American Philosophical Quarterly
Volume 50, Number 1, January 2013

DOXASTIC VOLUNTARISM, EPISTEMIC DEONTOLOGY, AND BELIEF-CONTRAVENING COMMITMENTS

Michael J. Shaffer

1. Introduction

A number of epistemologists have defended doxastic voluntarism, the view that we have voluntary control over what we believe. Defenders of this view allow then for the possibility that we can voluntarily commit ourselves to propositions. This appears to include belief-contravening propositions. Typically, belief-contravening commitments are themselves taken to be beliefs. Thus, these defenders of doxastic voluntarism allow that we can at least sometimes efficaciously choose to believe propositions that are negatively implicated by our evidence. One of the main reasons that doxastic voluntarism enjoys a degree of popularity is that it is supposed to allow for the evaluation of an agent's epistemic status from a deontological perspective. In other words, one can conceive of epistemology as normative and based on categorical principles that define meritorious and nonmeritorious epistemic behavior. The basic idea that is alleged to support this view is that we think that it is appropriate to hold epistemic agents responsible for what they believe and if doxastic voluntarism is false, then it is inappropriate to hold epistemic agents responsible for what they believe. The conditional premise of this argument is supposed to follow from the epistemic analog of the Kantian moral principle that "ought" implies "can." Put more simply, if we cannot control our beliefs, then we are not responsible for our beliefs. If we are not responsible for our beliefs, then it would be inappropriate to evaluate our beliefs in terms of any deontological epistemic principles.

In this essay it will be argued that the conjunction of epistemic deontology and doxastic voluntarism as it applies to ordinary cases of belief-contravening propositional commitments is incompatible with canonical formulations of evidentialism—the view that one should only commit to a proposition for which one has adequate evidence. The combination of these views entails practical or deontic contradictions if belief-contravening commitments are understood normally (i.e., as beliefs). In this essay ED and DV will be assumed and this negative result will be used to suggest that voluntary belief-contravening commitments are not themselves beliefs and that these sorts of commitments are not governed by evidentialism. So the apparent incompatibility of the package views noted above can be resolved without ceding evidentialism with respect to beliefs.

©2013 by the Board of Trustees of the University of Illinois

2. Epistemic Deontologism and Evidentialism

Epistemic deontologism is the view that:

> (ED) Our beliefs are justified if, and only if, what we believe is epistemically permissible for us to believe. And our beliefs are unjustified if, and only if, what we believe is epistemically impermissible (Steup 2000, p. 25).

This view has enjoyed increasing prominence among contemporary epistemologists and it is prima facie plausible in that it treats epistemic evaluation as a normative endeavor based on categorical epistemic principles in very much the same way that moral evaluation is a normative endeavor based on categorical moral principles. So according to the epistemic deontologist we can lay out these principles, and following them entails that our beliefs will be epistemically meritorious.

One such principle is that of evidentialism. Evidentialism is the view that one's justification for a proposition is a function of one's evidence. It is canonically understood as follows:

> (EV) Doxastic attitude *D* toward proposition *p* is epistemically justified for *S* at *t* if and only if having *D* toward *p* fits the evidence *S* has at *t* (Feldman and Conee 2004, p. 83).

So EV is an epistemic norm that governs our doxastic attitudes including belief, and it implies that one should believe a proposition if and only if one has adequate evidence for that proposition. As it applies to belief, EV is the view that,

> (EV-B) The belief that *p* is epistemically justified for *S* at *t* if and only if believing *p* fits the evidence *S* has at *t*.

There are, of course, many other such norms that might constitute elements of the full ethics of belief, but for the purposes of this essay, we need only have this derived principle at our disposal.

3. Doxastic Voluntarism

Doxastic voluntarism is the view that whether we believe something is a voluntary matter. This view is importantly related to ED, because it is widely (although not universally) believed that it is appropriate to subject an agent's doxastic attitude(s) to evaluation in terms of the normative principles that constitute the ethics of belief only if holding those doxastic attitudes is a voluntary matter for S.[1] We can only say, for example, that S ought or ought not to believe that p, when S's belief that p is a voluntary matter. In other words, we can only evaluate S's belief in this manner when believing that p is up to S. So doxastic voluntarism is the view that

> (DV) whether an agent believes that p or not is voluntary.

On standard accounts S has voluntary control over φ-ing if and only if S can control whether or not she will φ by exertion of S's will.[2] So DV is the view that, in at least some cases, whether an agent believes p or not is a matter of voluntary control on the part of S. This view has as a consequence that in some contexts c, whether S believes that p is under the control of S's will.

However, doxastic voluntarism comes in two important forms: direct doxastic voluntarism and indirect doxastic voluntarism. Direct doxastic voluntarism is the view that we have direct voluntary control over our beliefs, and this means that whether an agent believes that p or not is voluntary and that whether an agent believes that p is under the control of that agent such that the agent does not need to anything other than choose to believe that p in order to believe that p.[3] Indirect doxastic voluntarism is the view that we have indirect control over our beliefs. This means that whether an agent believes that p is under the control of the agent such that there is some q, r, s . . . such that the agent has direct control over q, r, s, . . . and q, r, s, . . . bring about

the belief that p.[4] ED appears to require only that indirect doxastic voluntarism is true, but as we shall see, at least some epistemic rules appear to involve the assumption of direct doxastic voluntarism.

4. Belief-Contravening Commitments

It is beyond dispute that there are a wide variety of propositional commitments that we can have toward propositions. This includes believing, wishing, desiring, considering, and so on. However, there is a distinct tendency in philosophical discussions of human cognitive behavior involving representation—particularly in epistemology—to treat belief as the common default attitude involved in numerous types of such behaviors. So on this default view, most propositional commitments in epistemological contexts are taken to be beliefs. One view of belief, then, that is widely shared is that the norm of belief is truth and thereby EV-B is thought to be a crucial element of the ethics of belief in virtue of this fact. This is because evidence is the only reliable indicator of truth. The problem that then arises in the context of DV and EV-B is that there appear to be a variety of contexts where epistemic agents have epistemically meritorious propositional commitments that contravene evidentially grounded beliefs already held by that agent.[5] But this appears to be epistemically impermissible given EV-B.

Given ED, there will be a host of specific rules that define the duties that govern our epistemic behavior in various contexts. One such rule arises out of the AGM theory of belief dynamics that Carlos Alchourrón, Peter Gärdenfors, and David Makinson developed in the 1980s, and a number of related theories have arisen as a consequence.[6] These theories are fundamentally based on the concept of a belief state K satisfying the following minimal conditions:

> (BS) A set of sentences, K, is a belief state if and only if (i) K is consistent and (ii) K is objectively closed under logical implication.

The content of a belief state is then defined as the set of logical consequences of K. Given this basic form of epistemic representation, the AGM-type theories are intended to be a normative theory about how a given belief state that satisfies the definition of a belief state is related to other belief states satisfying that definition relative to (1) the addition of a new belief b to K_i or (2) the retraction of a belief b from K_i, where $b \in K_i$. Belief changes of the latter kind are termed *contractions*, but belief changes of the former kind must be further subdivided into those that require giving up some elements of K_i and those that do not. Additions of beliefs that do not require giving up previously held beliefs are termed *expansions*, and those that do are termed *revisions*. Specifically, for our purposes here it is the concept of a revision that is of crucial importance to the issue of providing an account of rational commitment for conditionals.

What is important to the topic of this essay is that on the basis of such theories of belief revision, the defenders of this approach to belief dynamics have also proposed a theory of rational conditional commitment.[7] This theory is a normative account of the conditions under which one ought to believe a conditional. The core concept of this theory is the Ramsey Test:[8]

> (RT) Believe a sentence of the form 'If p, then q' in the state of belief K if and only if the minimal change of K needed to believe p also requires believing q.[9]

Even in this quasi-formal form we can see what the AGM and other theorists have in mind. The Ramsey Test requires that we conservatively modify our beliefs by adding p into our standing system of beliefs and then see what the result is. What this theory then appears to require of us is that our actual

system of beliefs must be altered in order to believe a conditional. We must believe the antecedent, make the appropriate minimal modifications to our belief system, and then see if the consequent is believed in that new doxastic state. Consider, however, how this would work in the case of the large class of conditionals with false and known to be false antecedents. For example, consider the following conditional: if John McCain had won the 2008 presidential election, then the U.S. would have engaged in air strikes against Iran's nuclear facilities in 2011. According to ED we ought to follow the correct epistemic rules and let us suppose that this includes RT. According to RT we ought to believe this conditional if and only if upon believing the antecedent and making the appropriate minimal changes in one's doxastic state, one would also believe the consequent. This then involves a belief-contravening commitment to the proposition that John McCain won the 2008 presidential election. According to DV and the standard view that belief-contravening commitments are themselves beliefs, we are then in a situation where we can believe that John McCain won the 2008 presidential election in order to follow RT and that commitment is itself a belief. Suppose also that in fact that RT yields the result that the McCain conditional should be believed. All of this then gives rise to the following undesirable situation. By EV-B we ought not to believe that John McCain won the 2008 presidential election but according to RT we ought to believe it because according to RT we ought to believe the McCain conditional. We cannot believe the McCain conditional unless we believe that he won the 2008 election and so we ought to believe that. So we are faced with a practical contradiction. Moreover, we can again see that in such cases ED problematically tells us to follow the categorical epistemic rules, but we cannot possibly do so because there are two conflicting imperatives, and we have an outright contradiction entailed by the combination of these views.

5. Possible Solutions and the Concept of Acceptance.

There are several ways that one might deal with this situation without ceding ED, DV, and/or EV-B. So let us consider them in turn. First, one might attempt to avoid this outcome by ranking imperatives. This might then allow the defender of ED, DV, and EV-B to argue that RT should be violated and EV-B satisfied when attempting to ascertain whether we should believe conditionals. One would then of course have to propose a plausible alternative to RT for counterfactuals. The basis for this maneuver might then be that EV-B is a more fundamental epistemic obligation than RT, and so this imperative is ranked lower in the hierarchy of our epistemic obligations. As such, the contradictions that follow from the conjunction of ED, DV, and EV-B when conjoined to RT might be resolved by appeal to the relative rankings of our epistemic obligations in this way. However, there does not seem to be any obvious and non–ad hoc way to do this. There is no clear or natural principled ranking of epistemic obligations available to us and, more importantly, even if there were such a ranking that identified EV as more fundamental than RT, it would paradoxically entail that we ought never to follow those lower-ranked rules when it comes to belief-contravening commitments, thus rendering them seriously deficient as epistemic rules. In the case of RT, we would specifically have to conclude that it has no application to counterfactuals and that there is some other appropriate rule for those conditionals. This is of course unacceptable, as that rule would seemingly have to conflict with RT because it is intended to apply to all conditionals.

A second way to avoid the implications noted above might be to simply deny EV-B. If we cede EV-B, then there is no conflict in our

following RT. In accord with ED and DV, we could then voluntarily adopt the belief-contravening beliefs that are necessary to satisfy RT without violating the epistemic norm EV-B. However, this response is not one we would like to adopt in light of the widely held view that truth is the (unique) norm of belief and of justified belief. If we reject EV-B, then it will be epistemically permissible to believe propositions independent of evidence (which is presumably truth-tropic in nature), and it is hard to see how those who champion the ethics of belief in terms of ED could endorse this severe weakening of their view.

A third way to defuse the problem of belief-contravening epistemic rules would be to reject all epistemic rules that entail violations of EV-B. So one might simply argue that RT is not an acceptable epistemic rule and thus does not specify an epistemic obligation. As we saw in the case of response one, this view is ultimately implausible. The defenders of rules like RT that involve belief-contravening commitments have presented significant and often compelling defenses of those epistemic norms, and it seems to be a blatantly ad hoc response on the part of defenders of ED and DV to suggest that no such view can be correct because they are at odds with EV-B. One might suggest that there are good reasons to endorse EV-B, but that does not appear to be a particularly good reason to reject RT given the sorts of reasons that support it. The defender of ED and DV would have to have some serious reasons to question RT that are independent of its conflicting with EV-B in order to adopt this response, and it is unlikely that this would work in all cases of epistemic rules that involve belief-contravening commitments.

A fourth attempt to defuse the problem of belief-contravening epistemic rules would be to reject rules like RT on the basis of the specific claims that they assume direct voluntarism and that direct voluntarism is implausible.[10] Recall RT says that for an arbitrary counterfactual conditional, we must believe the antecedent of that conditional in order to see if the conditional is itself to be believed. This means then that whatever antecedent we might find as a component of any conditional, it must be up to the agent considering that conditional whether to believe the antecedent proposition or not if the agent is to be able to use RT effectively and comprehensively as a rule in the sense of ED. However, if this means that believing the antecedent is under the agent's control in the direct sense, then one might be tempted to believe that RT assumes an indefensible form of voluntarism about propositional commitments and so ought to be rejected as a proper epistemic rule. This response is, however, deeply problematic.

First off, we should note that RT seems to be an independently plausible rule for believing conditionals. However, if we accept that any rule that assumes a form of direct voluntarism is illegitimate, then we are faced with the following problem. Either we must reject the independently plausible epistemic rule RT on this basis or we must reinterpret RT so as to be acceptable. But rejecting RT on this basis, given its independent plausibility, seems to be wrongheaded. So, as a result, it would appear to be the case that RT cannot be understood in terms of direct voluntarism because that view is supposedly illegitimate as applied to belief. The natural response is to then suggest that perhaps RT ought to be interpreted in terms of indirect voluntarism. Interpreting RT in this way, however, entails that applying RT involves only indirect control over commitments to the antecedents of arbitrary conditionals. But this interpretation of RT is itself totally implausible in the case of counterfactuals. It would amount to the view that agents cannot entertain arbitrary counterfactual conditionals by simply choosing to commit to the ante-

cedent of a given conditional in the direct sense of control. So suppose that John is trying to determine whether or not he ought to believe the following conditional: if JFK had not been assassinated, then the Vietnam conflict would have been shorter in duration. Given this interpretation of RT, John does not have direct control over his commitment to the proposition that JFK was not assassinated. He only has indirect control over this commitment. On this interpretation of RT, if it is up to John whether to commit to that proposition, John must have direct control over some other states that bring about his commitment to the proposition that JFK was not assassinated, and in order to apply RT, he must bring about those states in order to yield a verdict about whether or not to believe the conditional. Note that given EV-B, the states that John has direct control over cannot be the relevant evidence or grounds for belief in the antecedent proposition, for he knows it is false and most certainly does not need to look for *misleading* evidence in order to entertain the hypothetical claim.[11] The assumption that John would have to find misleading evidence in order to engage in committing to the antecedent of the counterfactual in question so that he could evaluate the conditional via RT is utterly implausible and psychologically pathological. The latter charge derives from the recognition that this view entails that John would have to engage in knowing and willful self-deception to apply RT so interpreted. More straightforwardly, it simply does not seem to be the case that John needs to do anything at all other than (directly) commit to the proposition that JFK was not assassinated in order to assess the conditional. He most certainly does not need to seek the advice of a therapist, read conspiracy books, or undergo brainwashing, and so on in order to adopt that commitment and thus assess whether or not to believe the conditional. So, contrary to the basis of this objection, we do seem to be able to entertain the belief-contravening antecedents of counterfactual conditionals at will in the direct sense, and RT is a reasonable epistemic rule for how to do this. Thus, objecting to rules like RT on the basis that direct doxastic voluntarism is implausible in this case. Some form of direct voluntarism does not seem reasonable in such cases, and so RT is in that respect a reasonable deontological epistemic rule in addition to its plausibility in other respects. But as we shall see momentarily, what this objection really suggests is the third possibility, that RT does not involve belief in the antecedents of conditionals at all and the above discussion has been carefully constructed so as to suggest this in its strategic uses of "commitment" instead of "belief" and "direct/indirect voluntarism" instead of "direct/indirect doxastic voluntarism."

So for the reasons just given, none of three responses suggested so far appears to be especially workable as a solution to the problem presented here. There is, however, a fourth response open to those that defend ED, DV, and EV-B. This response involves denying the standard account of belief-contravening commitments, entertaining a restriction on EV as a result and a small modification of RT. RT mentions belief but it is not clear that this is really an essential element of the rule, and this suggestion arises out of the fourth objection dealt with immediately above. One might adopt the view that rules like RT do not involve belief-contravening *beliefs* at all. In effect, one can deny the standard account of belief-contravening commitments. Thus, one can defuse the problematic implications of ED, DV, and EV-B without (i) introducing an ad hoc and problematic ranking of epistemic obligations, (ii) rejecting all epistemic rules that involve belief-contravening commitments, or (iii) being forced to accept direct doxastic voluntarism. This can be realized by interpreting belief-contravention as involving some commitment other than belief that is not veristic. Veristic propositional attitudes, then, are those that have truth (uniquely) as

norms.[12] Belief and knowledge are examples of veristic commitments. The aim of veristic propositional commitments is truth, and this is why EV applies to them. So it is crucial to this response that we must restrict EV in the following way:

> (EV') *Veristic* doxastic attitude *D* toward proposition *p* is epistemically justified for *S* at *t* if and only if having *D* toward *p* fits the evidence *S* has at *t*.

Importantly, nonveristic propositional commitments include forms of acceptance. There are several extant accounts of acceptance, but the best-known account was introduced by Cohen (1992). On this view, accepting p is a propositional commitment characterized by the following important features (among others):

> A1. Accepting p is voluntary.
> A2. Accepting p is nonevidential.
> A3. Accepting p is not a commitment to the literal truth of p.
> A4. Accepting p is subjectively closed under implication.

Accepting p is then something like adopting p for pragmatic reasons or "trying out" p in a cognitive sense. What is most important here is that since acceptance is governed by A2 and A3, it is clearly not a veristic doxastic attitude. So EV' does not apply to the accepting of propositions. This, then, suggests that we can refine RT as follows in a way that makes this rule perfectly compatible with EV':

> (RT') Believe a sentence of the form 'If *p*, then *q*' in the state of belief *K* if and only if the minimal change of *K* needed to *accept p* also requires *accepting q*.

Nothing in principle is lost in recasting this epistemic principle in this way. So understood, it is an epistemic obligation in accordance with ED, it does not involve violating DV, and it does not conflict with EV-B. So one can maintain that we have epistemic obligations involving belief-contravening commitments while accepting ED, DV, and EV-B, provided one is willing to reject the standard account of belief-contravening commitments and reinterpret epistemic rules that involve them in terms of nonveristic commitments like acceptance.

NOTES

1. See Alston 1988 and Steup 1986.

2. See Steup 2000, p. 28, in support of this analysis.

3. See Steup 1986, 71.

4. See ibid.

5. Here the example developed will concern the Ramsey Test for conditional acceptance. Other reasonable examples of such rules include the plausible scientific rule that we should adopt theories that, while known to be false, are approximately true, and the rule that we should adopt theories that, while known to be false, depend on acceptable idealizations.

6. See Alchourrón, Gärdenfors, and Makinson 1985, Gärdenfors 1988, and Levi 1996.

7. See Gärdenfors 1981 and 1988.

8. See Ramsey 1929 (1990).

9. One might think that this presentation of RT is uncharitable. This is because RT is ambiguous in the following sense. What RT requires of us is either (1) that our actual system of beliefs must be altered

by believing the antecedent of a conditional in order to believe a conditional, (2) that we *hypothetically modify* our beliefs in order to accept a conditional, or (3) that we add the *hypothetical belief* that *p* to our belief system in applying RT. But neither interpretations (2) nor (3) can possibly be acceptable. Interpretation (2) of RT entails a vicious regress. If we take interpretation (2) of the Ramsey Test to mean that in considering whether to accept $A > B$, we should *hypothetically add A* to our standing system of beliefs *K*, make the appropriate revisions in terms of the AGM postulates (or other similar postulates), and then see if *B* is in the resulting system of beliefs, then in order to accept $A > B$, we must accept the following additional conditional (D): if I were to add *A* to my standing belief system *K*, then I would believe *K*' . However, in order to see if we should believe D, we must apply RT to D. In the case of D, we must follow RT and thus we get a vicious regress. So (2) is not a defensible interpretation of RT. If we instead adopt interpretation (3) of RT and read it as meaning that in order to see if we should accept $A > B$ we must add the *hypothetical belief* A, then we are owed an account of what hypothetical beliefs are; how they interact with ordinary beliefs; and how we can assess conditionals using them without introducing the sort of vicious infinite regress noted here. This is, however, a difficult task to satisfy in terms of the AGM theory in particular, as that is a theory of belief revision the postulates of which all concern beliefs and not hypothetical beliefs. But no such account has been offered. So interpretation (1) of RT seems to be correct, and this is further supported by what Ramsey actually himself says in Ramsey 1990. In the relevant footnote Ramsey says that, "If two people are arguing 'If *p*, then *q*?' and are both in doubt as to *p*, they are adding *p* hypothetically to their stock of knowledge and arguing on that basis about *q* (pp. 154–155)." Similarly, Stalnaker (1968) interprets RT as, "First, add the antecedent (hypothetically) to your stock of beliefs; second, make whatever adjustments are required to maintain consistency (without modifying the hypothetical belief in the antecedent), finally, consider whether or not the consequent is then true" (p. 106). These claims are themselves ambiguous, but for the reasons cited above, neither interpretation (2) nor interpretation (3) of these claims is acceptable, and so the objection that my interpretation of RT is uncharitable is rendered moot. As things stand, RT must be interpreted in terms of (1). Otherwise, it is unsatisfiable (in terms of (2)) or it crucially depends on an undefined concept that the AGM postulates say nothing about (in terms of (3)). See Shaffer 2011 and Shaffer 2012 for more on this contention.

10. See Steup 1986, pp. 70–74.

11. See ibid., pp. 71–72.

12. We should be careful to note that there are two possible interpretations of veritism with respect to propositional attitudes. Pure veristic propositional attitudes are those that have truth as a *unique* norm. Impure veristic propositional attitudes are those that have truth as a norm, but may also have other norms. The view assumed here is that belief and knowledge are pure veristic propositional attitudes and this is the normal view of the matter for those who adopt representational views of belief and who defend EV. Truth is the norm of belief and evidence is evidence for truth. For example, Conee and Feldman's (2004) version of EV appears to be purely veristic in nature in this sense. Pure veritism about belief might, however, be denied, but it is hard to see how that would avoid the problem posed here. Impure veritism about belief might permit pragmatic (or other nonevidential) reasons for belief and so might permit the possibility that rules like RT might be satisfiable without violating EV-B. But these sorts of reasons would have to be nontruth-tropic. In support of this point, Zemach (1997) has shown that pragmatic reasons cannot be evidential in the sense of providing reasons for truth. So the problematic epistemic rules that involve belief-contravening commitments would all have to involve purely nonveristic factors, but this is not what the defenders of RT, for example, have in mind. RT is designed to show us when we should believe that a conditional is such that it should be believed to be true. So for the purposes of this essay, belief and knowledge will be assumed to be purely veristic.

REFERENCES

Alchourrón, C., P. Gärdenfors, and D. Makinson. 1985. "On the Logic of Theory Change: Partial Meet Functions for Contraction and Revision," *Journal of Symbolic Logic*, vol. 50, pp. 510–530.

Alston, W. 1988. "The Deontological Conception of Epistemic Justification," *Philosophical Perspectives*, vol. 2, pp. 257–299.

Cohen, L. J. 1992. *An Essay on Belief and Acceptance* (Oxford: Clarendon Press).

Conee, E., and F. Feldman (2004). *Evidentialism*. (Oxford: Oxford University Press).

Gärdenfors, P. 1981. "An Epistemic Approach to Conditionals," *American Philosophical Quarterly*, vol. 18, pp. 203–211.

_____. 1988. *Knowledge in Flux* (Cambridge, Mass.: MIT Press).

Levi, I. 1996. *For the Sake of the Argument: Ramsey Test Conditionals, Inductive Inference, and Nonmonotonic Reasoning* (Cambridge: Cambridge University Press).

Ramsey, F. P. 1929 (1990). "Laws and Causality," repr. in *F. P. Ramsey: Philosophical Papers*, ed. D. H. Mellor (Cambridge: Cambridge University Press).

Shaffer, M. 2011. "Three Problematic Theories of Conditional Acceptance," *Logos & Episteme*, vol. 1, pp. 117–125.

_____. 2012. *Counterfactuals and Scientific Realism* (New York: Palgrave MacMillan).

Stalnaker, R. 1968. "A Theory of Conditionals," *Studies in Logical Theory, American Philosophical Quarterly Monograph*, vol. 2, pp. 98–112.

Steup, M. 1986. "The Deontic Conception of Epistemic Justification," *Philosophical Studies*, vol. 53, pp. 65–84.

_____. 2000. "Doxastic Voluntarism and Epistemic Deontology," *Acta Analytica*, vol. 15, pp. 25–56.

Zemach, E. 1997. "Pragmatic Reasons for Belief?" *Nous*, vol. 4, pp. 525–527.