

## DUTY AND KNOWLEDGE

Yoav Isaacs  
Princeton University

### 1. Motif

Deontological ethics needs a decision theory. But it has a hard time getting one.

### 2. Motivating Deontological Decision Theory

It's fairly easy to see what a deontological ethic demands of an omniscient agent. If killing an innocent person were absolutely forbidden, then the ethic would require that the agent never choose to kill an innocent person no matter what. If promise-breaking were absolutely forbidden, then the ethic would require that the agent never choose to break a promise no matter what. And so on. An omniscient agent can follow a deontological ethic just by abstaining from any action which is categorically forbidden.<sup>1</sup>

It's far less easy to see how an agent acting under uncertainty would follow a deontological ethic. If killing an innocent person were absolutely forbidden, under what conditions can an agent who is unsure whether or not some person is innocent kill that person? If promise-breaking were absolutely forbidden, under what conditions can an agent who is unsure whether or not some act would break a promise perform that act? What should a deontologist do when he is unsure whether or not some possible act violates his deontological scruples? If deontological ethics has no decision theory—if there's no way for deontology to tell an agent how he should act in light of the limited information he has available rather than in light of the fullness of all truth—then deontology leaves its adherents rather in the lurch. Without a decision theory, deontology is a theory of rightness and wrongness which is of no use to agents trying to decide what to do. And conscientious deontologists need some way to figure out what to do.

If deontological ethics doesn't have a decision theory, then deontological ethics is on rather worse footing than consequentialist ethics, which does have a

decision theory. Expected utility theory provides a well-established way to take an agent's state of uncertainty (represented as a credence function) and facts about value (represented as a utility function) and thereby specify which course of action an agent should undertake.<sup>2</sup>

Consider a trivial case. Suppose that you're given the opportunity to bet your life for a dollar on the outcome of a fair coin toss. If it landed Heads you die, if it landed Tails you get a buck. Should you take the bet? One answer (sensible in its own, odd way) is that it depends: if the coin landed Heads then you shouldn't take the bet, and if the coin landed Tails then you should take the bet. Everyone can agree that taking the bet results either in death or in a dollar, and everyone can agree that death is a bad outcome and that a dollar is a good outcome. But if no other sort of answer to that question were available, then there'd be no way for an uncertain agent to determine whether or not he should take that bet. You'd be in bad shape indeed if you had no way to evaluate that bet without unavailable information.

Deontologists have an easy time characterizing the objective ought, but a much more difficult time characterizing the subjective ought. And the subjective ought is where the real action is—the subjective ought is strictly more informative than the objective ought. One can derive the objective ought from the subjective ought. Suppose that an agent must decide whether or not to perform some act  $\phi$ . That agent objectively ought to  $\phi$  if and only if an agent who was certain of the actual state of the world subjectively ought to  $\phi$ .<sup>3</sup> The limiting case of the subjective ought is the objective ought. But while the subjective ought can accommodate any degree of uncertainty (down to none at all), the objective ought cannot accommodate any uncertainty whatsoever.

One note in passing: The import and irreducibility of the subjective ought shows how misguided the project of “consequentializing” ostensibly non-consequentialist ethical theories is.<sup>4</sup> Since the “consequentializing” occurs in the limited domain of the objective ought, it amounts to nothing more than constructing a total order over worlds relative to which a maximizing choice function can be defined. But the subjective ought must be a choice function over probability functions over worlds, and it's not even clear what properties the “consequentialized” theory's subjective ought should have.<sup>5</sup> By considering the objective ought but not the subjective ought, “consequentializers” create the impression that consequentialist and deontological ethical theories are more similar than they actually are.

### 3. The Problem

How should a deontological ethic influence decision-making in conditions of uncertainty? There are, broadly speaking, three norms through which deontological scruples could be applied—and none of them are any good.<sup>6</sup>

**Permissive Norm:** The ethic forbids  $\phi$ -ing if and only if the agent is certain that  $\phi$ -ing would violate a deontological scruple.

**Restrictive Norm:** The ethic forbids  $\phi$ -ing if and only if the agent is not certain that  $\phi$ -ing would not violate a deontological scruple.

**Intermediate Norm:** The ethic sometimes forbids and sometimes permits  $\phi$ -ing when the agent is not certain whether or not  $\phi$ -ing would violate a deontological scruple.

The permissive norm is no good because it's *hugely* too permissive. Certainty is rare. If the prohibition against killing innocents applies only when an agent is certain that someone is innocent, then that agent may kill just about anyone without violating the prohibition. If the prohibition against lying applies only when an agent is certain that an assertion is false, then that agent may assert just about anything without violating the prohibition. The permissive norm pretty much never prohibits any action. That's bad.

The restrictive norm is no good because it's *hugely* too restrictive. Certainty is rare. If the prohibition against killing innocents applies whenever an agent is not certain that someone is not innocent, then that agent is basically prohibited from killing. If the prohibition against lying applies whenever an agent is not certain an assertion is true, then that agent is basically prohibited from making assertions. The restrictive norm pretty much never permits any action. That's bad.

We are therefore left with the intermediate norm. And while the intermediate norm isn't as terrible as either the permissive norm or the restrictive norm, it still looks pretty bad. It faces two major difficulties, which I will dub the *Arbitrariness Problem* and the *Conjunction Problem*.

The *Arbitrariness Problem* is that there seems to be no good reason for some particular intermediate level of confidence to separate permissibility from impermissibility. Why should it be permissible to kill someone when one has a credence of .6372 that they are innocent but impermissible to kill someone when one has a credence of .6373 that they are innocent? Whether there is a constant threshold or whether there are different thresholds for different acts (or even different thresholds for different circumstances) there doesn't seem to be a good reason for *any* threshold. The thresholds seem arbitrary, and arbitrary thresholds seem implausible.<sup>7</sup>

The *Conjunction Problem* is that thresholds interact poorly with multiplicities of acts. Jackson and Smith press this problem forcefully:

Suppose that there are two skiers, X and Y, whose probability of innocence is individually just below the threshold . . . and whose paths threaten separate groups of people. On the threshold view, you ought to shoot X if the number saved is large enough because the probability of innocence is below the level at which the absolute prohibition kicks in; ditto for Y. (Jackson and Smith 2006).

Since X's and Y's probabilities of innocence are each individually just below the threshold, the probability that at least one of them is innocent is over the

threshold.<sup>8</sup> Is it, then, permitted to kill each of them with two distinct actions (say, with two bullets) but forbidden to kill both of them with a single action (say, with a grenade)? This seems ridiculous. While there are various possible norms for multiplicities of acts, no extant candidate seems particularly palatable.<sup>9</sup>

#### 4. Knowledge Norms

Let's take it for granted that there's no good way for a deontological ethic to influence decision-making in conditions of uncertainty.<sup>10</sup> Let's also take it for granted that deontological ethics are in serious trouble if they cannot accommodate a decision theory. Even granting all of this it does not follow that deontological ethics are in serious trouble, but merely that to avoid serious trouble deontological ethics will need a decision theory which does not deal in uncertainties. Perhaps deontology will need its decision theory to be formulated in terms of some other epistemological ideology. I contend that a knowledge-based decision theory will fit the bill.

How should a deontological ethic influence decision-making in conditions of limited knowledge? Again there are, broadly speaking, three norms through which deontological scruples could be applied—but these look *much* better.

**Permissive Norm:** The ethic forbids  $\phi$ -ing if and only if the agent knows that  $\phi$ -ing would violate a deontological scruple.

**Restrictive Norm:** The ethic forbids  $\phi$ -ing if and only if the agent doesn't know that  $\phi$ -ing would not violate a deontological scruple.

**Intermediate Norm:** The ethic sometimes forbids and sometimes permits  $\phi$ -ing when the agent does not know whether or not  $\phi$ -ing would violate a deontological scruple.

Both the permissive norm and the restrictive norm look *much* better when framed in terms of knowledge than when framed in terms of certainty. Agents both know a great deal and fail to know a great deal. Agents often know that some act would violate a deontological scruple and often do not know that some act would violate a deontological scruple, so the permissive norm is non-trivial. Agents often know that some act wouldn't violate a deontological scruple and often do not know that some act wouldn't violate a deontological scruple, so the restrictive norm is non-trivial. Both the permissive norm and the restrictive norm are non-arbitrary, and since both the permissive norm and the restrictive norm are viable, there is no need to worry about the intermediate norm's problems with arbitrariness. But while both the permissive norm and the restrictive norm evade the arbitrariness problem, each can face the conjunction problem, so let's turn to that.

The conjunction problem arises for the permissive norm if it's possible for an agent to have more knowledge about a conjunction of acts than he does about either of the acts conjoined.<sup>11</sup>

Permissive Problem Case: The agent doesn't know that  $\phi$ -ing would violate a deontological scruple and doesn't know that  $\psi$ -ing would violate a deontological scruple but does know that both  $\phi$ -ing and  $\psi$ -ing would violate a deontological scruple.

It's obviously possible to have more knowledge about a conjunction of acts than about either of the acts conjoined. There's nothing questionable about saying that an agent doesn't know that buying the first ticket in a two ticket lottery would lead to his winning and that the agent doesn't know that buying the second ticket in the two ticket lottery would lead to his winning but that the agent does know that buying both the first ticket and the second ticket in the two ticket lottery would lead to his winning.

How would this sort of problem case play out? Let's return to the skier case. Suppose that the agent knows that exactly one of the two skiers is innocent, but that he doesn't know which. The agent would be permitted to kill either skier, but forbidden to kill both. This is a bit of an unpalatable result, but not completely horrible.<sup>12</sup>

The conjunction problem arises for the restrictive norm if it's possible for an agent to have less knowledge about a conjunction of acts than he does about each of the acts conjoined.

Restrictive Problem Case: The agent knows that  $\phi$ -ing wouldn't violate a deontological scruple and knows that  $\psi$ -ing wouldn't violate a deontological scruple but doesn't know that both  $\phi$ -ing and  $\psi$ -ing wouldn't violate a deontological scruple.

It's debatable whether or not it's possible to have less knowledge about a conjunction of acts than about each of the acts conjoined. Can an agent know that  $\phi$ -ing is licit and know that  $\psi$ -ing is licit without knowing that both  $\phi$ -ing and  $\psi$ -ing are licit? There's some measure of controversy over whether knowledge is closed under entailment,<sup>13</sup> and such a strong closure principle would obviously rule this sort of case out. A weaker principle can still help with this case. If agents are always in a position to know the entailments of what they know, this would suffice to ensure that whenever an agent knows that  $\phi$ -ing is licit and knows that  $\psi$ -ing is licit, he is at least in a position to know that both  $\phi$ -ing and  $\psi$ -ing are licit. Timothy Williamson's knowledge-first epistemology employs just such a framework.<sup>14</sup> One might, however, employ a notion of knowledge according to which it's possible for an agent to know  $p$ , know  $q$ , but still be quite agnostic about  $p \wedge q$ .

How would this sort of problem case play out? Let's return to the skier case again. Suppose that the agent knows that each of the two skiers is not innocent, but that he doesn't know that both skiers are not innocent. The agent would be permitted to kill either skier, but forbidden to kill both. This is a bit of an unpalatable result, but not completely horrible.

Although both the permissive norm and the restrictive norm are viable, I prefer the restrictive norm. Williamson's knowledge-first epistemology is the best

knowledge-based epistemology in which knowledge does real work.<sup>15</sup> Under the auspices of knowledge-first epistemology the restrictive norm evades the conjunction problem while the permissive norm does not. More importantly, the restrictive norm seems more appropriate for deontological ethics than the permissive norm. Deontologists take the idea of killing an innocent person *very* seriously. A deontologist would not knowingly take an innocent life *per se* *mundus*—though the world perish.<sup>16</sup> It therefore seems inappropriately cavalier for a deontologist to think that he is permitted to kill someone because he doesn't know whether or not that person is innocent.<sup>17</sup> If the deontological subjective ought is formulated through the permissive norm, then an agent who obeyed the subjective ought might still violate the objective ought. But if the deontological subjective ought is formulated through the restrictive norm, then an agent who obeyed the subjective ought would be guaranteed to obey the objective ought.

## 5. Battling Bayesianisms: Certainty-Based vs. Knowledge-Based

Knowledge norms are best thought of as part of a larger, knowledge-first epistemology. That framework is therefore worth exploring in a bit more detail. And it's best explored through contrast.

The standard framework for decision theory employs Bayesian probabilities. Probabilities provide enough detail about an agent's epistemic state to do the fine-grained work that decision theory requires.<sup>18</sup> While there are various probabilistic epistemologies in the offing, they share a common formal structure.

Traditional Bayesians impose two kinds of coherence requirements on rational credences, a synchronic constraint which applies at each time and a diachronic constraint which applies over different times. The synchronic requirement is **probabilistic coherence**: that the agent's credences at any time should form a probability function. That is, an agent's degrees of belief must (at a minimum) conform to the following axioms of the probability calculus:

- (1)  $\Pr(p) \geq 0$ , for any proposition  $p$ .
- (2)  $\Pr(t) = 1$ , for any tautology  $t$ .
- (3)  $\Pr(p \vee q) = \Pr(p) + \Pr(q)$ , for any inconsistent propositions  $p$  and  $q$ .

The diachronic requirement is **conditionalization**. This constraint relies on a notion of conditional probability, which we can define given the axioms above. The probability of  $p$  conditional on  $q$  is equal to the probability of  $p$  and  $q$  divided by the probability of  $q$  (assuming that  $\Pr(q) > 0$ ).

$$\Pr(p \mid q) =_{df} \frac{\Pr(p \wedge q)}{\Pr(q)}.$$

Conditionalization mandates that when an agent has evidence  $E$  at time  $a$  and evidence  $E+$  at time  $b$  (where  $E+$  implies  $E$ ), then the probability of any proposition  $p$  at  $b$  should equal the conditional probability of  $p$  given  $E+$  at  $a$ . That is,  $\text{Pr}_b(p) = \text{Pr}_a(p \mid E+)$ .

Putting this all together, the basic story of Bayesian epistemology goes something like this. Rational agents assign non-negative credence to various possible worlds in such a way that the sum of those credences is 1. When rational agents gain evidence which is inconsistent with some of those possible worlds, they adjust their credences in those worlds to 0, and proportionally increase their credences in the remaining worlds so that the sum of their credences returns to 1.

Conditionalization presupposes a notion of evidence. But what, exactly, is evidence? Different epistemologies offer different answers. Just as earlier there were certainty-based norms and knowledge-based norms, there is the certainty-based conception of evidence and knowledge-based conception of evidence.

The more traditional notion of evidence (at least among decision theorists) is certainty-based.<sup>19</sup> The basic idea is that an agent's evidence is what that agent has complete, utter, bet-your-life-for-a-nickel confidence in. An agent takes his evidence for granted. The traditional thought is that agents are certain of their phenomenal states. (This view is often called the phenomenal conception of evidence.) Suppose an agent is looking at a box and that the box looks red. The agent might be unsure whether the box really is red or whether the box merely looks red due to some tricky lighting, but (the traditional thought goes) the agent can be certain that the box looks red. Agents are supposed to have privileged, infallible access to their phenomenal states.

The certainty-based conception of evidence faces a serious problem. There's a strong case to be made that agents (or at least agents like us) are never justified in being certain of anything non-trivial. We may well have privileged access to our own phenomenal states, but it's quite clear that that access is not infallible. Suppose that an agent takes a quick glance at a painting with 137 black dots.<sup>20</sup> That agent's phenomenal state will have all 137 black dots represented (at least supposing that the agent's eyesight is decent). Nonetheless, it seems completely reasonable for the agent to be unsure of exactly how many black dots there are on the painting after the quick glance. The quick glance of the 137 black dots contained all the information the agent would have needed to be sure that the painting contained 137 black dots, but an agent need not have access to all the information contained in his phenomenal state.

Think about the staggeringly huge number of possible phenomenal states. For each phenomenal state there are myriad other phenomenal states just barely different from it. There are then other phenomenal states just barely different from those (though perhaps more than barely different from the original phenomenal state). And so on. Agents like us are not justified in being certain of which phenomenal state we're in—we're not good enough at distinguishing different phenomenal states.

Now one might think that while we are not justified in being certain of our exact phenomenal state, we can be justified in being certain that our phenomenal state falls into some range of phenomenal states. But I don't think this is plausible. There does not seem to be a sharp cutoff for how mistaken agents like us can be when trying to identify our phenomenal states. Consider the continuum of visible colors. It's obvious that I can mistake one shade of red for another shade of red. But I have no reason to think that it's impossible for me to mistake a shade of red for a shade of red/orange, or to mistake a shade of red for a shade of orange, or to mistake a shade of red for a shade of green. I think I'm unlikely to mistake a shade of red for a shade of green, but there's no last shade I might mistakenly pick out—I might mistake any shade for any other shade at all.<sup>21</sup>

Let's take for granted that agents like us are never warranted in being certain of anything non-trivial. The certainty-based conception of evidence would thus leave agents like us insuperably without evidence, and that would be bad. Epistemology should not decree that agents like us never have any reason to change our minds about anything. We therefore need an alternative notion of evidence.

Enter the knowledge-based conception of evidence. Knowledge is *much* easier to come by than certainty.<sup>22</sup> A key principle of Williamson's knowledge-first epistemology is that  $E = K$ , that an agent's evidence is the totality of what that agent knows. While knowledge can play the role of evidence, since knowledge is quite different from certainty, knowledge-based epistemology will be quite different from certainty-based epistemology.

A few differences between the certainty-based conception of evidence and the knowledge-based conception of evidence are notable. Although the role played by an agent's sense certainties in the former case is played by what the agent knows in the latter case, the role changes quite a bit depending on its occupant. Complete certainty that  $p$  would warrant betting one's life that  $p$  in order to get a nickel—one would think of such a bet as a free nickel. But the knowledge that  $p$  does not warrant betting one's life that  $p$  in order to get a nickel. Knowing  $p$  is not the loftiest epistemic state an agent can have regarding  $p$ . An agent could know that  $p$  while failing to know that he knows that  $p$ , or know that  $p$  and know that he knows that  $p$  while failing to know that he knows that he knows that  $p$ , and so on. The loftiest epistemic state an agent can have regarding a proposition would be all possible iterations of knowing<sup>23</sup>, but the arguments against justified certainties would show that agents never have all possible iterations of knowledge regarding any non-trivial proposition. Knowing that  $p$  is easier than having complete certainty that  $p$ , which is why the knowledge-based framework is not forced to be too permissive, too restrictive, or arbitrary. But just as knowing that  $p$  is easier than having complete certainty that  $p$ , knowing that  $p$  is not as helpful as having complete certainty that  $p$ . It was a hope of certainty-based decision theory that it would provide agents with rules they would always know how to follow. But since agents are not always in a position to know what they know, knowledge-based decision theory cannot possibly provide agents with rules they would always know how to follow. The proponent of



knowledge-based decision theory therefore needs some way to distinguish blameless violations of the theory's norms from blameworthy violations of the theory's norms.<sup>24</sup> For example, we might censure an agent for violating the norms when in a position to know that he was violating them but not when he had every reason to think that he was following them.<sup>25</sup> These complications impose serious costs for a knowledge-based decision theory.<sup>26</sup> But deontology is simply stuck with those costs. Deontology needs a decision theory, and there is no adequate alternative.

## 6. An Inadequate Alternative

Altering the formal structure of standard decision theory in order to accommodate deontological scruples leaves us with a decision theory which is either too permissive, too restrictive, or arbitrary. A knowledge-first decision theory solves these problems, but it does cause a few others. And while I do think that a knowledge-first decision theory is the best option for deontological ethics, it's worth making sure that we can't solve all our problems with a bit of creativity. Perhaps instead of altering the formal structure of standard decision theory we could just impose some new constraints within the formal structure of standard decision theory. Perhaps an agent with deontological scruples is just an agent whose utility function has certain specific features. If killing an innocent person were absolutely forbidden, then the agent would prefer any world in which he does not kill an innocent person to any world in which he does kill an innocent person.<sup>27</sup> If promise-breaking were absolutely forbidden, then the agent would prefer any world in which he does not break a promise to any world in which he does break a promise. And so on.

Unfortunately, this attempt to situate deontological scruples within standard decision theory causes its own insuperable problems.<sup>28</sup> The less grievous problem is that the agent's deontological scruples involve essentially first-personal facts. The agent prefers any world in which *he* does not kill an innocent person to any world in which *he* does kill an innocent person. Thus two different deontologists would have to have different utility functions—if one of them must be a murderer, then each must prefer that it be the other. A mere general abomination of killing innocents will not do, for such a general abomination would favor murdering the innocent child of a serial killer if by doing so one could demoralize the serial killer enough to stop him from killing, and this is precisely the sort of thing that deontological scruples forbid. The inclusion of first-personal facts is not an unreasonable feature for a decision theory. Any theory meant to allow for self-interest must include first-personal facts, and standard decision theory is meant to allow for self-interest.<sup>29</sup> I note, however, that the pertinence of first-personal facts for deontological ethics would then represent a decisive departure from consequentialist ethics for which first-personal facts must be irrelevant. One more nail in the coffin for "consequentializing" deontological ethical theories.

The more serious problem is that restricting utility functions to accommodate deontological scruples makes them ridiculous. Suppose that an agent prefers any world in which he does not kill an innocent person to any world in which he does kill an innocent person. That agent would therefore prefer for others to commit the worst horrors in the history of the world, spreading unprecedented misery throughout the Earth before ultimately extinguishing life itself than for him to kill an innocent person. Such an agent will, for all practical purposes, be an absolute pacifist. To his mind, the risk of killing an innocent person is too staggering to take given any plausible level of uncertainty. If promise-breaking were so forbidden, the agent would have to consider the mere act of making a promise—thus risking the most horrific possibility that he wind up breaking that promise—to be unconscionably reckless. This restriction on utility functions is too restrictive, and validates the scruples that deontologists ought to accept only by also validating other scruples that deontologists ought to reject.

Restrictions to utility functions warp deontological sensibilities. It's not the case that deontologists think that killing innocents always makes things worse and is therefore wrong. Deontologists think that killing innocents is wrong whatever its consequences may be. Restricting utility functions is just not a viable option for deontologists.

## 7. Some Formal Analysis

I've argued that the subjective ought of a deontological ethic should be formulated through the restrictive knowledge-based norm. This norm forbids  $\phi$ -ing if and only if the agent does not know that  $\phi$ -ing would not violate a deontological scruple. This subjective ought is not any sort of expected utility maximization, but is instead expected utility maximization subject to a constraint. (We saw in the last section that unconstrained expected utility maximization does not provide a viable context for a deontological decision theory.) Since I'm advocating that deontologists eschew expected utility maximization (albeit for something closely related to it) I owe some explanation for why doing so is sensible. After all, there is a grand tradition in decision theory which considers representability as an expected utility maximizer to be the very hallmark of practical rationality.

Representability is cheap. I can represent my socks as expected utility maximizers if I want to. The significance of representation theorems is that they represent someone as an expected utility maximizer *in some specific, non-trivial way*. Standardly, an agent's preferences may or may not be representable as maximizing expected utility relative to some credence function and some utility function. There are a number of important representation theorems that have been proven, but they all share their basic structure with the first such theorem—the Von Neumann–Morgenstern utility theorem. The Von Neumann–Morgenstern utility theorem is also the easiest to understand, so I'll focus my analysis on it.

Suppose that an agent is faced with some lotteries. For these purposes, we'll define a lottery as being constituted by some set of mutually exclusive possibilities, with each possibility assigned a probability such that the sum of the probabilities is 1. One lottery might be a 50% chance of eating a delicious, ordinary cupcake and a 50% chance of eating a delicious, cyanide-laced cupcake, and another lottery might be a 50% chance of throwing a delicious, ordinary cupcake into the trash and a 50% chance of throwing a delicious, cyanide-laced cupcake into the trash. As you can see, one can think of an action as providing an agent with a lottery which corresponds to the credences he would have if he performed the action.

Von Neumann and Morgenstern proved that an agent's preferences over lotteries are representable as maximizing expected utility relative to some credence function and some utility function if and only if that agent's preferences satisfy four constraints:<sup>30</sup>

- (1) Totality: For any lotteries  $L$  and  $M$ , exactly one of the following holds: (1)  $L \succ M$ , (2)  $M \succ L$ , or (3)  $L = M$ .
- (2) Transitivity: If  $L \succeq M$  and  $M \succeq N$ , then  $L \succeq N$ .
- (3) Independence: If  $L \succ M$ , then for any  $N$  and any probability  $p \in (0, 1]$ ,  $pL + (1 - p)N \succ pM + (1 - p)N$ .
- (4) Continuity: If  $L \succeq M$  and  $M \succeq N$ , then there is a probability  $p \in [0, 1]$  such that  $pL + (1 - p)N = M$ .

The deontological decision theory I've proposed satisfies the first three constraints, violating only the fourth. Let's see why.

Totality requires that an agent either strictly prefer one lottery to another or be indifferent between them. Disallowing acts which an agent does not know to be deontologically licit doesn't interfere with this at all.

Transitivity requires that an agent's preferences over lotteries correspond to some sort of overall ranking. Again, disallowing acts which an agent does not know to be deontologically licit doesn't interfere with this at all.

Independence requires an agent's preference of one lottery over another to hold regardless of the independent possibility of some other outcome. Once more, disallowing acts which an agent does not know to be deontologically licit doesn't interfere with this at all. If both of the lotteries being composed with a third alternative are known to be licit, whatever preference the agent has between the two lotteries will transfer over to the higher-order lotteries of which they are a part. And if one of the two lotteries is known to be licit while the other is not, that preference will transfer over as well.<sup>31, 32</sup>

Continuity requires that there be some mixture of the best of three options and the worst of three options which will be of equal value to the middle of three options. But the restrictive knowledge norm disallows any act with nonzero chance of violating a deontological scruple, and that disallowance cannot

be diluted. The risk of violating a deontological scruple poisons a possibility, rendering it unfit for choice.

Continuity is, however, the least important of the Von Neumann/Morgenstern axioms. It has no real theoretical basis. The axiom of continuity requires either that utility functions be bounded or that there be some restriction on the composition of lotteries, but the only motivation for either requirement is that they make the math easier.<sup>33</sup> I am not troubled that the theory I have proposed violates continuity.<sup>34</sup>

On the whole, the formal structure of knowledge-first deontological decision theory is extremely close to that of traditional decision theory.

## 8. Conclusion

Why should you accept this package of deontological ethics and knowledge-first epistemology? Well, frankly, I don't think you should. I like neither deontological ethics nor knowledge-first epistemology. It's stable to like neither. Similarly, it's stable to like just knowledge-first epistemology and favor some non-deontological ethic. But a predilection for deontological ethics should strongly incline one to knowledge-first epistemology. A major reason to like standard Bayesian epistemology is that it's good for decision theory.<sup>35</sup> But in fact standard Bayesian epistemology is only good for consequentialist decision theory. The way to make deontological decision theory work is to adopt a knowledge-first epistemology. So if you like deontological ethics, that's a major motivation to go knowledge-first in epistemology. And if you don't like knowledge-first epistemology, that's a major motivation to reject deontological ethics.<sup>36</sup>

## Notes

1. There are some niceties to be worked out regarding moral dilemmas, but the basic cases are unproblematic.
2. There are, of course, debates in decision theory about how expected utilities should be calculated. But the vast majority of cases are universally agreed upon.
3. Note that an agent who was certain of the actual state of the world might have inaccurate credences. For example, if in the actual world no agent was certain of anything, then this agent (who would obviously have to inhabit some other possible world) would be certain that no agent was certain of anything.
4. "Consequentializers" attempt to show that ostensibly non-consequentialist ethical theories are (or at least can be thought of as being) consequentialist ethical theories. For more, see Brown (2011).
5. A *prima facie* plausible standard for a "consequentialized" subjective ought is that it ought to make the theory representable as maximizing expected utility. But there are two reasons why this standard is unavailable to contemporary ethicists. First, John von Neumann and Oskar Morgenstern showed the conditions under

- which an agent is representable as an expected utility maximizer in 1947, so there's not much work for contemporary ethicists to do on that front. Secondly, deontological ethical theories are wildly incompatible with the prerequisites of expected utility maximization. More on this later.
6. Here I assume that deontological agents should choose from among their permissible acts so as to maximize expected utility, and that the role of the deontological norm is to determine which acts are permissible. I will consider an alternative approach to deontological norms in section 6.
  7. In contrast, consequentialists can offer a perfectly sensible explanation for the cutoffs they endorse—the cutoffs are where the risks and the rewards equal out.
  8. Assuming that their probabilities of innocence are not maximally correlated, which is easily stipulated.
  9. See Aboodi, Borer, and Enoch (2008).
  10. I will expand on this case in the next section.
  11. For these purposes I assume that licitness respects agglomeration. That is, I assume that if  $\phi$ -ing is licit and  $\psi$ -ing is licit then  $\phi$ -ing and  $\psi$ -ing is licit. I thus leave aside situations in which an agent is explicitly permitted to do either of two acts but forbidden to do both. Such situations can arise, but they just don't have anything to do with the conjunction problem. Strictly speaking, the formalism of decision theory only allows an agent to perform a single act in any decision situation, thus making acts fully fine-grained. Nonetheless, thinking in a coarser grain is often helpful. This coarse grain is merely a heuristic; in situations in which licitness does not respect agglomeration the heuristic should be dropped.
  12. Since an agent might perform a sequence of acts which he knew to violate a deontological scruple even though no act within it was known to violate a deontological scruple, it is likely necessary to forbid agents from performing acts which will cause them to know that they violated a deontological scruple.
  13. See Lewis (2006) and Hintikka (1962).
  14. Although Williamson allows that knowledge is extendible under competent deduction, this principle is not necessary for a solution to the conjunction problem. Even if a conjunction were not known despite its conjuncts being known, the knowledge of the conjuncts would guarantee that the conjunction had an evidential probability (that is, a probability conditional on what the agent knows) of 1. And probability 1 makes for a non-arbitrary threshold.
  15. In contrast to, say, David Lewis' knowledge-based epistemology, which leaves the real work to be done by credences.
  16. *Fiat justitia, et pereat mundus*—let there be justice, though the world perish—was the personal motto of the Holy Roman Emperor Ferdinand I. This motto is famously (and approvingly) quoted by Immanuel Kant in his *Perpetual Peace*.
  17. I note that the restrictive norm is favored by those who support a knowledge norm for assertion. Surely the norm for killing someone should not be more lax than the norm for assertion.
  18. Outright beliefs and deduction are insufficient. For more, see Harman (1988).
  19. For example, here's (Lewis 1952).
  20. I owe the structure of this case to Roderick Chisholm's problem of the speckled hen.
  21. For more, see Williamson, (2002).

22. Barring extreme skeptical worries, of course.
23. An epistemically ideal agent would also have absence-luminosity: if an ideal agent didn't know that  $p$  he would know to all possible iterations that he didn't know that  $p$ .
24. And similarly between praiseworthy conformity to the theory's norms and non-praiseworthy conformity to the theory's norms.
25. There are other plausible methods for apportioning decision-theoretic praise and blame. Consider the following passage from John Hawthorne and Amia Srinivasan: "Suppose two tennis players are each in situations where they know they can't reach the ball in time. One player—call him Andi—gives up. Another, call him Raphael, chases the ball though (unsurprisingly) fails to get to it in time. Andi might deride Raphael as a pathetic figure, someone who gives chase while knowing the chase is futile. But we can see that because the absence of knowledge isn't luminous, Andi risks turning himself into a player who fails to chase the ball when he might, in fact, reach it in time. For cases will arise sooner or later where Raphael *believes* he knows he won't reach the ball, but in fact doesn't know this. In such cases, thanks to the habit he has inculcated in himself, Raphael will sometimes end up reaching the ball. On the other hand, Andi doesn't chase the ball both in cases where he knows that he won't reach it, *and* in those cases in which he falsely takes himself to know that he won't reach it. In this way, Andi—in failing to chase the ball even when he knows he won't reach it—inculcates in himself an undesirable habit . . . If we focus on the local scenario, Andi's actions seem unassailable; after all, he gives up when he knows it is futile. By contrast, Raphael's behaviour seems a little lamentable; after all, he keeps chasing when he knows he can't reach the ball. But if we shift to a more global outlook, we can see Andi's course of action as problematic and blameworthy, Raphael's as noble and praiseworthy. It is far from clear that one of these outlooks has any special authority over the (Hawthorne and Srinivasan 2013)".
26. Although such complications could also aid rivals to knowledge-based decision theory, I believe that knowledge-based decision theory makes better use of complications than its rivals do.
27. Provided that the worlds in which he does not kill an innocent person are worlds in which he does not violate any other deontological scruple.
28. It does not, however, cause the problem ascribed to it by Frank Jackson and Michael Smith. They claim that such an approach violates agglomeration, such that according to this approach it can be the case that an agent ought to  $\phi$  and ought to  $\psi$  despite it not being the case that the agent ought to  $\phi$ —and— $\psi$ . But this is not so. It's easy to see that any agent following standard decision theory will obey agglomeration (at least finite agglomeration). We can apply the standard Kratzer semantics for "ought", taking the agent's available acts as our modal base and taking the expected utilities of those available acts as our ordering source. The Kratzer semantics validates finite agglomeration, so since we can fit this approach into the Kratzer semantics this approach obeys finite agglomeration. Of course, Jackson and Smith might be thinking of some non-standard notion of "ought", but they give no indication that they have a Kratzer-violating notion of "ought" in mind.

29. More complicated worries about personal identity and self-locating belief might require even more formalism, but the framework of centered worlds can provide it.
30. I express an agent having a strict preference for lottery  $L$  over lottery  $M$  with  $L > M$ , and an agent being indifferent between lottery  $L$  and lottery  $M$  with  $L = M$ .
31. Note that I continue to set aside moral dilemmas, which in this case means that I only consider cases in which an agent has at least one available act which he knows to be deontologically licit. There are a variety of ways to handle moral dilemmas, the most obvious of which are (1.) for the agent to minimize the probability of violating a deontological scruple, and (2.) for the agent to maximize expected utility without regard to the risks of violating his deontological scruples. Approach (1.) obeys Independence, whereas approach (2.) does not.
32. The permissive knowledge norm would violate independence. The agent could risk nonzero probabilities of violating deontological scruples so long as the agent maintained a nonzero probability of not violating deontological scruples. The agent would not kill an innocent person to prevent the world from being destroyed, but could nonetheless prefer a 50% chance of killing an innocent person and a 50% chance of nothing happening to a 50% chance of the world being destroyed and a 50% chance of nothing happening. Knowledge-first decision theories do accord knowledge a special significance, but this result is unpalatable nonetheless. One more reason to prefer the restrictive knowledge norm to the permissive knowledge norm.
33. If utilities are unbounded and there are no restrictions on the composition of lotteries then the St. Petersburg lottery will lead to violations of continuity.
34. A permissive knowledge norm would also violate continuity, but for different reasons. The agent would not kill an innocent person to prevent the world from being destroyed, but could nonetheless prefer any chance less than 1 of killing an innocent person to a corresponding chance of the world being destroyed.
35. Alan Hájek puts things well: “[O]nce we get to probabilism, it provides us with many fruits. Above all, it forms the basis of a unified theory of decision and confirmation—it combines seamlessly with utility theory to provide a fully general theory of rational action, and it illuminates or even resolves various hoary paradoxes in confirmation theory. (Hájek 2008).”
36. For helpful feedback, I thank John Hawthorne, Adam Elga, Michael Smith, and Adam Lerner. This publication was made possible by support from ANU.

## References

- Aboodi, Ron, Borer, Adi, and Enoch, David, 2008. “Deontology, Individualism, and Uncertainty: A Reply to Jackson and Smith,” *Journal of Philosophy* 105: 259–72.
- Brown, Campbell. 2011. “Consequentialize This.” *Ethics* 121: 749–71.
- Chisholm, Roderick. 1942. “The Problem of the Speckled Hen.” *Mind* 51: 368–73.
- Hájek, Alan. 2008. “Arguments For—Or Against—Probabilism?” *British Journal for the Philosophy of Science* 59: 793–819.
- Harman, Gilbert. 1988. *Change in View: Principles of Reasoning*. Cambridge: MIT Press.

- Hawthorne, John and Srinivasan, Amia. 2013. "Disagreement Without Transparency: Some Bleak Thoughts," in *The Epistemology of Disagreement: New Essays* edited by David Christensen and Jennifer Lackey, 9–30. Oxford: Oxford University Press.
- Hintikka, Jaakko. 1962. *Knowledge and Belief: An Introduction to the Logic of the Two Notions*. Ithaca: Cornell University Press.
- Jackson, Frank and Smith, Michael. 2006. "Absolutist Moral Theories and Uncertainty." *The Journal of Philosophy* 103: 267–83.
- Kant, Immanuel. 1917. *Perpetual Peace: A Philosophical Essay*. London: George Allen and Unwin.
- Kratzer, Angelika. 1991. "Modality," in *Semantics: An International Handbook of Contemporary Research* edited by A. von Stechow and D. Wunderlich, 639–50. Berlin: Walter de Gruyter.
- Lewis, David. 1996. "Elusive Knowledge." *Australian Journal of Philosophy* 74: 549–67.
- Lewis, Clarence Irving. 1952. "The Given Element in Empirical Knowledge." *The Philosophical Review* 61: 168–75.
- Neumann, John von and Morgenstern, Oskar. 1953. *Theory of Games and Economic Behavior*. Princeton: Princeton University Press.
- Williamson, Timothy. 2002. *Knowledge and Its Limits*. Oxford: Oxford University Press.