking the evidence. Simply to deny that ambivalence exists, however, is a non-starter. We need to distinguish ambivalence (having contrary evaluative dispositions) from indifference (having no evaluative dispositions). Catullus’ “*Odi et amo*” seems to me a telling counterexample to the unguarded claim that ambivalence does not exist. A more incisive riposte is required. One way of accepting the existence of ambivalence while avoiding the sting of *modus tollens* would be to say that, in ambivalence, the object of approbation is not the

⁵ I recognize that this analysis of “red” is subject to quibbles from almost every quarter. Its purpose is merely to illustrate the sort of account response-dependence theories provide, not to accurately represent Locke’s or Hume’s or Strawson’s or anyone’s actual theory of color terms.

⁶ Response-dependence semantics for moral terms must be carefully distinguished from emotivism, the thesis that moral terms *merely* express sentiments and, hence, that sentences containing them are not truth-evaluable (see Ayer (1990)). According to the “sensitivity theory” of Prinz, by contrast, “Moral judgments are truth-apt [...] they refer to response-dependent properties, just as their surface form would suggest” (2006, p. 35).

same as the object of disapprobation. “Love the sinner, hate the sin,” as it were.⁷ While I concede that this strategy may succeed in some cases, I doubt it will work across the board. Does Catullus love Lesbia but hate what she does? That seems to cheapen his verse.

Another potential dodge would be to deny that *moral* ambivalence exists. Perhaps we feel both approbation and disapprobation towards some things, but we never feel both *moral* approbation and *moral* disapprobation towards them. Someone might admire Karl Rove for his organizational and rhetorical prowess (amoral) while abhorring him for his motivation and ends (moral). Conversely, someone might admire a bonhomme for his benevolence yet scorn him for his dull-wittedness. The main challenge facing this strategy is distinguishing moral from amoral approbation (and moral from amoral disapprobation) without abandoning response-dependence semantics or talking in circles. For the sake of argument, let’s assume that challenge can be met. In that case, I again claim that it will work in some cases but not all. Catullus loves Lesbia. He hates her too. Both sentiments are as moral as can be.

Finally, one might deny that in ambivalence the same *person* harbors both sentiments. The lover and the hater, though they occupy the same body, are not of approbation all compact. This disintegration into two persons may result from failure, in Dennett’s (1992) terms, to share the same center of narrative gravity. Consider someone in the midst of a moral transition: he was raised to be homophobic but is becoming convinced that there is nothing wrong with homosexuality. Even upon careful reflection, he tends to feel both positive and negative emotions when he thinks about homosexual intimacy.⁸ To explain his behavior, we need to posit two different moral persons: the homophobe who hates gays (or homosexuality) and the liberal who tolerates them (or it). A similar idea can be found in Freud (1916-1917, p. 216). He compares the state of ambivalence to “an amalgamation of two separate people,” illustrating his view with the following fable:

A good fairy promised a poor married couple to grant them the fulfillment of their first three wishes. They were delighted, and made up their minds to choose their three wishes carefully. But a smell of sausages being fried in the cottage next door tempted the woman to wish for a couple of them. They were there in a flash; and this was the first wish-fulfillment. But the man was furious, and in his rage wished that the sausages were hanging on his wife’s nose. This happened too; and the sausages were not to be dislodged from their new position. This was the second wish-fulfillment; but the wish was the man’s, and its fulfillment was most disagreeable to his wife. You know the rest of the story. Since after all they were in fact one – man and wife – the third wish was bound to be that the sausages should come away from the woman’s nose.⁹

⁷ Or, conversely, hate the saint, love the saintliness.

⁸ I owe this idea to Prinz (2008). As he put it, “People in transition suffer from a kind of evaluative schizophrenia.”

⁹ This story structure has been wonderfully dramatized by W. W. Jacobs in “The Monkey’s Paw” (1902).

This Jekyll-and-Hyde (or, to vary the figure, Dorian-and-picture) strategy may succeed in some cases, but I doubt that many people are so hyper-ambivalent that their behavior can be explained only by positing two selves for each of them. Has Catullus endured psychological fission? In addition, it seems *ad hoc* to rescue a meta-ethical theory like response-dependence by committing oneself to a substantive theory of personal identity. Those who use this dodge cannot endorse the popular continuity-of-brain theory of personal identity.

These splitting strategies for avoiding the conclusion that some things are both good and bad share a common core. Since harboring a sentiment is a dyadic relation between an agent and an intentional object, there are three ways to weaken 15: splitting the right-hand relatum (the object), splitting the relation itself (the sentiment) and splitting the left-hand relatum (the agent). Splitting the object amounts to saying that the regimentation of 15 should be

$$(15') (\exists x)(\exists y)(\exists z)(Az \wedge zPx \wedge zNy \wedge x \neq y)$$

Splitting the relation amounts to saying that the regimentation should be either

$$(15'') (\exists x)(\exists y)(Ay \wedge yP_M x \wedge yN_{-M} x)$$

or

$$(15''') (\exists x)(\exists y)(Ay \wedge yP_{-M} x \wedge yN_M x)$$

Splitting the agent amounts to saying that the regimentation should be

$$(15''''') (\exists x)(\exists y)(\exists z)(Az \wedge Ay \wedge zPx \wedge yNx \wedge z \neq y)$$

Since there are only two relata and one relation in 15, these three splitting strategies are exhaustive. Moreover, they are mutually compatible, and together they may account for many, perhaps even most, *prima facie* cases of moral ambivalence. Provided, however, that there is a single instance of full-fledged moral ambivalence, my argument still goes through.¹⁰ Many readers' ambivalence towards Laura in Rossetti's *Goblin Market* seems to me a case in point. Trolley cases are another. Reactions to the story of Pero – the daughter who breastfeeds her father to keep him alive – yet another. Indeed, if the CAD hypothesis – according to which approbation (and disapprobation) are caused by emotions of contempt (for violation of societal rules), anger (for

¹⁰ Indeed, if it is even *possible* that there is a case of full-fledged moral ambivalence, the advocate of the splitting strategies is committed to saying that some contradictions are possible, i.e., $\Diamond \perp$. Such a result would be a Pyrrhic victory for defenders of response-dependence.

violation of individuals' rights) or disgust (for violations of nature)¹¹ – is correct, it should be an undergraduate textbook exercise to construct cases where an agent experiences approbation with respect to one dimension (filial piety, say) and disapprobation with respect to another (disgust, say).¹² The story of Pero seems to do just this, pitting her fulfillment of her duties to her father (invoking rules governing relations to society and to individuals) and her violation of nature. Note further that response-dependence theorists cannot wiggle their way out of this contradiction by pointing out that the cause of (or reason for) approbation (following social and personal norms) and the cause of (or reason for) disapprobation (violating a norm of nature) are different. Response-dependence theory – at least as characterized by 12 and 13 – makes no reference to causes or reasons.

One final way to resist the evidence would be to say that lack of ambivalence is a *necessary condition* for having carefully reflected (see 12 and 13). If this is right, the inference to 16 is blocked because 15 is false. This response, though it would render my argument unsound, strikes me as *ad hoc*. It disallows ambivalence by *fiat*, even though the question whether ambivalence exists seems empirical in nature. Psychologists certainly theorize as if ambivalence were a real phenomenon.¹³

From these considerations I conclude that attacking the evidence will not rescue response-dependence. Another option for dealing with such a dilemma is to deny that the intuitive contraries really are contraries. In this case, that would mean rejecting 14. I expect this tactic would be at best cold comfort to defenders of response-dependence. If one has already gone so far as to say that some things are both good and bad, one might as well go whole-hog and accept dialetheism; if one could rescue response-dependence from contradiction only in this way, it would cease to be a plausible moral theory.

If I am right so far, the defender of response-dependence has four remaining options. First, she can abandon either 12 or 13 while retaining 14. That would mean being able to say when something is good (bad) but not when something is bad (good). Presumably, though, any moral theory worth the candle can say both. Kemp (1954) has argued convincingly that metaethics owes us an account not just of goodness, rightness, pleasure and utility, but also of badness, wrongness, displeasure and disutility. Second, the defender of response-dependence can replace one of the definitions to make the contraries into contradictories. In other words, she could say that something is good just in case it is not bad, or bad just in case it is not good. Again, such a move does not seem to me a live option.

¹¹ See Shweder et al. (1997) and Rozin et al. (1999).

¹² For better or worse, we should probably resign ourselves to thought experiments and anecdotes in constructing such examples. It would be a rare IRB indeed that would approve such an experiment.

¹³ See for instance Newby-Clark, McGregor & Zanna (2002).

Third, and more plausibly, she could turn the contraries into trouser words. In other words, she could say either

- (12') Something is good just in case it is not bad and some agent is disposed to have a positive moral sentiment towards it upon careful reflection, i.e., $(\forall x)(Gx \equiv (\neg Bx \wedge (\exists y)(Ay \wedge yPx)))$

or

- (13') Something is bad just in case it is not good and some agent is disposed to have a negative moral sentiment towards it upon careful reflection, i.e., $(\forall x)(Bx \equiv (\neg Gx \wedge (\exists y)(Ay \wedge yNx)))$

The difficulty of saying which of these is right results in an arbitrary and therefore *ad hoc* choice. Is goodness prior to badness or badness prior to goodness? Nietzsche (1989) would say it depends. I can think of no plausible arguments for opting for 12' over 13' or *vice versa*.

3. Good and bad in the fitting-attitudes theory

Ambivalence poses a similar problem for the fitting-attitudes semantics for the moral predicates “good” and “bad.” According to the fitting-attitudes theory, something is good just in case it would be fitting (appropriate) to take an approbative attitude (e.g., love, choice, desire pursuit, approval, admiration, liking) towards it and bad just in case it would be fitting (appropriate) to take a disapprobative attitude (e.g., hate, avoidance, disapproval, contempt, disliking) towards it.

Brentano (1969, p. 18) seems to be the originator of this theory, defining “good” as “that which is worthy of love, that which can be loved with a love that is correct.” He goes on to declare (p. 22) insight “worthy of love” and error “worthy of hate, and therefore that the one is good and the other bad.” Brandt (1941, pp. 50-51) claims that “‘admirableness’ [...] is] the property of being worthy of admiration, and ‘despicableness’ [...] the property of deserving to be despised.” Ewing (1959, p. 85) says that bad “is just what ought to be the object of an unfavorable attitude, as good is what ought to be the object of a favorable.” Garcia (1987, p. 394) claims “that everything that is, strictly speaking [...] good [...] is *also* desirable [...] and everything that is, strictly speaking, bad [...] is *also* undesirable.” McDowell (1998, pp. 131-150; see also 1985), in his characteristically enigmatic way, defines the valuable in terms of what merits a specific modification of the human sensibility and claims that *being morally right* is “best understood” as being such as to warrant a particular emotion. Presumably he also thinks that *being morally wrong* is best understood as being such as to warrant a different, contrary particular emotion. Blackburn (1984, 1993, 1998) argues that to make an evaluative judgment is to endorse a sensibility or evaluative attitude. In (1998, p. 106) he

defines “ x is good” with the open sentence “ x is such as to elicit approval from good people under the ideal circumstances.” Presumably, then, he also thinks that x is bad just in case x is such as to elicit disapproval from good people under ideal circumstances. Gibbard (1990) thinks that judgments of right (wrong) essentially involve judgments of warranted praise (blame). He argues (p. 42) that “what a person does is *morally wrong* if and only if it is rational for him to feel guilty for doing it, and for others to resent him for doing it,” and (p. 128) that “a thing is shameful if it makes sense to be ashamed of it and for others to disdain one for it. It is admirable if it makes sense for others to admire one for it, and for one to feel self-approbation because of it.” Finally, Wiggins (1991, p. 206) defines the open sentence “ x is good” by “ x is such as to make a certain sentiment of approbation appropriate,” and says (1990-1991, p. 69) that “sentiments of approbation/disapprobation” stand in “one-one correspondence with the diversity of thoughts that sustain them.” He even hints (1991, p. 205) that if the fittingness of approbation and disapprobation were somehow to change it would “make what is now right wrong, what is now good bad.”

Provided that ambivalence is sometimes fitting, the fitting-attitudes theory is committed to some things’ being both good and bad, as the following *reductio* shows:

- (17) Something is good just in case it would be fitting to take an approbative attitude towards it, i.e., $(\forall x)(Gx \equiv Ax)$
- (18) Something is bad just in case it would be fitting to take a disapprobative attitude towards it, i.e., $(\forall x)(Bx \equiv Dx)$
- (19) Nothing is both good and bad, i.e., $\neg(\exists x)(Gx \wedge Bx)$
- (20) Ambivalence is sometimes fitting; in other words, some things are such that it would be fitting to take both an approbative and a disapprobative attitude towards them, i.e., $(\exists x)(Ax \wedge Dx)$
- (21) Hence, something (the thing towards which ambivalence is appropriate) is both good and bad, i.e., $(\exists x)(Gx \wedge Bx)$

With the possible exception of splitting the agent, the dodges discussed in the previous section are available to the fitting-attitudes theorist, but, as before, I think these strategies ultimately prove insufficient and that the purely logical revisions are non-starters. I shall discuss only the one new move opened up by the differences between the response-dependence theory and the fitting-attitudes theory.

Whereas the response-dependence theory analyzes moral predicates in terms of dispositions of actual agents, the fitting-attitudes theory analyzes such predicates in terms of objective appropriateness. Goodness is instantiated not because *someone would* love but because *everyone should* love. The dodge discussed above – making lack of ambivalence a *necessary condition* for having carefully reflected – may be revived and revised here. The crux of the debate then becomes the account of fittingness: what makes (dis)approbation

fitting? I am not so heroic as to attempt to answer that question. I argue merely with particulars. Regardless of what in general makes approbative and disapprobative attitudes fitting, if I can make a case that ambivalence is sometimes appropriate, my argument goes through.

Consider the character Ximene in popular retellings of the exploits of El Cid. She loves El Cid passionately, yet he kills her father in a duel of honor. Her father's dying wish is that she wreak vengeance on El Cid, and she promises to fulfill that wish. Carr (2002, p. 4) has argued persuasively that ambivalence, far from being inappropriate, is the *only* appropriate emotional state for Ximene. More such examples are easy to spin off. An unfaithful lover causes immense pain precisely because she is still appropriately loved, even as fitting hatred for her grows. The Spartan mothers were reputed to tell their beloved sons to return from battle either with their shields or on them; presumably deserters were objects of intense yet fitting ambivalence.

Once again, it will not do to say that the *reasons* for Ximene's or the cuckold's or the Spartan mothers' appropriate ambivalence are different, for the fitting-attitudes theory as articulated makes no reference to reasons. By the lights of the fitting-attitudes theory, if it would be appropriate to love and appropriate to hate the same thing – that thing is both good and bad. Once again, then, an account of polar predicates founders on the rocks of ambivalence.

4. Letting “I dare not” wait upon “I would” / Like the poor cat in the adage

Lady Macbeth scolds her husband for letting “I dare not” wait upon “I would” like the cat in the adage that wants to eat the fish but does not dare to wet its paws in the act of capturing it. One can easily picture the conflicted beast, unable to bring itself to do the deed, tensely pawing at the water where the fish glides in placid ignorance. Explaining the cat's behavior necessarily involves reference to ambivalence, but also to the reasons for ambivalence: it would because it wants to eat but it dare not because it fears the water.

I have argued that the existence and appropriateness of ambivalence spell trouble for both the response-dependence and fitting-attitudes theories of good and bad. In both cases, however, a final argumentative move may suffice: rejiggering the definitions of “good” and “bad” to make explicit reference to reasons.¹⁴ Doing this will add another epicycle to the already doubly subjunctive response-dependence theory (being disposed to have a sentiment means being disposed to be disposed to feel an emotion). In addition, making explicit reference to causes or reasons is not in the spirit of sentimentalism, which aims to avoid the cognitive overload it decries in other moral theories. Referring to reasons may work better for the fitting-attitudes theory;

¹⁴ This move has been independently suggested for the fitting-attitudes theory, though to counter a different problem, by Rabinowicz & Ronnow-Rasmussen (2004).

its use of appropriateness conditions for emotions is already starkly rationalistic, so perhaps the infusion of a little more reason would do no harm. I leave it an open question whether such a project can be successfully carried out.

Mark Alfano
CUNY Graduate Center
Program in Philosophy
mark.alfano@gmail.com

Bibliography

- Austin, J. L. (1964). *Sense and sensibilia*. Warnock (ed.). Oxford: Oxford University Press.
- Ayer, A. J. (1990). *Language, Truth, and Logic*. New York: Penguin Books.
- Blackburn, S. (1984). *Spreading the Word*. Oxford: Oxford University Press.
- Blackburn, S. (1993). *Essays in Quasi-Realism*. Oxford: Oxford University Press.
- Blackburn, S. (1998). *Ruling Passions*. Oxford: Oxford University Press.
- Brandt, R. (1941). "An emotional theory of the judgment of moral worth." *Ethics*, 52:1, 41-79.
- Brentano, F. (1969). *Origins of Our Knowledge of Right and Wrong*. (trans.) Chisholm & Schneewind. New York: Humanities.
- Carr, D. (2002). "Feeling in moral conflict and the hazards of emotional intelligence." *Ethical Theory and Moral Practice*, 5:1, 3-21.
- Dennet, D. (1992). "The self as a center of narrative gravity." In Kessel, Cole, & Johnson (eds.) *Self and Consciousness: Multiple Perspectives*. Hillsdale, NJ: Erlbaum.
- Dreier (1990). "Internalism and speaker relativism." *Ethics* 101: 6-26.
- Ewing, A. (1959). *Second Thoughts in Moral Philosophy*. London: Routledge & Kegan Paul.
- Freud, S. (1916-17). "Introductory Lectures on Psycho-Analysis." *The Standard Edition of the Complete Psychological Works of Sigmund Freud*, Volume XV.
- García, J. (1987). "Goods and evils." *Philosophy and Phenomenological Research*, 47:3, 385-412.
- Gibbard, A. (1990). *Wise Choices, Apt Feelings*. Cambridge: Harvard University Press.
- Jacobs, W. W. (1906). "The monkey's paw." In *The Lady of the Barge*. New York: Harper & Brothers.
- Johnston, M. (1989). "Dispositional theories of value." *Proceedings of the Aristotelian Society* 63 (Supplement), 139-74.
- Kemp, J. (1954). "Pain and evil." *Philosophy*, 29:108, 13-26.
- McDowell, J. (1985). "Values and secondary qualities." In Honderich (ed.) *Morality and Objectivity*. London: Routledge & Kegan Paul.
- McDowell, J. (1998). "Projection and truth in ethics." In *Mind, Value, and Reality*, 151-166. Cambridge: Harvard University Press.
- McNaughton (1988). *Moral Vision: An Introduction to Ethics*. Oxford: Blackwell.
- Newby-Clark, I., McGregor, I. & Zanna, M. (2002). "Thinking and caring about cognitive inconsistency: When and for whom does attitudinal ambivalence feel uncomfortable?" *Journal of Personality and Social Psychology*, 82:2, 157-166.
- Nietzsche, F. (1989). *The Genealogy of Morals*. Kaufmann (trans.). New York: Random House.
- Plato (1941). *Euthyphro*. Cooper (trans.). In Hamilton & Cairns (eds.) *The Collected Dialogues of Plato* (1999). Princeton: Princeton University Press.
- Priest, G. (2005). *Towards Non-Being: The Logic and Metaphysics of Intentionality*. Oxford: Clarendon.
- Prinz, J. (2005). "Passionate thoughts." In Pecher & Zwaan (eds.) *Grounding Cognition*. Cambridge: Cambridge University Press.
- Prinz, J. (2006). "The emotional basis of moral judgments." *Philosophical Explorations*, 9:1, 29-43.
- Prinz, J. (2007a). "Can moral obligations be empirically discovered?" *Midwest Studies in Philosophy*, 31, 271-291.
- Prinz, J. (2007b). *The Emotional Construction of Morals*. Oxford: Oxford University.
- Prinz, J. (2008). Personal communication. May 16.
- Rabinowicz, W. & Ronnow-Rasmussen, T. (2004). "The strike of the demon: On fitting pro-attitudes and value." *Ethics*, 114, 391-423.
- Rozin, P., Lowery, Imada, & Haidt (1999). "The CAD triad hypothesis: A mapping between three moral emotions (contempt, anger, disgust) and three moral codes (community, autonomy, divinity)." *Journal of Personality and Social Psychology* 76: 574-86.

- Schweder, R., Much, Mahapatra & Park (1997). "The 'big three' of morality (autonomy, community, and divinity), and the 'big three' explanations of suffering." In Brandt & Rozin (eds.) *Morality and Health*. New York: Routledge.
- Wiggins, D. (1990-1991). "Moral cognitivism, moral relativism, and motivating moral beliefs." *Proceedings of the Aristotelian Society*, 91, 61-85.
- Wiggins, D. (1991). "A sensible subjectivism." In *Needs, Values, Truth: Essays in the Philosophy of Value*. Oxford: Blackwell.
- Wright (1992). *Truth and objectivity*. Cambridge: Harvard University Press.