# Misleading Higher-Order Evidence, Conflicting Ideals, and Defeasible Logic

ALEKS KNOKS
*University of Zurich*

Thinking about misleading higher-order evidence naturally leads to a puzzle about epistemic rationality: If one's total evidence can be radically misleading regarding itself, then two widely-accepted requirements of rationality come into conflict, suggesting that there are rational dilemmas. This paper focuses on an often misunderstood and underexplored response to this (and similar) puzzles, the so-called *conflicting-ideals view*. Drawing on work from defeasible logic, I propose understanding this view as a move away from the default metaepistemological position according to which rationality requirements are strict and governed by a strong, but never explicitly stated logic, toward the more unconventional view, according to which requirements are defeasible and governed by a comparatively weak logic. When understood this way, the response is not committed to dilemmas.

**Keywords**: epistemic normativity, higher-order evidence, rationality ideals, epistemic dilemmas, defeasible logic

## 1. A Puzzle

Thinking about misleading higher-order evidence—roughly, evidence about our evidence and our capacities for evaluating it—naturally leads to a puzzle about epistemic rationality.[1] The puzzle involves a clash between two widely-accepted requirements of rationality. According to the first—the *evidential requirement*—epistemic rationality requires that you have the doxastic attitudes that are

---

1. This puzzle has been noticed independently by several authors, including Lasonen-Aarnio (2020), Littlejohn (2018), and Worsnip (2018).

---

**Contact:** Aleks Knoks <aleks.knoks@gmail.com>

supported by your (total) evidence. According to the second—the *inter-level coherence requirement*—rationality requires that your doxastic attitudes are in line with your beliefs about whether or not these attitudes are supported by your evidence. We will state the two requirements more precisely in the next section, but for now simply note that they are distinct, intuitively plausible, and have been supported by strong arguments.[2] The trouble is that they come into conflict if we think—as I think we should—that it's possible for evidence to be radically misleading regarding what it itself supports. That is:

**Misleading total evidence (MTE):**

It's possible that (1) your total evidence supports some doxastic attitude; and that (2) your total evidence supports believing that your total evidence doesn't support this doxastic attitude.[3]

The best, if still contested, illustrations of this sort of evidence from the literature are cases in which agents receive superb, yet misleading higher-order evidence. Here's one such:

*Prof. Moriarty's Drug*. Sherlock Holmes is a master sleuth famously good at assessing evidence. After investigating a murder that took place at the manor on the hill he concludes, correctly, that the maid is the one who did it. Not long afterwards Holmes finds out, from a very reliable source, that his archnemesis Professor Moriarty had slipped a drug into his morning tea. Holmes knows the drug's effects too well: In all but 5% of detectives, the ability to assess evidence gets distorted, and in a way that is not noticeable to them. What Holmes doesn't know is that he is among the lucky 5% that are immune to the drug.[4]

And for a more realistic scenario, consider:

*Implicit Bias*. Joanna is white. She's been exposed to evidence that very strongly suggests that most white people are subject to implicit bias which causes them to believe, incorrectly, that members of another race are ruder than they really are. Joanna meets a member of another race

---

2. For a defense of the evidential requirement see Conee and Feldman (2004) and Lasonen-Aarnio (2020), among others. For extended defenses of the inter-level coherence requirement see Feldman (2005), Horowitz (2014), Lasonen-Aarnio (2020), Sliwa and Horowitz (2015), and Worsnip (2018), among others.

3. The label "Misleading total evidence" comes from Worsnip (2019).

4. This vignette amalgamates cases from Christensen (2007) and Coates (2012).

and judges them rude on the basis of several personal interactions. It turns out, however, that Joanna is unusual in that she doesn't suffer from this particular implicit bias. But she doesn't know it.[5]

Advocates of (MTE) are pulled to the following analysis of these cases. Holmes's first-order evidence—roughly, the clues lying around the house—is sufficient evidence to believe that the maid did it, while his higher-order evidence—roughly, the information speaking to his capacity for evaluating clues—is sufficient evidence to believe that his (total) evidence does not support believing that the maid did it. Similarly, Joanna's first-order evidence—roughly, her first-hand experience with the particular member of another race—is sufficient evidence to believe that she's dealing with a rude person, while her higher-order evidence—roughly, the information about the bias—is sufficient evidence to believe that her (total) evidence does not support believing that she's dealing with a rude person. So there would seem to be cases instantiating both (MTE1) and (MTE2).[6]

But once this much is granted, we quickly run into trouble—I'll use Moriarty's Drug as the running example in what follows, but nothing hinges on the particular case we use: Holmes's evidence supports believing that the maid did it, as well as believing that his evidence doesn't support believing that the maid did it. In light of this, the evidential requirement appears to demand that Holmes believes that the maid did it, as well as that his evidence doesn't support believing that she did. The coherence requirement, on the other hand, appears to demand that Holmes doesn't believe that the maid did it if he also believes that his evidence doesn't support believing that she did. A minute's reflection, then, reveals that were Holmes to comply with the first requirement, he would thereby violate the second, and the other way around. So, supposing the two requirements are indeed genuine, Holmes simply can't do what epistemic rationality requires him to do, implying that Moriarty's Drug describes a rational dilemma.

---

5. Many thanks to the anonymous area editor of *Ergo* for this vignette and the suggestion to include it here.

6. For the most forceful arguments in favor of (MTE) see Lasonen-Aarnio (2020: §3) and Worsnip (2018; 2019); see also Coates (2012), Littlejohn (2018), and Weatherson (2010). It's been suggested to me that (MTE) is more controversial than my setup of the puzzle suggests, as multiple authors have argued against it. But while (MTE) may be less intuitive of a starting point than the two requirements, I've yet to see an argument against it that wouldn't be motivated by a prior commitment to one of these requirements. As a result, I think of those arguing against (MTE) as responding to the puzzle by denying it—see Footnote 20 for a list of references. Also, it's worth highlighting that the solution I'm going to present should be of interest independently of the particular puzzle—and thus the truth of (MTE)—as it naturally generalizes to *any* puzzle involving a conflict between epistemic requirements.

What should we say in response to this? That one of the requirements isn't genuine? That total evidence can never radically mislead about itself? That there are two independent domains of epistemic rationality, a domain concerned with having attitudes that track one's evidence and a domain concerned with having the right sort of fit between one's attitudes? Or should we, perhaps, bite the bullet and concede the existence of dilemmas? Well, although these responses have been explored in the literature, all of them incur intuitive costs and no single one has emerged as a clear winner.[7] And so we are still left with a genuine puzzle about rationality, a puzzle that has drawn and continues to draw much attention, in epistemology and outside it.

My main goal in this paper is to present a solution to this puzzle which—unlike any of the others—lets us preserve the two requirements, the possibility of radically misleading (total) evidence, as well as the unity of epistemic rationality, all while denying that there are dilemmas. There's a sense in which this solution is implicit in the *conflicting-ideals view* that David Christensen (2007; 2010; 2013) has been advocating for in his recent work. (According to this view, there are inherently unfortunate epistemic situations in which the agent can't act in accordance with all the "epistemic ideals" that apply to her, and not due to her cognitive limitations, but, rather, the way these ideals relate to each other.[8]) But the route that will get us to the solution is both new and nonstandard. What's more, taking it will help us understand the conflicting-ideals view itself much better.[9] The core idea is this: There's an implicit assumption in the literature about what we might call the *logic of interaction between requirements*, and the disconcerting result—that is, commitment to dilemmas—follows only if this assumption is in place. The conflicting-ideals view, then, can be naturally thought of as rejecting this assumption and holding that this logic is weaker than standardly thought, and that rationality requirements are, in fact, defeasible.

Here's the plan of attack. Sections 2 and 3 will introduce a simple formal notation and restate the puzzle in it. This exercise will help us unearth the hidden assumption, as well as put us in a position to draw on the work from the defeasible logic paradigm to devise a model which I see as the formal backbone of the conflicting-ideals view—this will be the task of Section 4. Section 5 will apply this model to the puzzle. Then, anticipating a potential worry, Section 6

---

7. I should say upfront that I will not engage with these responses in what follows, focusing on developing my preferred solution instead. I will, however, provide a fuller classification of responses with pointers to the literature in Section 3.

8. Various tragic epistemic scenarios of this sort are discussed in, e.g., Christensen (2007; 2013; 2016), Leonard (2020), Pryor (2018), and Schechter (2013).

9. It's noteworthy that the literature focusing on the puzzle usually takes Christensen to hold a view that concedes the existence of rational dilemmas. See, e.g., Lasonen-Aarnio (2020), Silva (2017), and Worsnip (2018) all of whom suggest that he embraces dilemmas. All three authors propose it as *one* interpretation of Christensen's views, but no other interpretation is ever discussed.

*Misleading Higher-Order Evidence, Conflicting Ideals, and Defeasible Logic* • 145

will add a further twist to the model. Section 7 will discuss some objection to the solution along with responses to them, and Section 8 will conclude.

## 2. Formal Preliminaries

Before we turn to the notation, three notes are in order. First, following Alex Worsnip (2018; 2019), we'll be thinking of *evidential support* as a two-place relation that obtains between bodies of evidence and doxastic attitudes. It's worth highlighting that this way of thinking comes with a certain sort of agent-neutrality: Once you fix the body of evidence, the doxastic attitudes it supports are also fixed. This doesn't mean that we're doing away with the agent—any body of evidence will be some agent's body of evidence—but it does mean that we can suppress the agent in the notation. Second, we'll try to remain as non-committal as possible about the nature of evidence and evidential support relation. What I'm going to say should be compatible with various ways of thinking about evidence—as a set of propositions, facts, or mental states—as well as various ways of spelling out the support relation. And third, we'll restrict attention to all-or-nothing attitudes—this is not to take a stance in the debate about the relation between credence and full belief, but to keep things manageable.[10]

As background, we assume the language of ordinary propositional logic with all the standard connectives and the customary turnstile $\vdash$ standing for classical logical consequence: Where $X$ and $Y$ are propositional formulas, $X \vdash Y$ means that $Y$ is a classical consequence of $X$. We supplement the language with the constant $E$ that will stand for an agent's total body of evidence. (Sometimes $E$ will stand for a particular body of evidence, and at other times for an arbitrary one—the context will always let disambiguate.) Further, we introduce three new operators: The one-place $B(\cdot)$ will be used to capture all-or-nothing doxastic attitudes. Thus, $B(p)$ says that the agent believes $p$, and $B(\neg q)$ says that the agent disbelieves $q$.[11] The two-place operator $\models$, in turn, will be used to formulate claims about the support relation. Thus, $E \models B(p)$ says that the body of evidence $E$ supports believing $p$, and $E \models B(\neg q)$ says that $E$ supports disbelieving $q$. Note that $\models$ is meant to stand for all things considered support. When an attitude $B(X)$ is not supported by the evidence, we will write

---

10. As is known too well, issues surrounding misleading higher-order evidence do not disappear once one moves to the credence-based framework—see, e.g., Christensen (2010). So sticking with the more parsimonious all-or-nothing attitude-based framework seems like a methodologically sound move.

11. We can also easily express suspension of judgment, either as $\neg B(p) \& \neg B(\neg p)$, or by introducing a new predicate $S(\cdot)$. But this particular attitude won't play a prominent role in the discussion.

$E \not\Rightarrow B(X)$. We will also form more complex expressions, including formulas that capture second-order beliefs—an agent's beliefs about whether a certain attitude is supported by their evidence—as well as formulas that assert the existence of support between $E$ and such beliefs. For example, $B(E \Rightarrow B(p))$ says that one believes that one's evidence supports believing $p$; and $E \not\Rightarrow B(E \Rightarrow B(p))$ says that the evidence doesn't support having this second-order belief.

Finally, we'll make use of the customary deontic operator $\bigcirc(\cdot)$. Together, the requirements of epistemic rationality or epistemic norms—I use the two terms interchangeably—set up a certain epistemic standard. A formula of the form $\bigcirc X$, then, should be read as saying that, according to the standards of epistemic rationality, it ought to be the case that $X$. In what follows, we will often refer to this standard as the *epistemic ought*.

With this notation in hand, the puzzle can be brought into plain sight. It's comprised of three claims. The first captures the evidential requirement:

**Evidential requirement (ER):**

$$[E \Rightarrow B(X)] \supset \bigcirc B(X)$$

In English: If your evidence supports believing $X$, then you ought to believe $X$. This thesis is very intuitive, with many epistemologists considering it a platitude.[12] Also, it—or something close to it—lies at the heart of the popular philosophical view called *evidentialism*.[13] The second claim captures the coherence requirement:

**Inter-level coherence requirement (ILC):**

(1) $\bigcirc[B(E \Rightarrow B(X)) \supset B(X)]$

(2) $\bigcirc[B(E \not\Rightarrow B(X)) \supset \neg B(X)]$

In English: (1) Rationality requires that you believe a proposition if you believe that your evidence supports that belief. (2) Rationality requires that you do not believe a proposition if you believe that your evidence doesn't support that belief. Much of the appeal of (ILC) derives from the intuition that there's something seriously wrong with agents who believe, act, or assert in accordance with mismatched attitudes of the sort (ILC) prohibits. Note the difference in the formal statement of (ER) and (ILC). In the former the ought occurs in the consequent,

12. Cf. Silva (2017) and Worsnip (2018).

13. In fact, we can define evidentialism via (ER): An agent is epistemically rational if and only if their attitudes satisfy (ER). For a concise introduction to evidentialism see Conee and Feldman (2008). For book-length defenses see Conee and Feldman (2004) and McCain (2014).

while in the latter it ranges over the entire conditional. This reflects the distinction between narrow- and wide-scope rationality requirements that goes back at least to Broome (1999).[14] Assigning (ER) and (ILC), respectively, narrow and wide scope is standard in the literature.[15]

The third and final claim is the possibility of total evidence that's radically misleading regarding itself. We have noted already that many epistemologists would characterize Holmes's evidence in Moriarty's Drug as evidence of just this sort. That is, they would say that his total evidence supports both believing that the maid did it, and believing that his evidence doesn't support believing that she did. Letting $m$ stand for the proposition that the maid committed the murder, we acquire:

**Misleading total evidence (MTE):**

(1) $E \Mapsto B(m)$

(2) $E \Mapsto B(E \nMapsto B(m))$

Note that Moriarty's Drug is only an example, and that we have a puzzle in case it's metaphysically possible that there's *some* body of evidence $E$ and *some* proposition $X$ with both $E \Mapsto B(X)$ and $E \Mapsto B(E \nMapsto B(X))$.[16]

## 3. Deriving the Disconcerting Result

Now all the pieces are in place, and we can construct a proof showing how exactly (ER), (ILC), and (MTE) lead to the conclusion that Moriarty's Drug describes a dilemma. We start with facts describing evidential support:

---

14. Narrow-scope requirements demand that a particular attitude is adopted once the conditional's antecedent is satisfied. Wide-scope requirements are supposed to give the agent more freedom. Consider Alice who finds herself believing that her evidence supports a belief in $p$. There are two ways for her to comply with (ILC1)—and we're supposing that complying with it is all that matters. Alice can either adopt a belief toward $p$, or she can drop her initial belief. For more on narrow vs. wide scope see, e.g., Broome (2007; 2013). For a recent criticism of the distinction see Fogal (2019).

15. See, e.g., Lasonen-Aarnio (2020), Silva (2017), Worsnip (2018). Those familiar with the debate about the normativity of rationality might have the following thought here: On some views in the debate, such structural requirements as (ILC) are either nonnormative—see, e.g., Kolodny (2005)—or simply nonexistent—see Kiesewetter (2017) and Lord (2018). So the proponents of such views wouldn't accept (ILC) and, thereby, also avoid the puzzle. But while they would avoid the particular puzzle we focus on here, they have to deal with a closely related one. Here's why. Denying (ILC) standardly goes in hand with accepting disjunctions of (nonstructural) requirements $\bigcirc\neg B\,(E \Mapsto B(X)) \lor \bigcirc B(X)$ and $\bigcirc\neg B(E \nMapsto B(X)) \lor \bigcirc\neg B(X)$, and, when these are combined with the other components of the puzzle, very similar problems arise. For a discussion of the kindred puzzle see Kiesewetter (2017: Ch. 9.6.2).

16. Cf. Lasonen-Aarnio (2020) and Worsnip (2018).

(1) $E \models\Rightarrow B(m)$            (MTE1)

(2) $E \models\Rightarrow B(E \not\Rightarrow B(m))$        (MTE2)

These determine the relevant instances of (ER):

(3) $[E \models\Rightarrow B(m)] \supset \bigcirc B(m)$       (ER)

(4) $[E \models\Rightarrow B(E \not\Rightarrow B(m))] \supset \bigcirc B(E \not\Rightarrow B(m))$       (ER)

Two applications of modus ponens result in (5) and (6)—which say, respectively, that Holmes ought to believe that the maid did it, and that Holmes ought to believe that the evidence doesn't support believing that the maid did it:

(5) $\bigcirc B(m)$            from (1) and (3)

(6) $\bigcirc B(E \not\Rightarrow B(m))$        from (2) and (4)

Next comes the following instance of (ILC):

(7) $\bigcirc[B(E \not\Rightarrow B(m)) \supset \neg B(m)]$       (ILC2)

This formula says that Holmes ought not to believe that the maid did it in case he believes that the evidence doesn't support believing that the maid did it. The subsequent step (8) is the crucial one for our purposes. To get further, we need to appeal to a logical feature of the deontic operator. More specifically, we need to assume that it satisfies the principle $\bigcirc(X \supset Y) \supset (\bigcirc X \supset \bigcirc Y)$.[17] Instantiating $X$ with $B(E \not\Rightarrow B(m))$ and $Y$ with $\neg B(m)$, we get (8):

(8) $\bigcirc[B(E \not\Rightarrow B(m)) \supset \neg B(m)] \supset [\bigcirc B(E \not\Rightarrow B(m)) \supset \bigcirc \neg B(m)]$    (Principle)

Then we have two more applications of modus ponens:

(9) $\bigcirc B(E \not\Rightarrow B(m)) \supset \bigcirc \neg B(m)$       from (7) and (8)

(10) $\bigcirc \neg B(m)$           from (6) and (9)

---

17. In the logic literature this is known as the *axiom K,* and it's a very natural principle for a deontic operator to satisfy. But see Broome (2013: Ch. 7) for an argument to the conclusion that a logic of requirements (and oughts) shouldn't be closed under *K*.

And, as a final touch, we combine the formulas acquired in steps (5) and (10):

(11) $\bigcirc B(m) \& \bigcirc \neg B(m)$                             conjunction, from (5) and (10)

This last formula is disconcerting because it entails that Holmes finds himself in a normative dilemma, or a situation in which it ought to be the case that $X$ and it ought to be the case that $Y$, while it is impossible for both $X$ and $Y$ to obtain.[18] If Holmes complies with either ought on line (11), he thereby violates the other. So there's no way he can have the attitudes he ought to have. And given that epistemic rationality is the source of norms in play, the derivation suggests that there are rational dilemmas.

Most epistemologists have tried to resolve the puzzle in ways that keep clear of dilemmas.[19] In particular, they have suggested rejecting (MTE),[20] rejecting (ER),[21] rejecting (ILC),[22] or treating the conflict as a clash between two irreducible types of epistemic rationality.[23] The derivation presented here provides a way to classify these responses: The first move denies one of the premises, the second denies either step (3) or (4), the third denies step (7), and the fourth rejects step (10).[24]

The real significance of the derivation, however, consists in its explicitly showing that the normative component picked out by $\bigcirc$ makes its own contribution to the puzzle. In order to get to the disconcerting result, we had to rely on a logical feature of $\bigcirc$. On the intuitive level, this can be understood thus: There's a nontrivial logic governing the interaction between epistemic requirements, and the puzzle causes real trouble only if it is further assumed—as most of the literatures has done hitherto—that this logic is relatively strong.

But this assumption can and should be questioned, as one attractive and underexplored route of response to the puzzle is to weaken the logic. And I'm suggesting that we understand the conflicting-ideals view as taking just this

---

18. This is the standard definition of normative dilemmas—see, e.g., Goble (2009: 450) and Horty (2003: 557).

19. The list of authors most likely to embrace the derivation's conclusion would include Hughes (2019) and Priest (2002).

20. See Feldman (2005), Horowitz (2014), Kiesewetter (2017: Ch. 9), Skipper (2019), Tal (2020), Titelbaum (2015), White (2007), among many others.

21. See Littlejohn (2018).

22. Authors who are often taken to reject (ILC) include Coates (2012), Lasonen-Aarnio (2020), and Weatherson (2010). But Lasonen-Aarnio appears to be the only one to explicitly advocate the rejection of (ILC) in response to the puzzle.

23. See Fogal (2019), Silva (2017: especially, en. 11), and Worsnip (2018).

24. How exactly does it reject step (10)? Well, two types of rationality means two distinct normative domains, and, hence, also two distinct deontic operators. Worsnip, for instance, distinguishes between the domains of evidence and coherence, and so implicitly also between the the oughts of evidence and coherence. The ought occurring on line (6) is an ought of evidence, while the one occurring on line (9) is an ought of coherence. Since the oughts are distinct, we can't apply modus ponens.

route. Let me emphasize that I am *not* suggesting that it simply denies that the epistemic ought satisfies the deontic principle we used in the proof. Instead, my suggestion is that we understand the view as not only changing the logic governing the interaction between epistemic requirements, but also distinguishing it from the logic of the epistemic ought. The result is a view of epistemic normativity, according to which epistemic oughts are determined through the interaction of *defeasible* epistemic requirements, or *ideals*. This response to the puzzle is hardly ad hoc. For, first, it naturally generalizes to other puzzles involving conflicts between requirements. And second, it parallels a familiar and well-respected move in the ethical literature on moral conflicts.

What we are going to do next is define a concrete defeasible deontic logic as a substitute for the implicitly assumed logic of $\bigcirc$ and, then, see how it can help us avoid the disconcerting result. As flagged above, I see this logic as the formal backbone of the conflicting-ideals view.

## 4. A Logic for Conflicting Imperatives

The logic we'll use isn't really new. Its core is a well-known and studied defeasible consequence relation defined in terms of classical consequence and maximally consistent subsets—I'll explain the latter notion in due time. Nicholas Rescher and Ruth Manor (1970) appear to have been the first to define and study consequence relations of this sort, and John Horty (1994; 2003) was the first to apply them in deontic setting, in the context of the debate over the existence of moral dilemmas. They also have close connections to Bas van Fraassen's (1973) deontic logic, Raymond Reiter's (1980) default logic, and other logics from the defeasible logic paradigm.[25] The particular notation we'll use comes from Horty (2003), and the consequence relation we'll define is a close cousin of what's called the *disjunctive account* there. For expository purposes, we'll discuss its application not only in the epistemic, but also in the moral domain.

The first thing we need to do is distinguish between a weaker and a stronger sense of *ought*. This move is standard in the literature on moral conflicts, where it is also customary to use technical terminology to keep the weaker and the stronger sense of the *moral ought* separate: Thus, Roderick Chisholm (1964) writes about *prima facie* and *absolute duties*, van Fraassen (1973) about *imperatives* and *oughts*, John Searle (1980) about *obligations* and *oughts*, Philippa Foot (1983) about *type 1* and *type 2 oughts*, Christopher Gowans (1987) about what we *ought* and what we *must* do, and Horty (2003) simply about *weak* and *strong oughts*. The terms used are different, but the underlying idea is always the same: Take

---

25. See Horty (1994) and Makinson (2005).

some situation that requires a nontrivial response on the part of the agent. The first term of the pair would be used to pick out various moral considerations the agent has and the responses they support; and the second term would be used to describe what the agent's response ought to be once all the relevant considerations are taken into account and weighed against each other. For illustration consider the well-known example from Sartre (1946/1996) in which a French youth at the time of Second World War is torn between the patriotic duty to fight for his country and the filial duty to care for his distressed and aging mother. All the parties to the debate about moral conflicts would agree—at least, once they left the terminological differences behind—that there's a weaker sense of *ought*, according to which the youth ought to fight for his country and also ought to stay with his mother. What they wouldn't agree on is whether the youth ought to do both of these things in the stronger sense of *ought* too—that is, whether he faces a genuine moral dilemma.

While drawing an analogous distinction between a weaker and a stronger sense of the *epistemic ought* is less common, nothing stands in the way of doing it. What's more, if I am right, the conflict-ideals diagnosis of the various inherently unfortunate or tragic epistemic scenarios discussed in the recent literature entails such a distinction.[26]

Given that we'll be concerned with both the epistemic and the moral domain, we need neutral terminology that fits both. Taking a cue from van Fraassen, we will refer to the weaker oughts—whether they be epistemic or moral—as *imperatives* and to the stronger ones as *all things considered oughts*, or *oughts* without qualification. At times, however, we will also refer to the weaker epistemic oughts as *ideals*. Now, it is both natural and common—at least, in the ethical literature—to take the all things considered oughts to be determined by the imperatives. And that's just how our logic will work: It will generate all things considered oughts from imperatives.

We will reserve $\bigcirc$ for the oughts and use a new two-place operator $!(\cdot\,|\,\cdot)$ to state conditional imperatives. Where $X$ and $Y$ are arbitrary propositions, $!(Y\,|\,X)$ says that there's a demand that $Y$ obtains under circumstances $X$, or, alternatively, that, ideally, $Y$ should obtain under circumstances $X$. Importantly, the demand is defeasible, as its force can get overridden. We'll denote imperatives using the letter $i$ (with subscripts). It will also be useful to introduce two functions for picking out their antecedent and consequent parts: Where $i$ stands for the imperative $!(Y\,|\,X)$, the output of *Antecedent*$[i]$ is the proposition $X$ and that of *Consequent*$[i]$ is the proposition $Y$. Note that

---

26. Let me add a note of caution: Since the moral ought in its weaker sense is roughly synonymous with the widely used notion of a moral reason, you might naturally expect that the epistemic ought in its weaker sense must be roughly synonymous with the notion of an epistemic reason. However, in the present context the latter two are definitely not synonymous.

an imperative's antecedent specifies its *triggering conditions*, or circumstances under which the imperative comes into force. An imperative's consequent, in turn, specifies its *satisfaction conditions*, or circumstances under which the imperative's demand gets fulfilled. We'll also lift the second function from individual imperatives to sets of imperatives: Where $\mathcal{I}$ is a set of imperatives, $Consequent[\mathcal{I}]$ is the collection of the consequents of all the imperatives from $\mathcal{I}$, that is, $Consequent[\mathcal{I}] = \{Consequent[i] : i \in \mathcal{I}\}$.

Oughts will be generated from pairs of the form $\langle \mathcal{W}, \mathcal{I} \rangle$, consisting of a set of propositional formulas $\mathcal{W}$ and a set of imperatives $\mathcal{I}$. The shape of $\mathcal{W}$ will differ, depending on whether the situation modeled is moral or epistemic: In the former case this set will contain formulas expressing the descriptive facts of the situation; in the latter it will contain formulas expressing facts about the evidential support relation. We will refer to pair of the form $\langle \mathcal{W}, \mathcal{I} \rangle$ as *contexts* and denote them using the letter $c$ (with subscripts). Given that $\mathcal{W}$ specifies the normatively relevant facts, it's natural to think that it should determine which imperatives from $\mathcal{I}$ are in force. The following function captures this thought formally:

Where $c = \langle \mathcal{W}, \mathcal{I} \rangle$ is a context, the imperatives that are *triggered* in it are those that belong to the set $Triggered(c) = \{i \in \mathcal{I} : \mathcal{W} \vdash Antecedent[i]\}$.

At this point an example is in order. Imagine that Holmes finds himself resolving a regular sort of case under normal circumstances—so no reasoning distorting drugs are involved—and that all the evidence points to one person: the good old butler. Letting $b$ stand for the proposition that the butler did it, we can express the proposition that Holmes's evidence supports believing that the butler did it with the formula $E \Mapsto B(b)$. The relevant ideal—an instance of the evidential requirement (ER) discussed above—is $i_1 = !(B(b) | E \Mapsto B(b))$, and it says that, ideally, Holmes should believe that the butler did it, in case his evidence supports believing that he did. The toy example itself can, then, be encoded in the context $c_1 = \langle \mathcal{W}, \mathcal{I} \rangle$ where $\mathcal{W}$ is the set $\{E \Mapsto B(b)\}$ and $\mathcal{I}$ is $\{i_1\}$. It's immediate to see that $Triggered(c_1) = \{i_1\}$.

The next question we must ask is how to get to all things considered oughts from this. Intuitively, such oughts should be determined by the satisfaction conditions of those imperatives that are in force, or triggered—in the particular case at hand, the satisfaction conditions of the only triggered imperative $i_1$. Once we recall that the function $Consequent[\cdot]$ tells us just what those satisfaction conditions are, the following definition should look natural:

Where $c = \langle \mathcal{W}, \mathcal{I} \rangle$ is a context, $\bigcirc X$ follows from $c$ if and only if $\mathcal{W} \cup Consequent[Triggered(c)] \vdash X$.

This definition is on the right track, and it gives us the intuitively correct result for $c_1$. Since $Consequent\big[\{i_1\}\big]$ equals $\{B(b)\}$, the formula $\bigcirc B(b)$ follows from the context. However, we need to amend the definition for the logic to actually deal with cases involving *conflicting* requirements. This becomes manifest once we apply the definition to another toy example, this time from the practical domain:

> *Twins.* You have inadvertently promised to have a private dinner with each of two twins. Both twins are equally important to you, and both would be equally disappointed by your cancellation.[27]

Let $p_1$ and $p_2$ stand for the propositions, respectively, that you have promised to dine with the first twin and that you have promised to dine with the second twin. And let $d_1$ and $d_2$ stand for the propositions, respectively, that you dine with the first twin and that you dine with the second. The case itself can be encoded in the context $c_2 = \langle \mathcal{W}, \mathcal{I} \rangle$ where $\mathcal{W} = \{p_1, p_2, \neg(d_1 \& d_2)\}$ and $\mathcal{I}$ contains the imperatives $i_2 = !(d_1 \,|\, p_1)$ and $i_3 = !(d_2 \,|\, p_2)$, with $i_2$ saying that you are under a standing demand to dine with the first twin, if you've promised to dine with them, and $i_3$ expressing a similar demand vis-à-vis the second twin. Note that the formula $\neg(d_1 \& d_2)$ serves as a constraint saying that it is impossible for you to dine with both twins. Both imperatives $i_2$ and $i_3$ get triggered in the context $c_2$, and now there's a problem. Combining their satisfaction conditions with $\mathcal{W}$ in the way the definition requires results in an inconsistent set. Such sets entail all formulas, and so our definition has the counterintuitive consequence that the context $c_2$ entails a formula of the form $\bigcirc X$ for any $X$ whatsoever.

What's at the root of the problem? Well, it's the fact that the definition requires that we use *all* the statements in $Consequent\big[Triggered(c)\big]$, a set that, in general, can be inconsistent (or inconsistent with $\mathcal{W}$). So we need some sort of restriction. And what we'll do is fall back from the entire set $Consequent\big[Triggered(c)\big]$ to its largest consistent parts. The formal concept we'll be relying on is that of a maximally consistent, or maxiconsistent, subset:[28]

> Where $\mathcal{G}$ and $\mathcal{H}$ are two sets of propositional formulas, a subset $\mathcal{F}$ of $\mathcal{H}$, $\mathcal{F} \subseteq \mathcal{H}$, is said to be maxiconsistent with $\mathcal{G}$ if and only if (i) $\mathcal{F}$ is consistent with $\mathcal{G}$, and (ii) there is no consistent set $\mathcal{F}'$ such that $\mathcal{F} \subset \mathcal{F}'$ and $\mathcal{F}' \subseteq \mathcal{H}$.

---

27. The example is adopted from Horty (2012: 30).
28. More precisely, we're relying on a generalization of the concept from Makinson (2005: 30ff.).

The concept is of most use when $\mathcal{G}$ and $\mathcal{H}$ are individually consistent, while $\mathcal{G} \cup \mathcal{H}$ isn't. Intuitively, a subset of $\mathcal{H}$ that's maxiconsistent with $\mathcal{G}$ is as big a subset of $\mathcal{H}$ as you can add to $\mathcal{G}$ without running afoul of inconsistency: Supplementing it with even one additional formula from $\mathcal{H}$ would render the result inconsistent—this is what clause (ii) ensures.

With this concept in hand, we can define which oughts follow from a given context $c$ by focusing *not* on the entire set $Consequent\big[Triggered(c)\big]$, but on those of its subsets that are maxiconsistent with $\mathcal{W}$. The plural is not accidental. An inconsistent set can have multiple maxiconsistent subsets, and our policy is to require that a statement $X$ must follow from *all* such subsets if $\bigcirc X$ is to qualify as an all things considered ought. That is:

> Let $c = \langle \mathcal{W}, \mathcal{I} \rangle$ be a context. Then $\bigcirc X$ follows from $c$ if and only if $\mathcal{W} \cup \mathcal{F} \vdash X$ for *every* subset $\mathcal{F}$ of $Consequent\big[Triggered(c)\big]$ that is maxiconsistent with $\mathcal{W}$.

Let's revisit our examples. First, notice that nothing has changed with regard to the context $c_1$. There's only one subset of $Consequent\big[Triggered(c_1)\big] = \{B(b)\}$ that's maximally consistent with $\mathcal{W}$, namely, $\{B(b)\}$ itself, and so $\bigcirc B(b)$ still follows from $c_1$. This illustrates the conservative character of the amendment. But what about $c_2$? As before, both imperatives $i_1 = !(d_1 \,|\, p_1)$ and $i_2 = !(d_2 \,|\, p_2)$ get triggered, and we get $Consequent\big[Triggered(c_2)\big] = \{d_1, d_2\}$. This set has two subsets that are maxiconsistent with $\mathcal{W} = \{\neg(d_1 \,\&\, d_2)\}$, namely, $\{d_1\}$ and $\{d_2\}$. What is it, then, that you ought to do in the Twins case? Since neither $d_1$, nor $d_2$ follows from both $\{d_1, \neg(d_1 \,\&\, d_2)\}$ and $\{d_2, \neg(d_1 \,\&\, d_2)\}$, the context doesn't entail either $\bigcirc d_1$ or $\bigcirc d_2$. So it's not the case that you ought to dine with the first twin, nor is it the case that you ought to dine with the second. However, the disjunction $d_1 \vee d_2$ does follow from both of these sets, and so the context entails $\bigcirc(d_1 \vee d_2)$. This means that you can't just walk away; you have to keep one of your promises. There's also a sensible rationale behind this recommendation. If you comply with the ought—in either of the two ways—you keep as many promises as is humanly possible in your situation. You also break as few promises as you possibly can.

## 5. Conflicting-Ideals Response to the Puzzle

With the logic at our disposal, we can return to the puzzle and Prof. Moriarty's Drug. We'll encode the case into the context $c_3 = \langle \mathcal{W}, \mathcal{I} \rangle$. Its first element $\mathcal{W}$ will contain the formulas expressing the relevant facts about evidential support, $E \mathrel{\vert\!\!\Rightarrow} B(m)$ and $E \Rightarrow B(E \not\Rightarrow B(m))$. Notice that this is the last component of the

puzzle (MTE). The context's second element, in turn, will contain the relevant instances of the requirements (ER) and (ILC). In Section 3, we expressed them as the strict requirements:

(3) $\left[E \Rightarrow B(m)\right] \supset \bigcirc B(m)$,

(4) $[E \Rightarrow B(E \not\Rightarrow B(m))] \supset \bigcirc B(E \not\Rightarrow B(m))$, and

(5) $\bigcirc[B(E \not\Rightarrow B(m)) \supset \neg B(m)]$

Adopting (3) and (4) to the new setting, we acquire $i_4$ and $i_5$:

$i_4$: $!(B(m) | E \Rightarrow B(m))$,

$i_5$: $!(B(E \not\Rightarrow B(m)) | E \Rightarrow B(E \not\Rightarrow B(m)))$.

The imperative $i_4$ says that, ideally, Holmes should believe that the maid did it in the circumstances where his evidence supports believing that she did. Similarly, the imperative $i_5$ says that, ideally, Holmes should believe that his evidence doesn't support believing that the maid did it in the circumstances where his evidence supports such a belief.

The requirement (ILC) is a little more tricky. We represent it as the defeasible imperative $i_6$:

$i_6$: $!((B(E \not\Rightarrow B(m)) \supset \neg B(m)) | \top)$.

The symbol $\top$ here stand for an arbitrary tautology; and given that $\top$ follows from any set of formulas, an imperative that has it in the antecedent is guaranteed to get triggered. So, on our rendering of (ILC), there's always a standing defeasible demand that one's second-order beliefs and first-order attitudes cohere with each other—note too that $i_6$ is equivalent to $!(\neg(B(E \not\Rightarrow B(m))) \,\&\, B(m)) | \top)$. This is also a very natural way to represent wide-scope requirements in our framework.[29]

All in all, then, we have the context $c_3 = \langle \mathcal{W}, \mathcal{I} \rangle$ where $\mathcal{W}$ contains the expressions $E \Rightarrow B(m)$ and $E \Rightarrow B(E \not\Rightarrow B(m))$, while $\mathcal{I}$ contains the imperatives $i_4$, $i_5$, and $i_6$. But how does this help with solving the puzzle? Well, recall that the problem was that we were able to conclude that Holmes both ought to believe that the maid did it and ought to avoid having this belief, $\bigcirc B(m)$ and $\bigcirc \neg B(m)$. Now let's see what all things considered oughts follow from $c_3$. As a first step,

---

29. Cf. Schroeder (2018).

notice that $i_4$, $i_5$, and $i_6$ all get triggered, and that $Consequent\big[Triggered(c_3)\big]$ has three subsets that are maxiconsistent with $\mathcal{W}$, namely,

$$\big\{B(m), B(E \not\Mapsto B(m))\big\},$$

$$\big\{B(m), B(E \not\Mapsto B(m)) \supset \neg B(m)\big\}, \text{ and}$$

$$\big\{B(E \not\Mapsto B(m)), B(E \not\Mapsto B(m)) \supset \neg B(m)\big\}.$$

A minute's reflection reveals that the third set doesn't entail $B(m)$ and that the first doesn't entail $\neg B(m)$. This means that neither of the problematic oughts, $\bigcirc B(m)$ and $\bigcirc \neg B(m)$, follow from $c_3$. It is not the case that Holmes ought to believe that the maid did it. Nor is it the case that he ought *not* to believe that the maid did it. However, as in the twins example, there's a disjunction that follows from all three sets, namely,

$$[B(m) \,\&\, B(E \not\Mapsto B(m))] \vee [B(m) \,\&\, \neg B(E \not\Mapsto B(m))] \vee [\neg B(m) \,\&\, B(E \not\Mapsto B(m))].$$

And this means that the context $c_3$ entails the following ought:

$$\bigcirc([B(m) \,\&\, B(E \not\Mapsto B(m))] \vee [B(m) \,\&\, \neg B(E \not\Mapsto B(m))] \vee [\neg B(m) \,\&\, B(E \not\Mapsto B(m))] ).$$

So Holmes can't adjust beliefs as he pleases; he has to do it in one of the specified ways. Unpacking the formula, he ought to either (i) believe both that the maid did it *and* that the evidence doesn't support believing that she did; or (ii) believe only that the maid did it; or (iii) believe only that the evidence doesn't support believing that the maid did it. Put differently, there are two relevant beliefs—that the maid did it and that the evidence doesn't support believing that she did—and Holmes is as he ought to be as long as he holds at least one of them.

Notice that this disjunctive recommendation can be supported by a rationale that parallels the one we appealed to in the twins case: By adjusting beliefs in *any* of the three specified ways, Holmes would satisfy as many instances of rationality requirements as he can and also violate as few of them as he can, given the situation. (In case you still find this recommendation utterly implausible, note that the formal model does *not* commit us to it. Equally well, it can deliver a stronger recommendation, that is, that *only one* of (i)–(iii) be followed. This is the topic of the next section.)

Even though Holmes would end up violating a requirement no matter what he does, this doesn't mean that we should classify his response as irrational. Think back to the twins case. Suppose you wind up calling off your rendezvous with the first twin and dining with the second. There's something unfortunate

about your response—you have broken one of your promises—and yet it is an optimal response to the situation you were in. Similarly, we can suppose that Holmes complies with (iii) and ends up believing that the evidence doesn't support believing that the maid did it. There's something unfortunate about his response—he violates the evidential requirement by not having a belief supported by the evidence—and yet it is an optimal response to the situation he is in. On the perspective that comes with the logic, requirements are defeasible. So, when assessing the agent's rationality, we should not be looking at whether or not she complies with all the requirements that are in force in her situation—formally, all the imperatives in $Triggered(c)$ —but rather at whether or not the agent complies with the generated oughts. Thus, if any of (i)–(iii) obtain, Holmes's response is fully rational, and that's enough to conclude that Moriarty's Drug is not a dilemma.

So we have a solution to the puzzle that concedes that the possibility of radically misleading total evidence (MTE) leads to a conflict between the epistemic requirement (ER) and the inter-level coherence requirement (ILC), and yet does not qualify it as a dilemma. Its advantages over the alternatives should be obvious: It lets us preserve (MTE), (ER), (ILC), as well as the unity of epistemic rationality.

The solution changes the logic governing the interaction between (ER) and (ILC), or, what's in effect the same thing, proposes that we think of them not as strict requirements specifying what doxastic attitudes one ought to have all things considered, but as defeasible requirements specifying what attitudes one should have ideally, requirements that interact to jointly determine the all things considered epistemic oughts—much like moral reasons are standardly taken to determine the all things considered moral oughts. I think it is very natural to call these defeasible requirements *ideals* and to think of the solution just developed as a mathematically precise characterization of the conflicting-ideals response to the puzzle. My suggestion, then, is that we haven't only solved the puzzle, but also taken some steps toward making the conflicting-ideals view itself more transparent. That is, I think, that it is best understood as a move away from the default metaepistemological opinion, according to which epistemic requirements are strict and governed by some rather strong, but never explicitly stated logic, toward the more unconventional metaepistemological view, according to which epistemic requirements are defeasible and governed by a comparatively weak logic. I don't think this interpretation of the view is standard: In the end, in the literature focusing specifically on the puzzle, Christensen (2007; 2013) is usually interpreted as conceding the existence of genuine rational dilemmas.[30]

---

30. See Footnote 9 for references.

## 6. Disjunctive Oughts and Weighted Requirements

The context encoding Moriarty's Drug entails a disjunctive ought, suggesting that there are three equally rational ways for Holmes to respond to the scenario. One might be worried by this on at least three grounds. First, one could feel that this diagnosis is simply intuitively implausible. Second, one could insist that this makes the logic I've presented a poor candidate for the formal backbone of the conflicting-ideals view, since the latter isn't normally associated with disjunctive recommendations. And third, one could point out that this comes with a commitment to permissivism—or, roughly, the view that there are at least some bodies of evidence that sanction holding different doxastic attitudes toward the same proposition—which we may have independent reasons to reject.[31] The goal of this section is to respond to these worries by showing how a simple twist to the logic can strengthen its recommendations, narrowing down the number of rational ways to respond to Moriarty's Drug to only one.

In the moral domain, there's a general, if not universal agreement, that imperatives aren't only defeasible, but can also have relative weights. Consider the following timeworn example:

> *Drowning Child.* You have promised a friend to have a dinner with her. Your route to the restaurant takes you past a pond, and, as you are walking past it, you notice that a child has fallen in. The child is crying in distress, and all your evidence suggests that she is going to drown, unless you do something about it. However, if you rescue the child, you will get your clothes wet and muddy, and won't make it to the dinner.

It seems clear that, in this case, what you ought to do all things considered is to save the child, thereby failing to keep your promise. But now let's take a look at how our logic would treat this case. Let $p$ and $d$ stand for the propositions, respectively, that you have a made promise to dine with your friend, and that you dine with her, and let $c$ and $s$ stand for the propositions that a child is drowning, and that you save the child. The imperative $i_7 = !(d \mid p)$ would then express the demand that you dine with the friend in case you've promised, and $i_8 = !(s \mid c)$ the demand that you save the child, given the need. We could try to encode the case into the context $c_4 = \langle \mathcal{W}, \mathcal{I} \rangle$ with $\mathcal{W}$ comprising $p$, $c$, and $\neg(d \& c)$ and $\mathcal{I}$ containing $i_7$ and $i_8$. However, this wouldn't give us the intuitive result, since what follows from $c_4$ is only the disjunctive ought $\bigcirc(d \lor s)$. Not

---

31. More precisely, the worry here is that the view doesn't simply comes with a commitment to permissivism, but a pretty radical version of it. We'll discuss this in more detail in the next section.

surprisingly, the problem with $c_4$ is that it doesn't reflect the relative weights of the two demands.

But this problem can be fixed by endowing the context with more structure. To this end we introduce a new formal device: a priority ordering $<$ over the set of imperatives $\mathcal{I}$.[32] Henceforth, $i < i'$ will mean that the imperative $i'$ has more weight than the imperative $i$, and the statement $i < \mathcal{J}$ will mean that every imperative in the set $\mathcal{J}$ has more weight than $i$. From now on, then, contexts will be triples of the form $\langle \mathcal{W}, \mathcal{I}, < \rangle$. The Drowning Child in particular will be expressed as $c_5 = \langle \mathcal{W}, \mathcal{I}, < \rangle$ which extends $c_4$ with the ordering on the imperatives $i_7 < i_8$.

But how shall we use an ordering when determining which all things considered oughts follow from a context? In a word, it will serve as a filter for triggered imperatives. When multiple triggered imperatives come into conflict, we'll keep those that rank the highest and discard those that rank the lowest. This simple idea is implemented in the following function:

> Where $c = \langle \mathcal{W}, \mathcal{I}, < \rangle$ is a context, the imperatives that are *binding* in it are those that belong to the set

$$Binding(c) = \{i \in \mathcal{I}: \; (1) \; i \in Triggered(c)$$
$$(2) \; \text{there's no } \mathcal{J} \subseteq Triggered(c) \text{ such that}$$
$$(a) \; i < \mathcal{J},$$
$$(b) \; Consequent[\mathcal{J}] \text{ is consistent with } \mathcal{W}, \text{ and}$$
$$(c) \; \mathcal{W} \cup Consequent[\mathcal{J}] \vdash \neg Consequent[i]\}.$$

In English: To enter the set $Binding(c)$ an imperative $i$ must (1) be triggered, and (2) there can't be a set of triggered imperatives $\mathcal{J}$ that are uniformly better than $j$, jointly consistent (with $\mathcal{W}$), and such that they entail the negation of $i$'s conclusion.

With this function in hand, the final amendment to the definition of the consequence relation from Section 4 is straightforward. All we do is substitute $Binding(c)$ for $Triggered(c)$:

> Let $c = \langle \mathcal{W}, \mathcal{I}, < \rangle$ be a weighted context. Then $\bigcirc X$ follows from $c$ if and only if $\mathcal{W} \cup \mathcal{F} \vdash X$ for *every* subset $\mathcal{F}$ of $Consequent[Binding(c)]$ that is maxiconsistent with $\mathcal{W}$.

---

32. Following the literature, I assume that $<$ is transitive and irreflexive, but not necessarily complete. See, for instance, Horty (2012: 27–32) for a discussion here.

Let's apply this definition to the Drowning Child scenario. Both $i_7$ and $i_8$ get triggered in $c_5$, but only the latter qualifies as binding. Notice how $i_7$ doesn't qualify because of the singleton set $\{i_8\}$. First, it is a subset of $Triggered(c_5)$. Second, every imperative in it has more weight than $i_7$, that is, $i_7 < \{i_8\}$. And third, $Consequent[\{i_8\}] = \{s\}$ is consistent with $\mathcal{W}$ and together with $\mathcal{W}$ it entails $\neg d$, the negation of $Consequent[i_7]$. From here it is but one step to see that $\bigcirc s$ does, while $\bigcirc d$ does not follow from $c_5$, as desired.

Now let's turn to the epistemic domain. Once we have stepped away from the assumption that rationality requirements have to be strict, it is natural to think that one requirement (or ideal) can have more weight than another, or that a particular instance of a requirement can have more weight than another. In particular, we might reasonably suspect that the requirements in force in Moriarty's Drug may have different (relative) weights. And if these weights are indeed different, then this should be reflected in the context capturing the scenario. So it seems perfectly reasonable to insist that the context $c_3$, which we used to capture Holmes's situation in Moriarty's Drug, has to be extended with an ordering if the scenario is to be captured in full, and that, once it is thus extended, Holmes will have only one rational response.[33] The difficult question, then, is what should this ordering be.

At first one might think that any assignment of relatives weights to instances of epistemic requirements has to be uniform and motivated on independent grounds. For us here this would mean that the ordering on $c_3$ would need to be supported by an argument for one of the following two claims: (ER) must always take precedence over (ILC). Or (ILC) must always take precedence over (ER). However, both claims are in tension with the spirit of the conflicting-ideals view and our solution to the puzzle. To see why, suppose that we did hold that in every situation of conflict between the defeasible (ER) and (ILC) the former wins out. It would be very natural to ask what sort of work the defeasibility of (ER) is even doing and to wonder if the resulting position isn't better thought of as one that denies that (ILC) is a genuine requirement. So I think that, once we have embraced the idea that (ER) and (ILC) are ideals that occasionally come into conflict, we can't uniformly prioritize one over the other.

What we can do, however, is hold that every particular conflict between instances of (ER) and (ILC) gets resolved, with (ER) winning out in some of them and (ILC) in others and with the winner being determined by the details of the case. Notice that this is just what we observe in cases of conflict between imperatives in the moral domain. In the Drowning Child scenario it was clear that saving the child and not keeping the promise was the only right thing to do. But

---

33. Generally, adding an ordering on imperatives results in a stronger consequence set, and so an ordering on imperatives in $c_3$ would lead to Holmes having fewer rational responses.

this doesn't mean that considerations of benevolence are always more important than those of promise-keeping. In fact, it's very easy to think of scenarios where keeping a promise and not doing the good seems like the only right thing to do. (Suppose you knew that the child wasn't in real danger, and that it would only calm down a bit if you helped.) So sometimes benevolence wins out, and at other times promise-keeping does.[34] And it seems perfectly reasonable to hold that every particular conflict between benevolence and promise-keeping has a correct resolution.

What are we to say about assigning weights to the imperatives in $c_3$ in light of this? I think the correct answer is that Moriarty's Drug is too underdescribed— and in a way that's typical of other examples illustrating (MTE) in the literature—for us to say what these weights should be. We know that Holmes's first-order evidence, or the clues that he finds at the manor, supports believing that the maid did it, but we don't know what exactly this evidence is. We know that Holmes knows how Moriarty's drug works, but we don't know what evidence exactly this knowledge is based on. We know that Holmes is an excellent detective, but we don't know if he cares more about believing truths or avoiding error. And we know that he wants to solve the case, but we don't know what's at stake for him in it.[35] A fuller description of the scenario would put us in a position to say what the weights of the relevant instances of (ER) and (ILC) are. But, in the absence of such a description, we can't (and shouldn't) say what they are or what exactly the one rational doxastic response to Moriarty's Drug is.

At this point one might object that, quite independently of the details suppressed in the description, it's intuitively very clear that Holmes's only rational response in Moriarty's Drug is to believe that his evidence doesn't support believing that the maid did it and not to believe that she did. But if there's such an intuition, it may stem from the implicit assumption that Holmes's first-order evidence

---

34. Interestingly, W. D. Ross appears to have thought that promise-keeping is standardly more important than doing the good: "Ross . . . held that ordinarily considerations of promising should take precedence over those of beneficence (it being more important ordinarily to keep one's promise than to do good); but he also supposed that there could be such a thing as a trivial breach of promise and a very large good to be done, in which case, he thought, the promise should be broken" (Dancy 2004: 27).

35. I'm only hinting at some factors that *could* play a role in determining the relative weights of requirements. I don't have a worked-out theory to offer here. However, I do find it plausible that evidential support could contribute to how requirements are weighed: When both $E \Rightarrow B(X)$ and $E \Rightarrow B(Y)$ obtain, it doesn't yet mean that $E$ supports the beliefs $B(X)$ and $B(Y)$ equally well. And it seems perfectly reasonable to hold that the relative degree of support that the evidence lends to $B(X)$ and $B(Y)$ is what will determine the relative weights of the corresponding requirements $!(B(X)|[E \Rightarrow B(X)])$ and $!(B(Y)| [E \Rightarrow B(Y)])$. I am also sympathetic to the idea that Holmes's valuing of the goals of believing truths and avoiding error might determine the weighing—I'll say more about this in the next section—and, in light of the debate about pragmatic encroachment, to the idea that pragmatic factors might be playing a role here too.

has to be fairly complex and that only relatively elaborate reasoning can get one from it to the conclusion that the maid did it—in the end, that's what we typically see in murder mysteries. But suppose that the evidence pointing to the maid as the likely culprit was overwhelming: She had a clear motive, there are no other suspects, multiple witnesses report her having had obsessive thoughts of violence, and, on top of that, there's a video of the murder caught on a security camera with the murderer looking just like the maid. Further, suppose that Holmes's knowledge of the drug's effects was based on cases where the affected detectives reasoned about complex bodies of evidence. Would we still say that it would be irrational for Holmes to believe that the maid did? My own intuition says no.

Let me close this section by discussing two sample assignments of weights to $c_3$. The first can be thought of as capturing the fuller description that I started sketching in the previous paragraph: Let $c_6 = \langle \mathcal{W}, \mathcal{I}, < \rangle$ be a context that extends $c_3$ with the following ordering on the imperatives: $i_6 < i_5 < i_4$. Here both instances of (ER) have more weight than the instance of (ILC), and the instance of (ER) concerned with the first-order belief that the maid did it has more weight than the one concerned with the second-order belief. A by now routine check will convince you that only the two strongest imperatives $i_4$ and $i_5$ get selected as binding, and that the context $c_6$ entails the formula $\bigcirc[B(m) \& B(E \not\Rightarrow B(m))]$. So there's only one way for Holmes to be rational. He must believe that the maid did it, as well as believe that his evidence doesn't support believing that she did. The second sample assignment would correspond to an alternative fuller description—perhaps one where the first-order evidence is indeed complex and, thus, calls for an elaborate chain of reasoning and where Holmes cares more about avoiding error than believing truths. Let $c_7 = \langle \mathcal{W}, \mathcal{I}, < \rangle$ be a context extending $c_3$ with the following ordering: $i_4 < i_6 < i_5$. So here the instance of (ILC) has more weight than one of the instances of (ER), but not the other. Yet again, only the two strongest imperatives $i_5$ and $i_6$ get selected as binding, but this time the context entails the formula $\bigcirc[\neg B(m) \& B(E \not\Rightarrow B(m))]$. So, there's, again, only one way for Holmes to respond rationally: He must have the second-order belief and avoid believing that the maid did it.[36]

---

36. An anonymous referee worries that the last three paragraphs reveal a problem with my solution to the puzzle: It lets us deal with cases where Holmes has misleading evidence about what his total evidence supports only at the expense of assuming that Holmes has no normative uncertainty about what he ought to believe. They wonder how much progress we have really made, given that the solution leaves us without an answer to puzzling cases in the vicinity of the original ones, namely, those where Holmes has misleading evidence about what he ought to believe all things considered. I have two things to say in response. First, while I do think that it's reasonable to hold that the details of the Moriarty's Drug scenario will determine which doxastic response is the uniquely rational one, I do *not* want to say—nor do I think that I'm committed to saying—that Holmes has access to all of these details, and that it'll be clear to him what the uniquely rational response is. Second, I'm not exactly sure what to make of an abstractly described

Let me emphasize that $c_6$ and $c_7$ are only illustrations, and that I'm not suggesting that either one of them reflects the correct weights. What I'm suggesting instead is that we can reasonably hold that the relative weights of the requirements in play in Moriarty's Drug—whatever they are—ensure that there's only one rational response to it. However, appealing to the relative weights of requirements isn't strictly necessary, in the sense that we have a solution to the puzzle even if we think that defeasible requirements don't have weights.

## 7. Permissivism and Controversial Commitments

We have now seen that my conflicting-ideals view is compatible with two different readings. According to the first, defeasible requirements do not have any weights. According to the second, they do. Call these, respectively, the *nonweighted* and the *weighted view*. You may recall that one of the motivations behind the latter was the former view's apparent commitment to permissivism. But, as a matter of fact, the nonweighted view is not simply committed to permissivism, but a particular version of it that even many permissivists have shunned. So the nonweighted view seems to come with a very controversial commitment. And to make matters worse, one can reasonably suspect that the weighted view doesn't escape this commitment either: The claim that requirements have weights doesn't exclude the possibility of conflicts between requirements whose weights are the same. And given that the weighted view resolves such conflicts in the same way the nonweighted view does—that is, by issuing disjunctive recommendations—it too may lead to the same version of permissivism. So both

---

problem of Holmes having misleading evidence about what he ought to believe, but the literature does talk about cases that would seem to exemplify it. Here's one due to Lasonen-Aarnio: "Assume that evidentialism is correct. Consider whether you ought to believe $p$ in a situation where you have excellent evidence for a nonevidentialist epistemological theory on which you are required to believe $p$ in your current situation, even though your evidence does not support $p$" (2020: 613). Now, it seems to me that the conflicting-ideals view and the formal model can be used to disentangle this case. The first thing we'd need to do is make sense of the claim that evidentialism is correct. And one promising way of going about here is to understand it as the claim that the instances of the evidential ideal (ER) have a lot of (relative) weight in this case, and that they engender all things considered oughts. So what you ought to do is not believe $p$, as well as believe the theory on which you are required to believe $p$. Presumably, this entails that you ought to both believe that you ought to believe $p$ and not believe $p$, meaning that you could be as you ought to be only at the expense of violating an inter-level coherence ideal in the vicinity of (ILC). I don't think that this results commits us to saying that what you ought to do will be transparent to you, and I still think that it's reasonable to hold that variations in your evidence—for instance, in how good your evidence for the nonevidentialist theory is—can lead to a different doxastic response being the uniquely rational one. But a more detailed analysis of cases like this must be left for future work.

versions of the conflicting-ideals view seem to come with the same potentially unacceptable commitment. The goal of this section is to convince you that this does not undermine the view. I'll start with a defense of the nonweighted version and the proceed to defend the weighted one.[37]

The literature on permissivism tends to distinguish between *inter*personal and *intra*personal permissivism.[38] Interpersonal permissivists think, roughly, that it is sometimes rationally permissible for *different individuals* to hold different doxastic attitudes toward some proposition on the basis of the same evidence. Intrapersonal permissivists, in turn, think that it is sometimes rationally permissible for the *same individual* to hold different doxastics attitudes toward some proposition on the basis of the same evidence. Supposing that the epistemic ought ○ satisfies the plausible principle that if something ought to happen, then it's also rationally permissible, it is straightforward to see that the nonweighted view is committed to the latter type of permissivism:[39] According to the nonweighted view, in Moriarty's Drug Holmes ought to hold at least one of the following: the belief that the maid did it and the belief that his evidence doesn't support believing that she did. By the principle, it's rationally permissible for Holmes to hold the first belief only, the second belief only, or both of them. A direct consequence of this is that it is rationally permissible for Holmes to believe that the maid did it, as well as to not believe that the maid did it, on the basis of the same evidence. And if we take not believing here to be a kind of doxastic attitude, perhaps, tantamount to withholding judgment—as I think we should—then we have *intra*personal permissivism.

In the earlier stages of the debate about permissivism, the inter/intra distinction wasn't clearly drawn and some of the better-known permissivists would ward off the objections coming from the anti-permmissivist camp by distinguishing their *inter*personal views from the seemingly less plausible *intra*personal permissivism. This, at any rate, is how Thomas Kelly (2014), Miriam Schoenfield (2014), and some others have handled the objections of Roger White (2005; 2014). So a brief look at the debate naturally leaves one with the impression that *intra*personal permissivism is ill-fated, inviting the conclusion that the nonweighted view is ill-fated too. But there are two sets of considerations that convince me that the situation is actually not so grim. First off, a closer look at the debate reveals that the dialectical landscape in it has undergone considerable changes in the last few years, and that now there's

---

37. This section was added in response to a set worries raised by an anonymous referee. My sincerest thanks for pushing me to explore the relations between the solution to the puzzle and permissivism.

38. Cf. Jackson and Turnbull (in press) and Kopec and Titelbaum (2016).

39. One could try to avoid the commitment by restricting the application of this principle. This response doesn't strike me as very promising, and so I won't pursue it further.

no agreement that *intra*personal permissivism is clearly inferior to its *inter*personal cousin: Authors including Elizabeth Jackson (2019) and Jonathan Weisberg (2020) explicitly defend *intra*personal permissivism.[40] Some of the most compelling examples of permissivist bodies of evidence that the literature currently has to offer would establish *intra*personal permissivism.[41] And, perhaps most importantly, multiple authors have argued that the familiar move toward *inter*personal permissivism—the move that Kelly and Schoenfield make—may not escape the most pressing of White's objections.[42]

The last point brings us to the second set of considerations: There's good reason to think that the sort of *intra*personal permissivism that the nonweighted view entails can respond to the standard objections against permissivism. I'll discuss two of them here.[43] The first one appeals to intuitions about evidential support and runs, roughly, as follows:[44] Clearly, it should only be permissible to have the beliefs supported by one's total body of evidence. But it's impossible for a body of evidence to support believing $X$, as well as disbelieving $X$. In the end, to whatever extent the evidence supports believing $X$, it supports disbelieving $\neg X$. And so it can't ever be rationally permissible for one to believe $X$, given one's evidence, as well as disbelieve $X$, given the same evidence.

The standard *inter*personal permissivist response to this objection is to say that evidential support is not a two-place, but a three-place relation. What's identified as the third relatum differs from one author to another: Subjective Bayesians say that it is the prior probability function,[45] Kelly (2014) that these are the agent's epistemic goals, and Schoenfield (2014) that these are her sets of epistemic standards. But the way the objection is blocked is always the same: $E$

---

40. Jackson defends what she calls "diachronic intrapersonal permissivism", and one might worry that her view is clearly different from the version of permissivism entailed by the nonweighted view. So it's worth highlighting that she is sympathetic to synchronic permissivism too—see Jackson (2019: 12).

41. See the *self-fulfilling beliefs* cases in Kopec (2015) and Raleigh (2017). Cf. Kopec and Titelbaum (2016).

42. More specifically, Simpson (2017), Stapleford (2019), and Weisberg (2020) have all argued that it fails as a response to the arbitrariness objection, which I'm going to discuss below.

43. Why these two? Well, among the many arguments against permissivism, these two have been central from the very beginning of the debate, they seem to have been discussed most widely, and they have standard *inter*personal permissivist responses— cf. Jackson and Turnbull (in press), Kopec and Titelbaum (2016: 193), Schoenfield (2014). Also, more recent anti-permissivist arguments have focused on metaepistemological considerations about the value of rationality on a permissive picture. Given that these arguments seem to apply to *intra*personal and *inter*personal permissivism to an equal extent and that my goal here is to only show that the former isn't worse off than the latter, not much would be gained by discussing them here. For an excellent overview of these metaepistemological objections, as well as responses to them see Thorstad (2019).

44. See Feldman (2007: 231) and White (2005: 447). My presentation follows Schoenfield (2014: 199–200).

45. See Meacham (2014).

doesn't support *X tout court*, but only relative to, say, a set of epistemic standards, and, once we realize that these standards can be different, it seems perfectly possible that there's a body of evidence *E*—peculiar as it may be—that supports *X*, given one set of standards, while supporting ¬*X*, given a different one. This response is of little help to the nonweighted view which is, for better or worse, committed to the idea that evidential support is a two-place relation.[46] However, it's actually not clear if the nonweighted view needs to be helped at all, since the objection is directed against a version of permissivism that is stronger than the one it entails. All that the nonweighted view commits one to is that it can sometimes be permissible for one to hold a belief that *X*, given one's evidence *E*, and also to not hold this belief, given *E*. The claim that it can be rationally permissible for one to believe *X* and disbelieve *X* given the same evidence *E* is much stronger.[47] But can't one reformulate the objection so that it would apply to the weaker claim too? Well, one could try, but I doubt that the result will have much dialectical force in the present context: While an advocate of the conflicting-ideals view may be willing to grant that it can only be permissible to have the beliefs supported by one's total evidence, she will not accept a parallel principle for withholding—which is what the new objection would have to appeal to. Since she thinks that the evidence requirement (ER) isn't strict, she'll want to say that there are cases where it is permissible to withhold toward *X* even when the total evidence supports believing *X*.

The second objection, or, rather, cluster of objections, centers around the idea that a permissivist is committed to thinking that her beliefs are arbitrary.[48] Suppose you think that it's rationally permissible for you to hold two different doxastic attitudes toward *X* on the basis of your evidence—for instance, you think that it's rationally permissible to believe *X* and withhold toward *X*. What can make you settle on one of these attitudes, as opposed to the other? Well, whatever your reasons might be, it seems that they won't have much to do with the truth of *X* or your evidence for it, and so they will be epistemically arbitrary. This is taken to be problematic. Some authors also highlight just how intuitively strange it is for an agent to, say, believe *X*, while simultaneously admitting that her evidence (also) supports withholding toward *X*, and point out that this stance is similar to believing *X*, while simultenously admitting that one's evidence doesn't support this belief.[49] What's more, anti-permissivists argue that

---

46. For some criticism of relativizing evidential support see Weisberg (2020).

47. The distinction at work here corresponds to the familiar distinction between extreme and moderate permissivism—see, e.g., Jackson (2019), Kelly (2014), and Kopec and Titelbaum (2016).

48. See Feldman (2007), Simpson (2017), Stapleford (2019), Weisberg (2020), and, especially, White (2005; 2014).

49. See Simpson (2017: 522) and Stapleford (2019: 343) who call the agent's stance "quasi-akratic".

allowing for arbitrariness in the agent's doxastic state commits one to further implausible claims. Perhaps, the most striking of them is that individuals are allowed to "toggle" between the doxastic attitudes that they are rationally permitted to have as it suits them.[50] Suppose you really do think that you're in a situation where believing and disbelieving are both rationally permissible. Then there should be nothing wrong with believing *X* at one point, then bringing it about that you disbelieve *X*—which, let's suppose, you can do by means of popping a belief-inducing pill—and then, again, going back to believing *X*. But, of course, it seems intuitively wrong to say that it can be rational for you to believe *X* at one time and then disbelieve it at another, without any change in your body of evidence.

According to the standard *inter*personal permissivist response to these worries, an individual will never think that it's rational *for her* to have two different doxastic attitudes toward the same proposition. The third relatum of the evidential support relation—whatever it is—always limits the range of attitudes that are permissible for her to have to only one.[51] But what should an advocate of the nonweighted view make of these worries? Well, I actually see nothing devastating in admitting that there's a degree of epistemic arbitrariness in situations where multiple doxastic responses are rational. The first point to highlight here is that those situations aren't typical and common, but, rather, peculiar and rare. And Moriarty's Drug can serve as an illustration of just how strange they can be: Holmes's settling on whether to believe that the maid did or to withhold might be arbitrary, but his evidence also radically misleads about what it itself supports! So allowing for some epistemic arbitrariness doesn't mean conceding that it's a pervasive phenomenon.

---

50. The literature discusses at least two other implausible commitments: A permissivist is said to be committed to denying the accuracy of her own belief-forming processes, as well as saying that appraising evidence and popping belief-inducing pills are equally good methods of arriving at truth. I set these two alleged commitments aside, as the arguments leading to them seem to be relying on a way of thinking about evidence and evidential support that the nonweighted view is not committed to. Also, the toggling objection seems to have received the most attention in the literature.

51. Some people suggest that this only pushes the worries away without addressing them. Thus, those *inter*personal permissivists who think that the third relatum is the set of agent's epistemic standards are faced with the following challenge: If an agent is aware that her evidence is permissive regarding *X*, she will still have to admit that there's something arbitrary about her rationally held attitude toward *X*, and that it could've been different if her epistemic standards were different. But, then, there would seem to be nothing wrong with bringing it about that those standards are different, and then toggling back again. Some *inter*personal permissivists are aware of this worry and have tried to respond to it. However, the jury is still out on whether any of their responses succeed. See Schoenfield (2014) and Simpson (2017) for such responses, as well as Stapleford (2019) and Weisberg (2020) for criticism.

Further, I don't think that an advocate of the nonweighted view has much reason to worry about the parallel between agents who acknowledge that their attitude is arbitrary and those who hold an attitude while simultaneously admitting that it is not supported by the evidence. For starters, she doesn't think that the latter type of agent is necessarily irrational. But more importantly, it looks like she could use the parallel to her own advantage, as her view has the resources to explain why the stances of such agents strike us as similar—at least, in cases that have the structure of Moriarty's Drug. On the conflicting-ideals view, an agent who believes $X$ while admitting that her evidence doesn't support believing that $X$ strikes us as doing something wrong because she violates the inter-level coherence ideal (ILC). Now suppose that Holmes responds to Moriarty's Drug by believing that his evidence doesn't support believing that the maid did it, $B(E \not\Rightarrow B(m))$, and withholding on whether or not she did, $\neg B(m)$. Presumably, Holmes would admit that his attitude toward $m$ could've been different, and, specifically, that he could've believed it. Why does this seem quasi-akratic to us? I see two explanations one might give. First, there's a rational response in the vicinity of Holmes's actual one where he does violate the inter-level coherence ideal. And second, were Holmes to satisfy the one ideal that he's currently violating—the instance of (ER) $i_4 =!(B(m)\,|\,E \Rightarrow B(m))$—he'd thereby violate (ILC).

Another fact that, I think, an advocate of the nonweighted view could use to her advantage is that a degree of arbitrariness is not only allowed for, but even thought to be necessary in the practical realm. This is witnessed by the Buridan's ass scenario and its analogues in the moral domain. Since the piles of hay are identical, the donkey's reason for choosing one over the other must be prudentially arbitrary. However, the facts that the donkey's settling on the left pile is arbitrary and that it could've easily settled on the right pile hardly speak against the donkey's choice or the possibility of cases where more than one response is rationally permissible. Similarly, in a situation where you can save (at most) one of two people $A$ and $B$ from imminent death—and where all else is equal: no special obligations, no differences in likelihood of success, and so on—your settling on which person to save will also contain an element of arbitrariness.[52] But it would be strange, to say the least, to point to the arbitrary component in your saving person $A$ (and not $B$) and the fact that you could've saved $B$ to argue that there's something wrong about your saving $A$, that tragic scenarios of this sort can never arise, or that morality doesn't prescribe that you save at least one

---

52. There's a debate about whether or not you are required to give $A$ and $B$ an equal chance to be saved in such cases by holding a lottery—see, e.g., Henning (2015). As far as I can see, this debate is orthogonal to the point I'm making. If a lottery is held, its outcome is the arbitrary element in settling on which person to save.

person.[53] So I think that an advocate of the nonweighted view can invoke these cases to push back against the general worry and demand for an explanation of what is it exactly that makes *any* degree of arbitrariness utterly unacceptable in the epistemic domain.

What's more, the practical domain suggests two possible responses to the toggling objection. First, we could draw a parallel with the moral domain to argue that the fact that it is rationally permissible for one to have multiple dox-astic attitudes doesn't entail that toggling between them is permissible. There is no moral reason to save *A* and not *B*. But this doesn't entail that "toggling" between saving them is morally permissible. Once you've settled on saving *A*, it appears to be morally wrong for you to switch to saving *B*. In the same way the fact that there are two attitudes that are rationally permissible for Holmes to hold toward the proposition that the maid did it doesn't mean that it's rationally permissible for him to switch from one of them to another as he likes. Second, we could draw a parallel with the prudential domain and argue in the opposite direction: In the Buridan's ass scenario, there would seem to be nothing wrong with the donkey eating from one pile of hay, then switching to the other, and then switching back to the first. Similarly, there might be nothing epistemically wrong with Holmes toggling between the doxastic states that are rationally per-missible for him in the Moriarty's Drug scenario—which, once again, is a strange case to begin with. As Jackson (2019) has pointed out, anti-permissivists say that toggling situations are absurd, but beyond that don't give much argument for thinking that toggling is epistemically irrational.[54]

All in all, then, the nonweighted view does lead to a version of *intra*personal permissivism, but there's good reason to think that this permissivism is not unacceptable. With the (recent) changes in the dialectical landscape, the posi-tions of the *inter*personal and *intra*personal views seem to have evened out. And we have seen some good reason to think that an advocate of the nonweighted view can respond to the standard objections against permissivism.

Now let's move on to the weighted version of the view. Recall the concern: In the absence of an independent reason to think that defeasible requirements can't have the same weight, the weighted view too may be committed to *intra*personal permissivism. I'll confine my response to three brief counterconsiderations here.

---

53. At least one author in the permissivism debate appears to suggest that, in at least some scenarios that we might be tempted to classify as permissive between two doxastic attitudes, epis-temic rationality in fact prescribes taking neither of them—see Stapleford (2019: 353).

54. Jackson refers to Roger White in particular—see Jackson (2019: fn. 48). In addition to sketching the two responses to the toggling objection that I discuss, she mentions a third one: The *intra*personal permissivist could concede that toggling is epistemically permissible, while insisting that it is *prudentially* impermissible. She writes, "If one Frequently takes belief-inducing pills that alter one's beliefs at random, this makes decision-making, goal setting, and life planning very difficult".

First, I think that this worry should look less pressing now that we have seen that *intra*personal permissivism is not necessarily an ill-fated position, but, rather, a defensible one. Second, I don't think the worry has as much dialectical force as it initially seems to have, since in so far as it reveals a genuine problem for the weighted view, it reveals a problem for some central *inter*personal permissivist views as well: Recall how the latter avoid the commitment to *intra*personal permissivism. Consider some body of evidence *E* that's permissive with respect to *X*, say, both believing *X* and withholding toward *X* are rationally permissible on its basis. Now zoom in on some particular agent who possesses *E*. Why is it rationally permissible for her to hold only one of these doxastic attitudes? Well, Kelly (2014) will say that this is due to her epistemic goals, Schoenfield (2014) that this due to her set of epistemic standards, and others will point to other factors. Usually, this is taken to suffice as an answer. But should it? Take Kelly's view as example—which I think is representative. Drawing on William James (1897), he distinguishes between the goals of not believing what's false and believing what's true, and suggests that how these goals are valued is agent-relative. Presumably, an agent who values believing truths over not believing falsehoods will be only permitted to believe *X*, given *E*. And an agent whose valuing is reversed will be only permitted to withhold toward *X*, given *E*. But following the objection, we should insist that this doesn't yet mean that an agent couldn't value both goals equally. And as long as this is an open possibility, Kelly's permissivism may still be *intra*personal. Clearly, a consideration can't be taken to undermine (only) the weighted view if it turns out to casts doubt on some of the central permissivist positions.

Finally, we shouldn't forget that the weighted view is quite abstract. While it says that ideals have weights, it's silent on where these weights come from, and it seems natural to expect that the story here could be told in a number of different ways. This is important for the question of whether two requirements can have the same weights because it suggests that we shouldn't expect an answer to it that's both general and informative. Instead, we should expect less abstract answer that appeal to concrete sources of weights. What's more, I suspect that an advocate of the weighted view will be able to reuse at least some of the answers that *inter*personal permissivists might provide to the shared problem that the worry reveals: On one fairly natural reading, the weighted view is an *inter*personal permissivist view formulated at a fairly high level of abstraction. Recall that in the formal model that gave rise to it epistemic situations are represented using weighted contexts, or structures of the form $\langle \mathcal{W}, \mathcal{I}, < \rangle$. We could hold that the first two elements $\mathcal{W}$ and $\mathcal{I}$ of a context represent the agent-neutral features of the situation, and that what the priority ordering $<$ adds to it are the agent-specific factors that *inter*personal permissivists talk about. These could be the agent's weighing of the goals of believing truths and avoiding error, her

epistemic standards, what's at stake to her, or whatever else they say reduces the range of rationally permissible responses to one.[55] On this way of thinking, the recommendations based on $\langle \mathcal{W}, \mathcal{I} \rangle$ reflect the agent-neutral perspective, and the stronger recommendations based on $\langle \mathcal{W}, \mathcal{I}, < \rangle$ reflect the perspective of a given agent. Now, given this connection between the weighted view and *inter*personal permissivism, I see no reason why the former couldn't make use of responses provided in defense of the latter.

## 8. Conclusion

My goal in this paper was to present a solution to an important puzzle that starts with acknowledging the possibility of total evidence that's radically misleading about itself, (MTE), proceeds to a clash between two plausible and widely accepted requirements of rationality, (ER) and (ILC), and arrives at the claim that there are rational dilemmas. Aiming to avoid this conclusion, the existing responses to the puzzle have proposed rejecting (MTE), rejecting one of the requirements, or treating the conflict as a clash between two irreducible types of rationality. The main idea behind my solution was to substitute a defeasible deontic logic for the relatively strong, but never explicitly stated logic governing the interaction between (ER) and (ILC), or, put less formally, suggest that we think of (ER) and (ILC) not as strict requirements specifying what ought to happen all things considered, but as defeasible requirements specifying what should happen ideally. We saw that such defeasible requirements are naturally thought of as epistemic ideals, and that the logic governing their interaction can be thought of as the formal backbone of the conflicting-ideals view. We also saw that this view is compatible with two different readings—the nonweighted one, on which ideals have no weights, and the weighted one, on which they do—as well as how we can address the worry that both of them come with a commitment to an unacceptable version of permissivism.

I envisage four directions in which the work begun here could (and should) be taken in the future. First, the framework we have discussed could be applied to many other puzzles in which epistemic requirements come into conflict. Second, it could be used in the context of the meta-level debate about the existence of epistemic dilemmas.[56] Third, we could explore particular theories of the sources of weights of ideals. And fourth, we might also think about

---

55. Cf. Stapleford (2019: 345) where it's suggested that epistemic standards can be treated as particular ways of assigning weights to various theoretical virtues like simplicity, unification, explanatoriness, fruitfulness.

56. See, e.g., Hughes (2019) and Leonard (2020).

alternative ways of making sense of defeasible rationality requirements, that is, not as ideals. In the end, the perspective that comes with the use of defeasible logic is general, and the conflicting-ideals view may be only one attractive way of filling in its details.

## Acknowledgments

## References

Broome, John (1999). Normative Requirements. *Ratio*, *12*(4), 398–419.
Broome, John (2007). Wide or Narrow Scope? *Mind*, *116*(462), 359–70.
Broome, John (2013). *Rationality through Reasoning*. Wiley Blackwell Publishing.
Chisholm, Roderick (1964). The Ethics of Requirement. *American Philosophical Quarterly*, *1*(2), 147–53.
Christensen, David (2007). Does Murphy's Law Apply in Epistemology? Self-Doubt and Rational Ideals. *Oxford Studies in Epistemology*, *2*, 3–31.
Christensen, David (2010). Higher-Order Evidence. *Philosophy and Phenomenological Research*, *81*(1), 185–215.
Christensen, David (2013). Epistemic Modesty Defended. In David Christensen and Jennifer Lackey (Eds.), *The Epistemology of Disagreement: New Essays* (77–97). Oxford University Press.
Christensen, David (2016). Conciliation, Uniqueness and Rational Toxicity. *Noûs*, *50*(3), 584–603.
Coates, Allen (2012). Rational Epistemic Akrasia. *American Philosophical Quarterly*, *49*(2), 113–24.
Conee, Earl and Richard Feldman (2004). *Evidentialism: Essays in Epistemology*. Oxford University Press.
Conee, Earl and Richard Feldman (2008). Evidence. In Quentin Smith (Ed.), *Epistemology: New Essays* (83–104). Oxford University Press.
Dancy, Jonathan (2004). *Ethics without Principles*. Oxford University Press.
Feldman, Richard (2005). Respecting the Evidence. *Philosophical Perspectives*, *19*(1), 95–119.
Feldman, Richard (2007). Reasonable Religious Disagreement. In Louise Antony (Ed.), *Philosophers without Gods* (194–214). Oxford University Press.

Fogal, Daniel (2019). Rational Requirements and the Primacy of Pressure. *Mind*, *129*(516), 1033–70.

Foot, Philippa (1983). Moral Realism and Moral Dilemma. *Journal of Philosophy*, *80*(7), 379–98.

Goble, Lou (2009). Normative Conflicts and the Logic of 'Ought'. *Noûs*, *43*(3), 450–89.

Gowans, Christopher (1987). Introduction: The Debate on Moral Dilemmas. In C. Gowans (Ed.), *Moral Dilemmas* (3–33). Oxford University Press.

Henning, Tim (2015). From Choice to Chance? Saving People, Fairness, and Lotteries. *Philosophical Review*, *124*(2), 169–206.

Horowitz, Sophie (2014). Epistemic Akrasia. *Noûs*, *48*(4), 718–44.

Horty, John (1994). Moral Dilemmas and Nonmonotonic Logic. *Journal of Philosophical Logic*, *23*, 35–65.

Horty, John (2003). Reasoning with Moral Conflicts. *Noûs*, *37*(4), 557–605.

Horty, John (2012). *Reasons as Defaults*. Oxford University Press.

Hughes, Nick (2019). Dilemmic Epistemology. *Synthese*, *196*, 4059–90.

Jackson, Elizabeth (2019). A Defense of Intrapersonal Belief Permissivism. *Episteme*. https://doi.org/10.1017/epi.2019.19

Jackson, Elizabeth and Margareta Turnbull (in press). Permissivism, Underdetermination, and Evidence. In Clayton Littlejohn and Maria Lasonen-Aarnio (Eds.), *The Routledge Handbook of the Philosophy of Evidence*. Routledge.

James, William (1897). *The Will to Believe and Other Essays in Popular Philosophy*. University Press, John Wilson and Son.

Kelly, Thomas (2014). Evidence Can Be Permissive. In Matthias Steup, John Turri, and Ernest Sosa (Eds.), *Contemporary Debates in Epistemology* (298–311). John Wiley and Sons.

Kiesewetter, Benjamin (2017). *The Normativity of Rationality*. Oxford University Press.

Kolodny, Niko (2005). Why Be Rational? *Mind*, *114*(455), 509–63.

Kopec, Matthew (2015). A Counterexample to the Uniqueness Thesis. *Philosophia*, *43*, 403–9.

Kopec, Matthew and Michael Titelbaum (2016). The Uniqueness Thesis. *Philosophy Compass*, *11*(4), 189–200.

Lasonen-Aarnio, Maria (2020). Enkrasia or Evidentialism? Learning to Love Mismatch. *Philosophical Studies*, *177*, 597–632.

Leonard, Nick (2020). Epistemic Dilemmas and Rational Indeterminacy. *Philosophical Studies*, *177*, 573–96.

Littlejohn, Clayton (2018). Stop Making Sense? On a Puzzle about Rationality. *Philosophy and Phenomenological Research*, *96*(2), 257–72.

Lord, Errol (2018). *The Importance of Being Rational*. Oxford University Press.

Makinson, David (2005). *Bridges from Classical to Nonmonotonic Logic*. King's College Publications.

McCain, Kevin (2014). *Evidentialism and Epistemic Justification*. Routledge.

Meacham, Christopher (2014). Impermissive Bayesianism. *Erkenntnis*, *79*, 1185–1217.

Priest, Graham (2002). Rational Dilemmas. *Analysis*, *62*(1), 11–16.

Pryor, James (2018). The Merits of Incoherence. *Analytic Philosophy*, *59*(1), 112–41.

Raleigh, Thomas (2017). Another Argument against Uniqueness. *Philosophical Quarterly*, *67*(267), 327–46.

Reiter, Raymond (1980). A Logic for Default Reasoning. *Artificial Intelligece*, *13*, 81–132.

Rescher, Nicholas and Ruth Manor (1970). On Inference from Inconsistent Premises. *Theory and Decision*, *1*(2), 179–217.

Sartre, Jean-Paul (1996). *L'existentialisme est un Humanisme*. Folio Essais. (Original work published 1946)

Schechter, Joshua (2013). Rational Self-Doubt and the Limits of Closure. *Philosophical Studies*, *163*(2), 429–52.

Schoenfield, Miriam (2014). Permission to Believe: Why Permissivism Is True and What It Tells Us about Irrelevant Influences on Belief. *Noûs*, *48*(2), 193–218.

Schroeder, Mark (2018). The Unity of Reasons. In Daniel Star (Ed.), *The Oxford Handbook of Reasons and Normativity* (46–66). Oxford University Press.

Searle, John (1980). Prima Facie Obligations. In Zak van Straaten (Ed.), *Philosophical Subjects: Essays Presented to P. F. Strawson* (238–59). Oxford University Press.

Silva, Paul, Jr. (2017). How Doxastic Justification Helps Us Solve the Puzzle of Misleading Higher-Order Evidence. *Pacific Philosophical Quarterly*, *98*(S1), 308–28.

Simpson, Robert (2017). Permissivism and the Arbitrariness Objection. *Episteme*, *14*(4), 519–38.

Skipper, Mattias (2019). Higher-Order Defeat and the Impossibility of Self-Misleading Evidence. In Mattias Skipper and Asbjørn Steglich-Petersen (Eds.), *Higher-Order Evidence: New Essays* (189–208). Oxford University Press.

Sliwa, Paulina and Sophie Horowitz (2015). Respecting all the Evidence. *Philosophical Studies*, *172*, 2835–58.

Stapleford, Scott (2019). Intraspecies Impermissivism. *Episteme*, *16*(3), 340–56.

Tal, Eyal (2020). Is Higher-Order Evidence Evidence? *Philosophical Studies*. Advance online publication. https://doi.org/10.1007/s11098-020-01574-0

Thorstad, David (2019). Permissive Metaepistemology. *Mind*, *128*(511), 907–26.

Titelbaum, Michael (2015). Rationality's Fixed Point (or: In Defense of Right Reason). *Oxford Studies in Epistemology*, *5*, 253–94.

van Fraassen, Bas (1973). Values and the Heart's Command. *The Journal of Philosophy*, *70*(1), 5–19.

Weatherson, Brian (2010). Do Judgments Screen Evidence? Unpublished manuscript.

Weisberg, Jonathan (2020). Could've Thought Otherwise. *Philosopher's Imprint*, *20*(12), 1–24.

White, Roger (2005). Epistemic Permissiveness. *Philosophical Perspectives*, *19*(1), 445–59.

White, Roger (2007). Epistemic Subjectivism. *Episteme*, *4*(1), 115–29.

White, Roger (2014). Evidence Cannot Be Permissive. In Matthias Steup, John Turri, and Ernest Sosa (Eds.), *Contemporary Debates in Epistemology* (312–23). John Wiley and Sons.

Worsnip, Alex (2018). The Conflict of Evidence and Coherence. *Philosophy and Phenomenological Research*, *96*(1), 3–44.

Worsnip, Alex (2019). Can Your Total Evidence Mislead about Itself? In Mattias Skipper and Asbjørn Steglich-Petersen (Eds.), *Higher-Order Evidence: New Essays* (298–316). Oxford University Press.