Moral Principles: Hedged, Contributory, Mixed

Aleks Knoks [1]

University of Zurich
Philosophisches Seminar, Zürichbergstrasse 43
CH-8044 Zürich, Switzerland

Abstract

It's natural to think that the principles expressed by the statements "Promises ought to be kept" and "We ought to help those in need" are defeasible. But how are we to make sense of this defeasibility? On one proposal, moral principles have *hedges* or built-in unless clauses specifying the conditions under which the principle doesn't apply. On another, such principles are *contributory* and, thus, do not specify which actions ought to be carried out, but only what counts in favor or against them. Drawing on a defeasible logic framework, this paper sets up three models: one model for each proposal, as well as a third model capturing a mixed view on principles that combines them. It then explores the structural connections between the three models and establishes some equivalence results, suggesting that the seemingly different views captured by the models are closer than standardly thought.

Keywords: moral principles, defeasibility, reasons, defeasible logic

1 Introduction

Consider the following moral principles:

(Promise-keeping) If an agent has promised to X, then she ought to X.

(Beneficence) If an agent can help someone in need by X-ing, then she ought to X. [2]

It's natural to think that these principles—or something in their vicinity—are getting at important truths, and that they should have some role to play in our accounts of morality. However, anyone who accepts them—in fact, anyone who thinks that there are *some* principles like them—faces a challenge: They must explain what happens in cases where such principles come into conflict, such as:

Drowning Child. You have promised a friend to meet her for dinner. En route to the restaurant, you come across a child who has fallen in a shallow pond. The child is crying in distress, and all the evidence suggests that she is going

[1] aleks.knoks@uni.lu

[2] Both principles are mentioned in W. D. Ross' list of "basic duties"—see [25, pp. 21–2].

to drown, unless you help her. However, if you rescue the child, you will get your clothes wet and muddy, and won't make it to the dinner

Applying the two rules to this case leads to the conclusion that you ought to both have dinner with your friend and save the child. Taking this at face value means classifying the scenario as a tragic dilemma, or a situation where the agent can't do what she ought to no matter how she acts.[3] And this in spite of the strong intuition that the right thing for you to do is to save the child.

There seem to be two plausible things to say about cases involving conflicts between Promise-keeping and Beneficence, and other principles like them. Both imply that these principles are defeasible.[4] First, one could hold that moral principles are *contributory* or that they do not (by themselves) specify which actions ought to be carried out, but only what counts in favor or against them. Applying this view to the scenario, one could say that, even though there's a genuine conflict between the two principles, it's not a dilemma, because Beneficence outweighs or overrides Promise-keeping. So what you ought to do all-things-considered is save the child. Alternatively, one could hold that moral principles have implicit *hedges* or unless clauses that specify the circumstances under which they don't apply. Applying this view to the Drowning Child, one could say that the conflict between Beneficence and Promise-keeping is only apparent because, say, Promise-keeping doesn't apply when helping those in need means saving their lives. So, on this view too, what you ought to do all-things-considered is save the child.[5]

We will state these views on principles more precisely in later sections. For now simply note that they are naturally thought of and usually presented as distinct, even rival.[6] My aim in this paper is to contribute to a systematic theory of moral principles by exploring the relations between these two views and a mixed view combining them. I will devise a formal model of each, drawing on a simple defeasible logic framework, and establish some results, suggesting that the views modeled are closer than one may think.

The remainder of this paper is structured as follows. Section 2 sets up the stage by formally stating the problem that conflicts between principles give rise to. Sections 3–5 present the models of, respectively, the view on which rules are contributory, the view on which they are hedged, and the mixed view. Section 6 establishes the main results of this paper: The model of the view on which rules are contributory turns out to be equivalent to a fragment of the model on which rules are hedged, and the latter turns out to be equivalent to a restricted fragment of the model of the mixed view. Section 7 concludes and discusses

---

[3] This is how dilemmas are usually characterized—see, e.g., [5].

[4] I'm using the term *defeasible* loosely here, meaning that a principle can engender an ought in a situation and then fail to engender it in a slightly different situation.

[5] For views on which moral principles have hedges see [6], [27], see also [1], [3], and [31] for kindred views in epistemology. For a classical defense of contributory moral principles see the work of W. D. Ross [25]. Views that are naturally thought of as ones on which principles are contributory, but can also have hedges include [11], [15], and [32].

[6] See, for instance, [2, Sec. 1.2] in ethics and [1] in epistemology.

some directions for future research.

2 Preliminaries and the naive view

As our background, we assume the language of propositional logic with the standard connectives. The turnstile $\vdash$ will stand for classical logical consequence. To avoid unnecessary clutter when formalizing particular cases, we assume that our background language allows for materially inconsistent atomic formulas that can't jointly be true, representing such statements as "It's Friday" and "It's Monday." All the formulas we'll encounter should be thought of as relativized to an agent in a situation. Also, we make use of the customary deontic operator. A formula of the form $\bigcirc X$ should be read as saying that it ought to be the case that X, or that morality requires that X. Also, the *ought* here is all-things-considered and not *pro tanto*.

As a first stab, we represent moral principles as (vertically ordered) pairs of formulas of the form $\dfrac{X}{\bigcirc Y}$, where X and Y are formulas of propositional logic.[7] The first expresses a descriptive feature of the situation, the second a normative one, an ought. Now think back to the Drowning Child. Letting p and d stand for the propositions, respectively, that you've made a promise to your friend to dine with her, and that you dine with her, we could represent the relevant instance of Promise-keeping as $\dfrac{p}{\bigcirc d}$. Think of this pair of formulas by analogy with (indefeasible) inference rules of logical systems. The idea is that it lets you infer $\bigcirc d$ whenever p obtains. From now on, then, we'll often refer to our formal representations of principles as *rules*. We denote them with the Greek letter δ, with subscripts, and also introduce two functions $Premise[\cdot]$ and $Conclusion[\cdot]$ for selecting their elements: Where δ stands for $\dfrac{X}{Y}$, the expression $Premise[\delta]$ will stand for the proposition X and $Conclusion[\delta]$ for Y. Also, where $\mathcal{D}$ is a set of rules, we let $Conclusion[\mathcal{D}]$ stand for $\{Conclusion[\delta] : \ \delta \in \mathcal{D}\}$.

We represent particular cases with the help of the notion of a context.

Definition 2.1 [Contexts] A *context* c is a structure of the form $\langle \mathcal{W}, \mathcal{D} \rangle$, where $\mathcal{W}$ is a set of propositional formulas—capturing the normatively-relevant descriptive features of the scenario and called the *hard information*—and $\mathcal{D}$ is a set of rules of the form $\dfrac{X}{\bigcirc Y}$.

To see the notion in play, let's use it to formalize our running example: Let p and d be as before, and let c and s stand for the propositions that a child is drowning, and that you save the child. (We assume that d and s are materially inconsistent.) The scenario can then be captured in the context $c_1 = \langle \mathcal{W}, \mathcal{D} \rangle$ where $\mathcal{W}$ is the set $\{c, p\}$ and $\mathcal{D}$ contains the familiar rule $\delta_1 = \dfrac{p}{\bigcirc d}$, as well as the relevant instance of Beneficence, namely, the rule $\delta_2 = \dfrac{c}{\bigcirc s}$, which says

[7] Principles are often formalized as pairs of formulas—see, e.g., [7,8,13,14].

that you ought to save the child in case she is drowning.

Why does it seem like Promise-keeping and Beneficence reveal important truths about morality? Well, one possible answer is that their instances—together with instances of all other principles—are what link the descriptive features of situations to the normative ones, or, roughly, what happens to what ought to happen.[8] It seems natural to explicate this intuitive idea in the present framework as follows: There's a context standing for every situation, and the (infinite) set of all contexts shares a common set of rules $\mathcal{D}$, containing every instance of Promise-keeping, Beneficence, and other schemas capturing moral principles. Now, one might hope that the logic governing the interaction between these principles is just the good old classical logic.[9] We can capture this view—the naive view alluded to in this section's title—in our framework in two steps. The first is to introduce the notion of triggered rules:

Definition 2.2 [Triggered rules] Let $c = \langle \mathcal{W}, \mathcal{D} \rangle$ be a context. The rules from $\mathcal{D}$ that are *triggered* in c are those that belong to the set $Triggered(c) = \{\delta \in \mathcal{D} : \mathcal{W} \vdash Premise[\delta]\}$.

And the second is to specify which ought formulas follow from a context:

Definition 2.3 [Consequence, first pass] Let $c = \langle \mathcal{W}, \mathcal{D} \rangle$ be a context. Then $\bigcirc X$ follows from c just in case $\bigcirc X$ follow from $Conclusion[Triggered(c)]$ by *standard deontic logic*.[10]

Now let's apply these definitions to the context capturing the Drowning Child. It's easy to see that both δ_1 and δ_2 are triggered in c_1, and that both $\bigcirc d$ and $\bigcirc s$ follow from it. And in light of the fact that d and s are materially inconsistent, a formula of the form $\bigcirc X$ follows from c_1 for any X whatsoever. This, of course, is no good. So we have to abandon the naive view and change either the way we think about moral principles, or the logic governing their interaction, or both.

3 Contributory principles

According to one prominent view, there is no real problem here because the moral principles in play in the Drowning Child and other scenarios like it are contributory: They do not—not by themselves anyway—specify which actions ought to be carried out, but only what speaks in favor or against carrying them out.[11] This section sets up a simple model of this view, drawing on the work of Horty [7,8].[12]

We represent contributory principles as default rules of the form $\frac{X}{Y}$. Intuitively, a rule of this form can be thought of as saying that X exerts some sort

[8] This idea is widely shared among ethicists.

[9] Again, the idea that moral principles (whatever their shape) are governed by classical logic is widespread among ethicists—see, e.g., the remarks in [6].

[10] For a nice presentation of standard deontic logic, see [16, Sec. 2].

[11] See footnote 5 for references.

[12] The model presented here is a fragment of the model set up in [8].

of normative pressure that Y obtains. Functionally, it will let us infer Y from X by default.

Contributory rules are usually associated with *relative weights*, and it's standard to represent these weights formally by means of a priority relation. So where δ and δ' are (contributory) rules, a statement of the form $\delta \leq \delta'$ will mean that δ' has at least as much weight as δ, or that δ' is at least as strong as δ. Following standard practice, we assume that the relation $\leq$ is reflexive and transitive, as well as write $\delta < \delta'$ when $\delta \leq \delta'$ and not $\delta' \leq \delta$.

The next natural step would be to adapt the notion of a context to the idea that rules are contributory. Before taking it, however, we need to introduce the notion of *contrary rules*. Our definition will draw on the concept of minimal inconsistency:

Definition 3.1 [Minimally inconsistent subsets of contrary rules] Let $\mathcal{D}$ be a set of contributory rules and $\mathcal{D}' \subseteq \mathcal{D}$. Then $\mathcal{D}'$ is a minimally inconsistent subset of $\mathcal{D}$ just in case $Conclusion[\mathcal{D}'] \vdash \bot$ and there's no $\mathcal{D}'' \subset \mathcal{D}'$ with $Conclusion[\mathcal{D}''] \vdash \bot$.

Definition 3.2 [Contrary contributory rules] Let $\mathcal{D}$ be a set of contributory rules and δ, δ' two rules from $\mathcal{D}$. Then δ and δ' are *contrary* against the background of $\mathcal{D}$, written as $contrary_{\mathcal{D}}(\delta, \delta')$, if and only if there's a minimally inconsistent subset $\mathcal{D}'$ of $\mathcal{D}$ with $\delta, \delta' \in \mathcal{D}'$.

Notice that Definitions 3.1 and 3.2 capture and generalize the intuitive idea that *two* rules are contrary when their conclusions are inconsistent. The recourse to minimally inconsistent subsets is needed to account for cases where there's a set of rules the conclusions of which are pairwise consistent, but jointly inconsistent. As an example, consider the rules $\frac{\top}{a}$, $\frac{\top}{b}$, and $\frac{\top}{\neg(a\&b)}$. Were we to say that two rules are contrary just in case their conclusions are inconsistent, the inconsistency of these three rules would slip through. And we don't want that to happen.

Now we can adjust our notion of a context, and we call contexts that represent moral principles as contributory rules *weighted*:

Definition 3.3 [Weighted contexts] A weighted context c is a structure of the form $\langle \mathcal{W}, \mathcal{D}, \leq \rangle$, where $\mathcal{W}$ is a set of propositional formulas, $\mathcal{D}$ is a set of contributory rules, and $\leq$ is a reflexive and transitive relation (a preorder) on $\mathcal{D}$. We assume that weighted contexts are subject to the following constraint:

No Dilemmas: For any $\delta, \delta' \in \mathcal{D}$ with $contrary_{\mathcal{D}}(\delta, \delta')$, either $\delta \leq \delta'$, or $\delta' \leq \delta$.

The Drowning Child can, then, be captured by the weighted context $c_2 = \langle \mathcal{W}, \mathcal{D}, \leq \rangle$ where $\mathcal{W}$ is the set $\{p, c\}$, where $\mathcal{D}$ contains the rules $\delta_3 = \frac{p}{d}$ and $\delta_4 = \frac{c}{s}$, and where $\delta_3 < \delta_4$. Intuitively, δ_3 says that your having made a promise to dine with your friend speaks in favor of you dining with her, while δ_4 says that the child's drowning (and needing your help) speaks in favor of you saving the child. The relation between the rules, $\delta_3 < \delta_4$, expresses the

idea that the latter has strictly more weight, or is strictly stronger, than the former. Notice that this doesn't mean that Beneficence always has more weight than Promise-keeping—there will be many other contexts in which instances of the latter take precedence over the instances of the former. The No Dilemmas constraint captures the following assumption: Moving to contributory principles in response to the problem that conflicts between moral principles give rise to suffices to show that such conflicts aren't tragic dilemmas. The assumption seems to me to be fully justified in the context of this paper.

Now we specify which ought statements follow from a context, relying on three simple definitions.

Definition 3.4 [Outweighed rules] Let $c = \langle \mathcal{W}, \mathcal{D}, \leq \rangle$ be a weighted context. The rules from $\mathcal{D}$ that are *outweighed* in c are those that belong to the set

$$Outweighed(c) \quad = \quad \{\delta \in \mathcal{D} : \quad \text{there is some } \delta' \in Triggered(c) \text{ such that } (1)\ \delta \leq \delta' \text{ and } (2)\ contrary_{\mathcal{D}}(\delta, \delta')\}.$$

So a rule is outweighed in a context just in case there's another rule that's triggered, contrary to it, and at least as strong. Notice that this formal notion doesn't match the intuitive sense of *outweighed* perfectly, since it qualifies a rule δ as outweighed if there's another rule that's strictly stronger than δ, as well as if there's another rule that's only as strong as δ. In the latter case, it'd be more fitting to say that δ is counterbalanced. If there was a word in English covering the senses of both *outweighed* and *counterbalanced*, it'd be perfect for our purposes. But given that there isn't one, we work with what we have.

Our next definition combines the notions of triggered and outweighed rules:

Definition 3.5 [Binding rules] Let $c = \langle \mathcal{W}, \mathcal{D}, \leq \rangle$ be a weighted context. The rules from $\mathcal{D}$ that are *binding* in c are those that belong to the set

$$Binding(c) \quad = \quad \{\delta \in \mathcal{D} : \quad \delta \in Triggered(c) \text{ and } \delta \notin Outweighed(c)\}.$$

So a rule is binding just in case it is triggered and *not* outweighed. Such binding rules are just what will give us the ought statements:

Definition 3.6 [Consequence, weighted] Let $c = \langle \mathcal{W}, \mathcal{D}, \leq \rangle$ be a weighted context. Then $\bigcirc X$ follows from c just in case $Conclusion[Binding(c)] \vdash X$.

Returning to the Drowning Child scenario, it's easy to see that, on Definition 3.6, $\bigcirc s$ follows from c_2. Both δ_3 and δ_4 get triggered, but only the latter qualifies as binding. Given that δ_4 is strictly stronger than δ_3, $\delta_3 < \delta_4$, and that the two are contrary, $contrary_{\mathcal{D}}(\delta_3, \delta_4)$, the rule δ_3 comes out outweighed in c_2. Since $Binding(c_2) = \{\delta_4\}$, we have it that $Conclusion[Binding(c_2)] = \{s\}$ and, therefore, that $\bigcirc s$ follows from c_2. Thus, we get the intuitive result that you ought to save the child.

It's worth noting that our model of the view on which moral principles are contributory gives rise to consistent oughts: [13]

[13] Many thanks to an anonymous reviewer for pressing me to prove this fact. Among other

Fact 3.7 *Let $c = \langle \mathcal{W}, \mathcal{D}, \leq \rangle$ be a weighted context. Then $\bigcirc\bot$ follows from c only if there's a rule δ in $\mathcal{D}$ with $Conclusion[\delta] = \bot$.*

Proof. Suppose that $\bigcirc\bot$ follows from the context c. This means that $Conclusion[Binding(c)] \vdash \bot$. Now let's zoom in on some minimally inconsistent subset $\mathcal{D}'$ of $Binding(c)$. (It is guaranteed to exist.) If $\mathcal{D}'$ is a singleton, we're done. So suppose that it isn't. Since $\mathcal{D}'$ is a minimally inconsistent subset of $Binding(c)$ and $Binding(c) \subseteq \mathcal{D}$, Definition 3.2 entails that, for any $\delta, \delta' \in \mathcal{D}'$, we have $contrary_{\mathcal{D}}(\delta, \delta')$. Now zoom in on two such rules δ, δ'. Since $\delta, \delta' \in Binding(c)$, we have $\delta, \delta' \in Triggered(c)$ and $\delta, \delta' \notin Outweighed(c)$. In light of the No Dilemmas constraint, we can be sure that either $\delta \leq \delta'$, or $\delta' \leq \delta$. Without loss of generality, we assume that $\delta \leq \delta'$. Now notice that we have $\delta' \in Triggered(c)$, $\delta \leq \delta'$, and $contrary_{\mathcal{D}}(\delta, \delta')$. This entails $\delta \in Outweighed(c)$, giving us a contradiction. □

Even though the model we just set up is very simple, it's expressive enough to capture some of the *reasons talk* that's so pervasive in contemporary ethical and meta-normative debates.[14] More specifically, the notion of a normative reason can be specified as follows: X is a normative reason for Y in the weighted context c if and only if there's a rule δ of the form $\frac{X}{Y}$ that's triggered in c. When there is such a rule, we say that X's being a reason for Y depends on it.[15] We say that X is outweighed as a reason for Y in c just in case the rule δ that X's being a reason for Y is triggered, but not binding in c. And we say that X is outweighed—or, more precisely, outweighed or counterbalanced—as a reason for Y *by the consideration* Z in c just in case the rule δ that X's being a reason for Y depends on is outweighed by some contrary rule δ' that has Z as a premise.

4 Hedged principles

On an alternative view, situations where moral principles seem to support conflicting recommendations are cases of only apparent conflicts, because principles have built-in hedges guaranteeing that at most one principle applies in any such situation.[16] This section sets up a model of this view.

In Section 2, we expressed principles as rules of the form $\frac{X}{\bigcirc Y}$. Now we do so using the slightly more complex $\frac{X : \{\neg Z_1, \neg Z_2, \ldots\}}{\bigcirc Y}$. The new element $\{\neg Z_1, \neg Z_2, \ldots\}$ is a set of negated propositional formulas standing for the rule's hedge—note that we will often abbreviate it as $\mathcal{Z}$. A hedged principle of this form should be read as, "If X obtains, then it ought to be the case that Y, unless either Z_1 obtains, or Z_2 obtains, or ...." Alternatively, it can be read as, "If X

---

things, this saved me from an embarrassing mistake.

[14] See, e.g., [20,23,26,28,29,30], and [7,8,18] for modeling reasons talk using defeasible logics.

[15] Compare to [8, Section 2.1].

[16] See footnote 5 for references.

obtains, and not-Z_1, not-Z_2, ..., then it ought to be the case that Y." [17] We retain the functions for selecting rule premises and conclusions. Additionally, we introduce a function for selecting a given rule's hedge: If δ is a rule of the form $\frac{X : \mathcal{Z}}{Y}$, let $Hedge[\delta] = \mathcal{Z}$, and if δ is of the form $\frac{X}{Y}$, let $Hedge[\delta] = \emptyset$.

Like we did in the previous section, here too we make use of the idea of contrary rules: [18]

Definition 4.1 [Minimally inconsistent subsets of hedged rules] Let $\mathcal{D}$ be a set of hedged rules and $\mathcal{D}' \subseteq \mathcal{D}$. Then $\mathcal{D}'$ is a *minimally inconsistent subset* of $\mathcal{D}$ just in case $\bigcirc\bot$ follows from $Conclusion[\mathcal{D}']$ in standard deontic logic and there's no $\mathcal{D}'' \subset \mathcal{D}'$ such that $\bigcirc\bot$ follows from $Conclusion[\mathcal{D}'']$ in standard deontic logic.

Definition 4.2 [Contrary hedged rules] Let $\mathcal{D}$ be a set of hedged rules and δ, δ' two rules from $\mathcal{D}$. Then δ and δ' are *contrary* against the background of $\mathcal{D}$, written as $contrary_{\mathcal{D}}(\delta, \delta')$, if and only if there's a minimally inconsistent subset $\mathcal{D}'$ of $\mathcal{D}$ with $\delta, \delta' \in \mathcal{D}'$.

With the notion of contrary rules in hand, we can adjust the definition of a context from Section 2 to the idea that rules expressing principles can have hedges:

Definition 4.3 [Hedged contexts] A *hedged context* is a structure of the form $\langle \mathcal{W}, \mathcal{D} \rangle$, where $\mathcal{W}$ is a set of propositional formulas and $\mathcal{D}$ is a set of rules, possibly hedged. We assume throughout that hedged contexts are subject to two constraints:

No Dilemmas: For any $\delta, \delta' \in \mathcal{D}$ with $contrary_{\mathcal{D}}(\delta, \delta')$, either $\neg Premise[\delta] \in Hedge[\delta']$, or $\neg Premise[\delta'] \in Hedge[\delta]$.

No Deviant Pairs of Rules: For any $\delta, \delta' \in \mathcal{D}$, in case $Premise[\delta] = Premise[\delta']$ and $Conclusion[\delta] = Conclusion[\delta']$, then $\delta = \delta'$.

As an illustration, the Drowning Child can be represented as the hedged context $c_3 = \langle \mathcal{W}, \mathcal{D} \rangle$ where $\mathcal{W} = \{p, c\}$ and where $\mathcal{D}$ is comprised of the rules $\delta_5 = \frac{p : \neg c}{\bigcirc d}$ and $\delta_6 = \frac{c}{\bigcirc s}$. The first rule says that you ought to dine with your friend if you've promised to dine with her, unless a child needs help; the second says that you ought to save the child, if the child needs help. [19] Now for the two constraints: No Dilemmas amounts to, again, the assumption that

[17] Compare to Reiter's default rules [24].

[18] You may wonder if I shouldn't use indices to keep track of the difference between Definitions 3.2 and 4.2. My reason for not using indices here and elsewhere is to avoid notational clutter. I think that the context always makes it clear whether we're discussing the view on which principles are hedged or the one on which they are contributory, helping disambiguate between the two notions of contrary rules. Parallel considerations apply to the notions of outweighed rules, Definitions 3.4 and 5.4, and undercut rules, Definitions 5.2 and 5.3.

[19] It's natural to wonder if the hedge of the rule δ_5 shouldn't also list the other circumstances in which this rule wouldn't apply—and similarly for δ_6. While I do think that this is a natural consequence of the view, nothing hinges on us working with simplified examples here. For more on the worry that hedged rules might end up being incredibly complex see, e.g., [1].

appealing to hedges succeeds as a response to the problem of conflicting moral principles. No Deviant Pairs of Rules, in turn, rules out pairs of principles that apply in the same circumstances and prescribe the same course of action, but have different hedges: It seems natural to think that any pair of such principles is deviant and should be substituted by one principle with a single hedge.

Our next two definitions determine which ought statements follow from hedged contexts.

Definition 4.4 [Admissible rules] Let $c = \langle \mathcal{W}, \mathcal{D} \rangle$ be a hedged context. The rules from $\mathcal{D}$ that are *admissible* in c are those that belong to the set

$$Admissible(c) \quad = \quad \{\delta \in \mathcal{D} : \quad \delta \in Triggered(c) \text{ and, for no } \neg Z \in Hedge[\delta], \mathcal{W} \vdash Z\}.$$

Definition 4.5 [Consequence, hedged] Let $c = \langle \mathcal{W}, \mathcal{D} \rangle$ be a hedged context. Then $\bigcirc X$ follows from c just in case $\bigcirc X$ follows from $Conclusion[Admissible(c)]$ by standard deontic logic.

Notice that δ_6 does, while δ_5 does not qualify as admissible in c_3. The latter rule gets triggered, but the fact that $\neg c \in Hedge[\delta_5]$ and $\mathcal{W} \vdash c$ precludes it from being added to $Admissible(c_3)$. And given that $Conclusion[Admissible(c_3)] = \{\bigcirc s\}$, we get the intuitive result that $\bigcirc s$ does, while $\bigcirc d$ does not follow from c_3: What you ought to do is save the child.

Let's also note that this model too gives rise to consistent oughts:

Fact 4.6 *Let $c = \langle \mathcal{W}, \mathcal{D} \rangle$ be a hedged context. Then $\bigcirc\bot$ follows from c only if there's a rule δ in $\mathcal{D}$ with $Conclusion[\delta] = \bigcirc\bot$.*

Proof. Suppose that $\bigcirc\bot$ follows from the hedged context c. This means that $\bigcirc\bot$ follow from $Conclusion[Admissible(c)]$ in standard deontic logic. Now consider some minimally inconsistent subset $\mathcal{D}'$ of $Admissible(c)$. In case $\mathcal{D}'$ is a singleton, we have a $\delta \in \mathcal{D}' \subseteq Admissible(c) \subseteq \mathcal{D}$ with $Conclusion[\delta] = \bigcirc\bot$, which would establish the fact. So suppose, toward a contradiction, that $\mathcal{D}'$ is not a singleton set. Given that $\mathcal{D}'$ is a minimally inconsistent subset of $Admissible(c)$ and $Admissible(c) \subseteq \mathcal{D}$, Definition 4.2 tells us that, for any $\delta, \delta' \in \mathcal{D}'$, we have $contrary_{\mathcal{D}}(\delta, \delta')$. Now consider two such rules δ, δ'. Since $\delta, \delta' \in Admissible(c)$, we know that $\delta, \delta' \in Triggered(c)$, that there's no $\neg Z \in Hedge[\delta]$ with $\mathcal{W} \vdash Z$, and that there's no $\neg Z \in Hedge[\delta']$ with $\mathcal{W} \vdash Z$. Given the No Dilemmas constraint, we can be sure that either $\neg Premise[\delta] \in Hedge[\delta']$ or $\neg Premise[\delta'] \in Hedge[\delta]$ holds true. Without loss of generality, we suppose it is the former. Since $\delta \in Triggered(c)$, we have it that $\mathcal{W} \vdash Premise[\delta]$. Then, however, we have $\neg Premise[\delta] \in Hedge[\delta']$ and $\mathcal{W} \vdash Premise[\delta]$, which contradicts the claim that there's no $\neg Z \in Hedge[\delta']$ with $\mathcal{W} \vdash Z$. □

Unlike the view on which moral principles are contributory, the view on which they are hedged is usually taken to be less congenial to reasons talk. [20]

[20] See, e.g., Dancy's [2, pp. 22–9] where it's argued that the view has no hope to account for the phenomena of residual reasons and reason aggregation. If the results established below

However, nothing stands in the way of using the model we just set up to talk and reason about reasons. Thus, we say that X is a normative reason for Y in the hedged context c if and only if there's a rule δ of the form $\frac{X : \mathcal{Z}}{\bigcirc Y}$ that's triggered in c. When there is such a rule δ, we say that X's being a reason for Y depend on δ. We say that X is *defeated* as a reason for Y in c just in case the rule δ that X's being a reason for Y in c depends on is triggered, but not admissible in c. And we say that X is defeated as a reason for Y *by the consideration* Z in $c = \langle \mathcal{W}, \mathcal{D} \rangle$, or, alternatively, that Z is a defeater of X as a reason for Y in c, if and only if the rule δ that X's being a reason for Y depends on is triggered in c, has $\neg Z$ in its hedge, and $\mathcal{W} \vdash Z$. So our some reason talk can be captured in our model of the hedged-principles view too.

Now we have two models of two seemingly very different views on moral principles and conflicts between them. But it turns out to be possible to establish a type of equivalence result between these models. More concretely, there's a many-one equivalence between weighted contexts and a special class of hedged contexts I call *simple*. The special character of these simple contexts has to do with the shape of the hedges of their rules:

Definition 4.7 [Simple hedged contexts] Let $c = \langle \mathcal{W}, \mathcal{D} \rangle$ be a hedged context. We say that c is *simple* just in case, for any rule δ in $\mathcal{D}$, the hedge of δ is the set $\{\neg Premise[\delta'] : \ \delta' \in \mathcal{D}'\}$ where $\mathcal{D}' \subseteq \{\delta' \in \mathcal{D} : contrary_{\mathcal{D}}(\delta, \delta')\}$.

On the intuitive level, this simplicity condition can be thought of as a restriction on what rule hedges can do: They can refer *only* to the premises of contrary rules, and, thus, all they can do is resolve conflicts between rules supporting conflicting recommendations.

And I call a weighted context c and a hedged one c' *equivalent* if and only if X is a reason for Y in c just in case X is a reason for Y in c'; X is outweighed as a reason for Y by Z in c just in case X is defeated as a reason for Y by Z in c'; and $\bigcirc X$ follows from c just in case $\bigcirc X$ follows from c'. So the connection between weighted and simple hedged contexts is very close. We'll state the result more precisely in Section 6, along with some other equivalences. Now we turn to a combination of the two views on principles we have discussed.

5 Mixed principles

There's a well-known objection to the views on which principles are contributory: They entail that the moral valences of features are fixed, or, roughly, that if a feature constitutes a reason for some type of action in one situation, then it's *always* a reason for that type of action. This, however, appears to be too strong. Suppose that you've made a promise to meet someone for dinner, but they extracted this promise from you by threat of force. A proponent of the contributory view is forced to think of this case as one where Promise-keeping yields to a weightier principle saying something along the lines of, "If a promise to X was extracted by threat of force, then that speaks against X-ing." And

---

are correct, Dancy's arguments can't work—see [9, Ch. 2] for more on this.

this implies that there's some reason for you to keep the promise. But, intuitively, the moral force that normally comes with promise-making is invalidated or voided in this situation, and there's absolutely no reason for you to keep the promise. [21]

Some philosophers take this objection to be fatal to the contributory view, and it does seem that this view can't handle cases involving what we could call *principle-invalidating conditions* adequately. However, there's a view that's naturally thought of as its extension that can, namely, the *mixed view* on which principles are contributory, but can also have hedges. [22] In the remainder of this section, we set up a model of this view.

We use rules of the form $\frac{X : \{\neg Z_1, \ldots, \neg Z_n\}}{Y}$ to express mixed principles. A rule having this form should be read as, "If X obtains, then there's normative pressure that Y obtains, unless either Z_1, or Z_2, or, ..., Z_n obtain." Adapting the by now familiar notion of contrary rules to the mixed setting is a trivial exercise: All we need to do is substitute *mixed rules* for *contributory rules* in Definitions 3.1 and 3.2 from Section 3. We won't do this explicitly, proceeding directly to the next step, that is, to defining mixed contexts:

Definition 5.1 [Mixed contexts] A *mixed context* c is a structure of the form $\langle \mathcal{W}, \mathcal{D}, \leq \rangle$, where $\mathcal{W}$ is a set of propositional formulas, $\mathcal{D}$ is a set of mixed rules, and $\leq$ is a preorder on $\mathcal{D}$. Yet again, we assume that mixed contexts are subject to two familiar constraints:

No Dilemmas: For any $\delta, \delta' \in \mathcal{D}$ with $contrary_{\mathcal{D}}(\delta, \delta')$, either $\delta \leq \delta'$, or $\delta' \leq \delta$.

No Deviant Pairs of Rules: For any $\delta, \delta' \in \mathcal{D}$, in case $Premise[\delta] = Premise[\delta']$ and $Conclusion[\delta] = Conclusion[\delta']$, then $\delta = \delta'$.

To see the notion at work, we formalize the case sketched above. Let p and d stand for the propositions that they stood for before, namely, that you've made a promise to dine with your friend, and that you dine with her. And let t express the proposition that the promise was extracted from you by threat of force. We can then express the case as the context $c_4 = \langle \mathcal{W}, \mathcal{D}, \leq \rangle$, where $\mathcal{W} = \{p, t\}$, where $\mathcal{D}$ is a singleton set containing the rule $\delta_7 = \frac{p : \neg t}{d}$, and where $\leq$ is empty.

Next, we introduce a new notion that operates on the information stored in the hedges of mixed rules: [23]

Definition 5.2 [Undercut rules] Let $c = \langle \mathcal{W}, \mathcal{D}, \leq \rangle$ be a mixed context. The rules from $\mathcal{D}$ that are *undercut* in it are those that belong to the set

[21] Particularists are particularly fond of raising this objection, in both moral and epistemic domains—see, e.g., [1,2,10]. Also, the ideas that duress and deceit invalidates promises, and that a promise's getting invalidated is very different from it being outweighed is widespread among moral philosophers—see, e.g., [4,19].

[22] Again, see footnote 5 for references.

[23] The term *undercut* comes from epistemology where it's customary to distinguish between rebutting and undercutting defeat—see, e.g., [21] and [22].

$$Undercut(c) \quad = \quad \{\delta \in \mathcal{D} : \quad \text{there's a } \neg Z \in Hedge[\delta] \text{ such that } \mathcal{W} \vdash Z\}.$$

This definition might remind you of the way we got to consequences of hedged contexts. However, the notion of undercutting is meant to serve a more specific function here, namely, to get certain rules completely out of play in determining which ought statements follow from a context. What's more, it can be defined in our model of the hedged-principles view too, as follows:

Definition 5.3 [Undercut rules, hedged] Let $c = \langle \mathcal{W}, \mathcal{D} \rangle$ be a hedged context. The rules from $\mathcal{D}$ that are *undercut* in it are those that belong to the set

$$\begin{array}{lcll} Undercut(c) & = & \{\delta \in \mathcal{D} : & \text{there's some } \neg Z \in Hedge[\delta] \text{ with } \mathcal{W} \vdash Z \text{ and} \\ & & & \text{there is no } \delta' \in \mathcal{D} \text{ such that} \\ & & & (1)\ Z = Premise[\delta'] \text{ and } (2)\ contrary_{\mathcal{D}}(\delta, \delta')\}. \end{array}$$

Notice that, on this definition, a hedge rule can be undercut only by a consideration that is not a premise of some rule that's contrary to it.

The addition of undercut rules means that we need to modify the notion of outweighed rules slightly. More specifically, we need to make sure that the rule that's responsible for outweighing is not itself undercut:

Definition 5.4 [Outweighed rules, mixed] Let $c = \langle \mathcal{W}, \mathcal{D}, \leq \rangle$ be a mixed context. The rules from $\mathcal{D}$ that are *outweighed* in it belong to the set

$$\begin{array}{lcll} Outweighed(c) & = & \{\delta \in \mathcal{D} : & \text{there is some } \delta' \in Triggered(c) \text{ such that} \\ & & & (1)\ \delta \leq \delta',\ (2)\ contrary_{\mathcal{D}}(\delta, \delta'), \text{ and} \\ & & & (3)\ \delta' \notin Undercut(c)\}. \end{array}$$

With this, the model is pretty much set up. We only need to combine the above definitions in the analogue of admissible/binding rules for mixed context and define consequence.

Definition 5.5 [Optimal rules] Let $c = \langle \mathcal{W}, \mathcal{D}, \leq \rangle$ be a mixed context. The rules from $\mathcal{D}$ that are *optimal* in it are those that belong to the set

$$\begin{array}{lcll} Optimal(c) & = & \{\delta \in \mathcal{D} : & \delta \in Triggered(c), \\ & & & \delta \notin Outweighed(c), \\ & & & \delta \notin Undercut(c)\}. \end{array}$$

Definition 5.6 [Consequence, mixed] Let $c = \langle \mathcal{W}, \mathcal{D}, \leq \rangle$ be a mixed context. Then $\bigcirc X$ follows from c just in case $Conclusion[Optimal(c)] \vdash X$.

Returning to the context c_4, it's easy to see that the statement $\bigcirc d$ doesn't follow from it. This is as it should be. What's more, we also get the intuitive result that there's no reason for you to meet with the person for dinner—that is, as soon as we specify how reasons are to be identified in the model. Here we reuse the idea from Section 3, with a small twist to it: X is a reason for Y in c if and only if there's a rule of the form $\dfrac{X\ :\ Z}{Y}$ that's triggered in c *and* not undercut. And X is outweighed as a reason for Y *by the consideration* Z in c just in case the rule δ that X's being a reason for Y depends on is outweighed

by a contrary rule δ' that has Z as a premise. With this, p, that you've made a promise to meet your friend for dinner, doesn't qualify as a reason for d, to have dinner with her, in c_4. So the mixed view can account for the case adequately.

We close the section by noting that the oughts that the model of the mixed view gives rise to are also consistent:

Fact 5.7 *Let $c = \langle \mathcal{W}, \mathcal{D}, \leq \rangle$ be a mixed context. Then $\bigcirc\bot$ follows from c only if there's a rule δ in $\mathcal{D}$ with $Conclusion[\delta] = \bot$.*

Proof. Similar to the proof of Fact 3.7. □

6 Relations

Having laid out the models of the three views on principles, we can explore the relations between them. This is the goal of this section.

First off, it should be clear that our model of the view on which principles are contributory corresponds to a fragment of the model of the mixed view: It's easy to see that there's a one-one correspondence between weighted contexts and a special class of mixed contexts, namely, those the hedges of all rules of which are empty.

More surprisingly, it's possible to establish a many-one correspondence between a special class of mixed contexts and hedged contexts. It's natural to think of this correspondence as the main result of this paper.[24] As a first step, we introduce an auxiliary notion that will help us avoid clutter in proofs:

Definition 6.1 [Rule counterparts] Let δ be of the form $\dfrac{X : \mathcal{Z}}{\bigcirc Y}$ and $c = \langle \mathcal{W}, \mathcal{D}, \leq \rangle$ some mixed context. If there's a rule $\delta' \in \mathcal{D}$ with $Premise[\delta'] = X$ and $Conclusion[\delta'] = Y$, we say that δ' is the (mixed) *counterpart* of δ in the context c, written as $counterpart_c(\delta) = \delta'$. Similarly, in case δ is of the form $\dfrac{X : \mathcal{Z}}{Y}$ and the set of rules of some hedged context $c = \langle \mathcal{W}, \mathcal{D} \rangle$ contains some rule δ' with $Premise[\delta'] = X$ and $Conclusion[\delta'] = \bigcirc Y$, we say that δ' is the (hedged) *counterpart* of δ in c, written as $counterpart_c(\delta) = \delta'$.

Now notice that there's a natural procedure for transforming mixed contexts into hedged ones:

Definition 6.2 [Derived hedged contexts] Let $c = \langle \mathcal{W}, \mathcal{D}, \leq \rangle$ be a mixed context. We construct a hedged context c' from it as follows. Let $c' = \langle \mathcal{W}', \mathcal{D}' \rangle$, where $\mathcal{W}' = \mathcal{W}$ and $\mathcal{D}'$ is acquired from $\langle \mathcal{D}, \leq \rangle$ by the following procedure:
For every rule $\delta \in \mathcal{D}$,

A. Let $\mathcal{D}_\delta = \{\delta' \in \mathcal{D} : \delta \leq \delta' \text{ and } contrary_{\mathcal{D}}(\delta, \delta')\}$;

B. set $\mathcal{Z} = \{\neg Premise[\delta'] : \delta' \in \mathcal{D}_\delta\}$;

C. finally, replace $\delta \in \mathcal{D}$ with the rule $\dfrac{Premise[\delta] : Hedge[\delta] \cup \mathcal{Z}}{\bigcirc Conclusion[\delta]}$.

We call the class of mixed contexts that have hedged counterparts *regular*:

[24] I structure the presentation of the result after Lewis' [12, Sec. 4]. Thanks to Paolo Santorio for pointing me to Lewis' paper.

Definition 6.3 [Regular mixed contexts] Let $c = \langle \mathcal{W}, \mathcal{D}, \leq \rangle$ some mixed context. We say that c is *regular* if and only if, for any $\delta, \delta' \in \mathcal{D}$ with $contrary_{\mathcal{D}}(\delta, \delta')$, neither $\neg Premise[\delta'] \in Hedge[\delta]$, nor $\neg Premise[\delta] \in Hedge[\delta']$.

What does this regularity condition do? Well, first, notice that the No Dilemmas constraint on mixed contexts guarantees that any two contrary rules are related by $\leq$, and, thus, that all conflicts between them get resolved. The regularity condition, then, ensures that rule hedges do not interact with this in any way. So it can be thought of as enforcing continuity between the view on which principles are contributory and the mixed one: What accounts for the resolution of conflicts between conflicting principles is their contributory character, not their hedges.

Either way, if a hedged context is acquired from a regular mixed context by Definition 6.2, they are equivalent:

Theorem 6.4 (Mixed-to-Hedged) *Let $c = \langle \mathcal{W}, \mathcal{D}, \leq \rangle$ be some* regular *mixed context and $c' = \langle \mathcal{W}', \mathcal{D}' \rangle$ a (hedged) context derived from c by the procedure specified in Definition 6.2. Then c and c' are* equivalent*, that is:*

1. X is a reason for Y in c if and only if X is a reason for Y in c';

2. X is outweighed as a reason for Y by a consideration Z in c if and only if X is defeated as a reason for Y by Z in c'; and

3. $\bigcirc X$ follows from c if and only if $\bigcirc X$ follows from c'.

Proof. As a first step, we establish the following claim: $\delta \in Undercut(c)$ if and only if $counterpart_{c'}(\delta) \in Undercut(c')$. $\Rightarrow$ Consider an arbitrary $\delta \in Undercut(c)$. There must be some Z with $\mathcal{W} \vdash Z$ and $\neg Z \in Hedge[\delta]$. Since c is regular, there's no $\delta^* \in \mathcal{D}$ with both $contrary_{\mathcal{D}}(\delta, \delta^*)$ and $Z = Premise[\delta^*]$. Now consider $counterpart_{c'}(\delta) = \delta'$. In light of Definition 6.2, there's no $\delta'' \in \mathcal{D}'$ with both $contrary_{\mathcal{D}'}(\delta', \delta'')$ and $\neg Premise[\delta'']$ in $Hedge[\delta']$. Yet $\neg Z \in Hedge[\delta']$ and $\mathcal{W}' \vdash Z$ (since $\mathcal{W}' = \mathcal{W}$). So $\delta' \in Undercut(c')$. $\Leftarrow$ Consider an arbitrary $\delta \in Undercut(c')$. This means that there's some $\neg Z \in Hedge[\delta]$ such that $\mathcal{W}' \vdash Z$ and there's no $\delta^* \in \mathcal{D}'$ with both $Z = Premise[\delta^*]$ and $contrary_{\mathcal{D}'}(\delta, \delta^*)$. Now refocus on $counterpart_c(\delta) = \delta'$. Either $\neg Z \in Hedge[\delta']$ or not. If yes, we're done. So suppose not. Then, in light of Definition 6.2, there has to be some $\delta'' \in \mathcal{D}$ such that $Z = Premise[\delta'']$, $contrary_{\mathcal{D}}(\delta', \delta'')$, and $\delta' \leq \delta''$. Then, however—again, by Definition 6.2—there must be a $\delta^* \in \mathcal{D}'$ with $Z = Premise[\delta^*]$ and $contrary_{\mathcal{D}'}(\delta, \delta^*)$, which gives us a contradiction.

1. The first clause follows directly from the above claim and the definitions of reasons in the two models.

2. Without loss of generality, we prove only the the left-to-right direction of the second clause: Suppose X is outweighed as a reason for Y by Z in c. This entails that there are rules $\delta = \frac{X \,:\, \mathcal{Z}}{Y}$ and $\delta^* = \frac{Z \,:\, \mathcal{Z}'}{W}$ in $\mathcal{D}$ such that $contrary_{\mathcal{D}}(\delta, \delta^*)$, $\delta \leq \delta^*$, δ, $\delta^* \in Triggered(c)$, and δ, $\delta^* \notin Undercut(c)$. By the construction of c', there are rules $\delta', \delta'' \in \mathcal{D}'$ such that $counterpart_c(\delta') = \delta$, $counterpart_c(\delta'') = \delta^*$, and $\neg Premise[\delta''] \in Hedge[\delta']$. Since $Premise[\delta''] = Z$ and $\mathcal{W}' \vdash Z$, the rule δ' gets defeated by Z in c'. Since $\delta \notin Undercut(c)$, we

can be sure that $\delta' \notin Undercut(c')$. Thus, X is defeated as a reason for Y by Z in c'.

3. Instead of proving the third clause directly, we prove a claim from which it quickly follows, namely, $\{\bigcirc X : X \in Conclusion[Optimal(c)]\} = Conclusion[Admissible(c')]$.

$\subseteq$ Consider some formula $\bigcirc X$ in the set on the left hand side. Clearly, $X \in Conclusion[Optimal(c)]$, and so there's a rule δ in $Optimal(c)$ such that $Conclusion[\delta] = X$. Definitions 6.2 and 5.5 (Optimal rules) entail that there's a δ' in $\mathcal{D}'$ such that $counterpart_c(\delta') = \delta$, $\delta' \in Triggered(c')$ and there's no $\delta'' \in Triggered(c')$ with $\neg Premise[\delta''] \in Hedge[\delta']$. The fact that $\delta \in Optimal(c)$ also entails that $\delta \notin Undercut(c)$, whence $\delta' \notin Undercut(c')$. This is actually enough to conclude that $\delta' \in Admissible(c')$. And given that $Conclusion[\delta'] = \bigcirc Conclusion[\delta] = \bigcirc X$, we're done. $\supseteq$ Take an $\bigcirc X \in Conclusion[Admissible(c')]$. Clearly, there's a rule $\delta \in Admissible(c')$ such that $Concluson[\delta] = \bigcirc X$. Now, $\delta \in Triggered(c')$ and for no $\neg Z \in Hedge[\delta]$ do we have $\mathcal{W} \vdash Z$. By Definition 6.2, there must be $\delta' \in \mathcal{D}$ with $counterpart_{c'}(\delta') = \delta$. It's easy to see that $\delta' \in Triggered(c)$, $\delta' \notin Undercut(c)$, and that $\delta' \notin Outweighed(c')$—otherwise, δ wouldn't be admissible. From here, $\delta' \in Optimal(c)$, and so $\bigcirc Conclusion[\delta'] = Conclusion[\delta] = \bigcirc X$ is in the set $\{\bigcirc X : X \in Conclusion[Optimal(c)]\}$. □

Although every regular mixed context is equivalent to some hedged context, two regular mixed contexts can be equivalent to the same hedged context. As an example, take the toy contexts $c_5 = \langle \mathcal{W}, \mathcal{D}, \leq \rangle$ where $\mathcal{W} = \{a, b\}$, $\mathcal{D} = \{ \delta_8 = \dfrac{a : \emptyset}{c}, \ \delta_9 = \dfrac{b : \emptyset}{d} \}$, and $\delta_8 < \delta_9$, and c_6 which is like c_5, except for, in its case, the ordering on rules is $\delta_9 < \delta_8$. Since c and d here are consistent (by assumption), the ordering on rules has no bearing on which ought statements follow from these contexts. So it seems natural to think of it as providing surplus information. It's not difficult to see that applying Definition 6.2 to c_5 and c_6 results in the same hedged contexts, from which the surplus information is absent. (Notice that a similar pair of contexts shows that the correspondence between weighted and simple hedged contexts is many-one.)

For the other direction, we can, again, define a procedure for transforming hedged contexts into mixed ones:

Definition 6.5 [Derived mixed contexts] Let $c = \langle \mathcal{W}, \mathcal{D} \rangle$ be a hedged context. We construct a mixed context c' from it as follows. First, we define an ordering on the rules from $\mathcal{D}$, using their hedges: For any two rules $\delta, \delta' \in \mathcal{D}$, let

$$\delta \preceq \delta' \text{ if and only if } \neg Premise[\delta'] \in Hedge[\delta] \text{ and } contrary_{\mathcal{D}}(\delta, \delta').$$

Now let c' be the mixed context $\langle \mathcal{W}', \mathcal{D}', \leq \rangle$, where
1. $\mathcal{W}' = \mathcal{W}$;
2. $\mathcal{D}'$ is the set of rules δ' obtained thus: For every rule $\delta = \dfrac{X : \mathcal{Z}_{Old}}{\bigcirc Y}$ in $\mathcal{D}$,
   A. Let $\mathcal{D}_\delta = \{\delta'' \in \mathcal{D} : \delta \preceq \delta''\}$;
   B. let $\mathcal{P} = \{\neg Premise[\delta''] : \delta'' \in \mathcal{D}_\delta\}$;

C. set δ' to be $\dfrac{X : \mathcal{Z}_{Old} \backslash \mathcal{P}}{Y}$;

3. $\delta \leq \delta'$ if and only if $\dfrac{Premise[\delta] : \mathcal{Z}}{\bigcirc Conclusion[\delta]} \preceq \dfrac{Premise[\delta'] : \mathcal{Z}'}{\bigcirc Conclusion[\delta']}$.

Any mixed context constructed by this procedure turns out to be equivalent to the hedged context it's constructed from:

Theorem 6.6 (Hedged-to-Mixed) *Let the context $c' = \langle \mathcal{W}', \mathcal{D}', \leq' \rangle$ be derived from some hedged context $c = \langle \mathcal{W}, \mathcal{D} \rangle$ by the procedure specified in Definition 6.5. Then c' and c are equivalent.*

Proof. Let $c'' = \langle \mathcal{W}'', \mathcal{D}'' \rangle$ be the result of applying Definition 6.2 to c'. By Theorem 6.4, the contexts c' and c'' are equivalent. What we need to do is show that $c = c''$. It's obvious that $\mathcal{W} = \mathcal{W}''$. So it remains to show that $\mathcal{D} = \mathcal{D}''$:

Without loss of generality, we establish only $\mathcal{D} \subseteq \mathcal{D}''$: Consider an arbitrary rule δ from $\mathcal{D}$. By Definition 6.5, there must be some rule δ' in $\mathcal{D}'$ such that $counterpart_c(\delta') = \delta$. By Definition 6.2, there must be a rule δ'' in $\mathcal{D}''$ such that $counterpart_{c'}(\delta'') = \delta'$. Clearly, $Premise[\delta''] = Premise[\delta]$ and $Conclusion[\delta''] = Conclusion[\delta]$. Now we need to show that $Hedge[\delta''] = Hedge[\delta]$.

$Hedge[\delta] \subseteq Hedge[\delta'']$: Consider an arbitrary $\neg Z \in Hedge[\delta]$. Either there's some rule δ^* such that $contrary_{\mathcal{D}}(\delta, \delta^*)$ and $Z = Premise[\delta^*]$, or there isn't. If yes, then, by Definition 6.5, there's a rule $\delta^\dagger \in \mathcal{D}'$ such that $counterpart_c(\delta^\dagger) = \delta^*$ and $\delta' \leq \delta^\dagger$. Also, notice that $contrary_{\mathcal{D}'}(\delta', \delta^\dagger)$. By Definition 6.2, $\delta^\dagger$ is in $\mathcal{D}'_{\delta'}$, and so $\neg Premise[\delta^\dagger] \in Hedge[\delta'']$. But $Premise[\delta^\dagger] = Premise[\delta^*] = Z$. So, $\neg Z \in Hedge[\delta'']$. If there's no rule δ^* with $contrary_{\mathcal{D}}(\delta, \delta^*)$ and $Z = Premise[\delta^*]$, then, by Definition 6.5, $\neg Z \in Hedge[\delta']$, and, from this and Definition 6.2, $\neg Z \in Hedge[\delta'']$.

$Hedge[\delta''] \subseteq Hedge[\delta]$: Take some $\neg Z \in Hedge[\delta'']$. Either there is a rule $\delta^* \in \mathcal{D}''$ with $contrary_{\mathcal{D}''}(\delta'', \delta^*)$ and $Premise[\delta^*] = Z$, or not. If yes, then, by Definition 6.2, there's a rule $\delta^\dagger \in \mathcal{D}'$ such that $counterpart_{c''}(\delta^\dagger) = \delta^*$, $contrary_{\mathcal{D}'}(\delta', \delta^\dagger)$, and $\delta' \leq \delta^\dagger$. In light of Definition 6.5, $\delta' \leq \delta^\dagger$ is enough to conclude that $\delta^\dagger \in \mathcal{D}'_{\delta'}$, and so that $counterpart_c(\delta') = \delta \preceq counterpart_c(\delta^*)$. This latter fact means that $\neg Premise[counterpart_c(\delta^*)]$ is in $Hedge[\delta]$. But $Premise[counterpart_c(\delta^*)] = Premise[\delta^*]$, and so $\neg Z \in Hedge[\delta]$. In case there's no rule $\delta^* \in \mathcal{D}''$ with $contrary_{\mathcal{D}''}(\delta'', \delta^*)$ and $Premise[\delta^*] = Z$, Definition 6.5 entails that $\neg Z \in Hedge[\delta']$, and this fact together with Definition 6.2 entail that $\neg Z \in Hedge[\delta]$. □

So there's a close connection between regular mixed and hedged contexts. But what about irregular mixed contexts that have no hedged counterparts? Notice that any such context will contain at least one pair of contrary rules δ, δ' with $\delta \leq \delta'$ and either $\neg Premise[\delta] \in Hedge[\delta']$, or $\neg Premise[\delta'] \in Hedge[\delta]$. The former type of pattern, that is, one where $\neg Premise[\delta] \in Hedge[\delta']$, strikes me as borderline incoherent: The principle instantiated by δ' is stronger than the one instantiated by δ, and yet the feature that makes the weaker principle apply is also what undercuts the stronger one. But the latter type of pattern

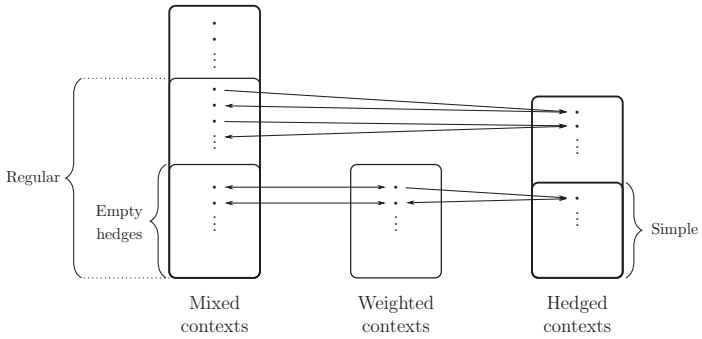

Fig. 1. Relations between various types of contexts

is a different matter; and if there are clear cases exhibiting it, the mixed view has an upper hand over the view on which principles are hedged. Why? Well, because only the mixed view has the resources to distinguish such cases from the mundane ones where the stronger principle outweighs the weaker one. [25]

In Section 4, we already discussed the relation between the model on which principles are contributory and the one on which they are hedged. Here we state it in the form of a corollary to Theorems 6.4 and 6.6: [26]

Corollary 6.7 (Contributory is simple hedged) *For every weighted context c, there's a (simple) hedged context c' such that c and c' are equivalent, and, for every simple hedged context c', there's a weighted context c such that c and c' are equivalent.*

The relations between all three types of contexts—mixed, weighted, and hedged—are summarized in Figure 1.

7 Conclusion and outlook

The main goal of this paper was to explore the connections between three different takes on moral principles. On the first, they are contributory; on the second, they are hedged; on the third, they are contributory, but can also have hedges. We saw that there are close connections between the models of these views. This strongly suggests that the views themselves are much closer than standardly thought. And while it might be obvious that the mixed view extends the contributory-principles view, it would be quite surprising

[25] Unfortunately, all the examples I could think of here are quite controversial. Here's one such, coming from the epistemic domain. First, notice that the rules "If an agent perceives that X, then that perception speaks in favor of believing that X" and "If an agent has outstanding testimony that X, then it speaks in favor of believing that X" are the epistemic analogues of contributory/mixed moral principles. Epistemologists sometimes invoke an infallible Epistemology Oracle reporting the truth to the agent—see, e.g., [34]—and we might imagine a situation where some agent sees a red-looking object, but is also told, by this Oracle, that the object is blue. One could hold that, in this case, the Oracle's testimony doesn't simply outweigh, but also undercuts the Perception rule.

[26] The proof runs parallel to those of the theorems. It's also simpler.

if it turned out that the contributory-principles view was only as expressive as the hedged-principles view, or if the differences between the mixed view and the hedged-principles view were only cosmetic. It may be too early to claim that the correspondence results established here show that, since the models we set up here are very simple. So one direction for future research is to extend the correspondence results established here to more expressive models of the views. Another one is to explore the ramifications of this result for claims about views on principles advanced in the philosophical literature. [27] Yet another is to explore the connections between hedged rules and "exclusionary rules" discussed in [8, Sec. 6], or, roughly, rules that take other rules out of consideration. On the face of it, having an exclusionary rule δ that gets triggered when X obtains, taking some other rule δ' out of consideration, isn't all that different from thinking of δ' as having $\neg X$ listed in its hedge. Finally, it would be interesting to explore how the models and results presented here might be relevant in the context of the debate between generalists and specificationists about rights. [28]

References

[1] Bradley, D., *Are there indefeasible epistemic rules?*, Philosopher's Imprint **19** (2019), pp. 1–19.
[2] Dancy, J., "Ethics without Principles," Oxford University Press, 2004.
[3] Elga, A., *How to disagree about how to disagree*, in: R. Feldman and T. Warfield, editors, *Disagreement*, Oxford University Press, 2010 pp. 175–86.
[4] Frederick, D., *Pro-tanto obligations and ceteris-paribus rules*, Journal of Moral Philosophy **12** (2015), pp. 255–66.
[5] Goble, L., *Normative conflicts and the logic of 'ought'*, Noûs **43** (2009), pp. 450–89.
[6] Holton, R., *Principles and particularisms*, in: *Proceedings of the Aristotelian Society, Suppl. Volume 76*, 2002, pp. 191–210.
[7] Horty, J., *Reasons as defaults*, Philosophers' Imprint **7** (2007).
[8] Horty, J., "Reasons as Defaults," Oxford University Press, 2012.
[9] Knoks, A., "Defeasibility in Epistemology," Ph.D. thesis, University of Maryland, College Park (2020).
[10] Lance, M. and M. Little, *Particularism and antitheory*, in: D. Copp, editor, *The Oxford Handbook of Ethical Theory*, Oxford University Press, 2006 pp. 567–94.
[11] Lance, M. and M. Little, *Where the laws are*, Oxford Studies in Metaethics **2** (2007), pp. 149–71.
[12] Lewis, D., *Ordering semantics and premise semantics for counterfactuals*, Journal of philosophical logic **10** (1981), pp. 217–34.
[13] Makinson, D. and L. van der Torre, *Input/output logics*, Journal of Philosophical Logic **29** (2000), pp. 383–408.
[14] Makinson, D. and L. van der Torre, *Constraints for input/output logics*, Journal of Philosophical Logic **30** (2001), pp. 155–85.
[15] McKeever, S. and M. Ridge, "Principled Ethics: Generalism as a Regulative Ideal," Oxford University Press, 2006.

[27] In [9, Ch. 2] I make some steps in this direction.

[28] See [17] and [33, Sec. 5.2] for an introduction to the debate and further pointers to the literature. Also, since this is the final footnote, I thank the anonymous reviewers for their feedback and insightful comments.

[16] McNamara, P., *Deontic logic*, in: E. N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*, Metaphysics Research Lab, Stanford University, 2019 .
[17] Mullins, R., *Moral conflict and the logic of rights*, Philosophical Studies **117** (2020), pp. 633–51.
[18] Nair, S. and J. Horty, *The logic of reasons*, in: D. Star, editor, *The Oxford Handbook of Reasons and Normativity*, Oxford University Press, 2018 pp. 67–84.
[19] Owens, D., *Duress, deception and the validity of a promise*, Mind **116** (2007), pp. 293–315.
[20] Parfit, D., "On What Matters: Volume One," Oxford University Press, 2011.
[21] Pollock, J., "Knowledge and Justification," Princeton: Princeton University Press, 1974.
[22] Pollock, J. and J. Cruz, "Contemporary Theories of Knowledge," Rowman & Littlefield Publishers, 1999.
[23] Raz, J., "Engaging Reasons: On the Theory of Value and Action," Oxford University Press, 1999.
[24] Reiter, R., *A logic for default reasoning*, Artificial Intelligece **13** (1980), pp. 81–132.
[25] Ross, W. D., "The Right and the Good," Oxford University Press, 1930.
[26] Scanlon, T. M., "What We Owe to Each Other," Cambridge, MA: Harvard University Press, 1998.
[27] Scanlon, T. M., *Principles and particularisms*, in: *Proceedings of the Aristotelian Society, Suppl. Volume 74*, 2000, pp. 301–17.
[28] Schroeder, M., "Slaves of the Passions," New York: Oxford University Press, 2007.
[29] Schroeder, M., "Reasons First," Oxford University Press, 2021, (forthcoming).
[30] Star, D., *Introduction*, in: D. Star, editor, *The Oxford Handbook of Reasons and Normativity*, Oxford University Press, 2018 pp. 1–21.
[31] Titelbaum, M., *Rationality's fixed point (or: in defense of right reason)*, Oxford Studies in Epistemology **5** (2015), pp. 253–294.
[32] Väyrenen, P., *A theory of hedged moral principles*, Oxford Studies in Metaethics **4** (2009), pp. 91–132.
[33] Wenar, L., *Rights*, in: E. N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*, Metaphysics Research Lab, Stanford University, 2021 .
[34] White, R., *Epistemic permissiveness*, Philosophical Perspectives **19** (2005), pp. 445–59.