KANTIAN DECISION MAKING UNDER UNCERTAINTY: DIGNITY, PRICE, AND CONSISTENCY

Adam Bjorndahl, Alex John London, Kevin J.S. Zollman

Carnegie Mellon University

Abstract

The idea that there is a fundamental difference in value between persons and things and that respecting this difference is an important moral requirement has strong intuitive appeal. Kantian ethics is unique in placing this requirement at the center of a moral system and for explicating the conditions for complying with it. Unlike challenges to Kantian ethics that focus on tragic cases that pit respect for one person against respect for another, this paper focuses on the question of how we can respect the value distinction between persons and things under conditions of uncertainty. After exploring why decision making under uncertainty is a neglected topic among Kantians and demonstrating how uncertainty challenges our ability to comply with this norm, we propose a notion of *morally insignificant risk* within a framework that allows agents to navigate real world decisions involving material benefit and some risk to dignity without violating the Kantian's core commitments. We conclude by exploring some of the challenges facing this approach.

1. Introduction

There is something compelling in the idea that there is a fundamental moral difference between persons and things, and that a significant portion of ethics is concerned with ensuring that our treatment of ourselves and others reflects this distinction. Kantian ethics places this intuitively appealing idea at the center of its moral system. A common challenge to Kantian frameworks concerns how they deal with situations in which all the choices open to an agent fail to respect the value of some individual.[1] There is, however, a different fundamental prob-

---

1. For instance, Marcus says, "But Kantian ethics is notably deficient in coping with dilemmas. Kant seems to claim that they don't really arise, and we are provided with no moral grounds for their resolution" (1980, pp. 125–126). In cases where each of the actions open to an agent will imperil the lives of different groups of people, Kantians face the challenge of how to decide which action to take without aggregating lives in the way a consequentialist would. Trolley problems present a similar challenge; the literature on doing/allowing, intending/foreseeing, killing/letting die, etc., represent strategies Kantians have

lem facing Kantian ethics that arises in far more mundane situations: how can the Kantian make decisions that do not equate the value of persons and things in circumstances of uncertainty?

Uncertainty is a ubiquitous feature of life. Decisions that are easy to make under certainty can become much more difficult and morally fraught under uncertainty. It is surprising, therefore, that the Kantian moral tradition has so little to say about decision making under uncertainty. This stands in sharp contrast to consequentialism, which has a long and robust tradition of theories encompassing uncertainty (von Neumann and Morgenstern, 1953; Savage, 1954).

This lack of development might be explained by two distinct but interrelated assumptions. First, the most detailed and mathematically sophisticated treatments of uncertainty have been developed under very different assumptions than those used in Kantian moral theories. Uncertainty is usually modeled as attaching to consequences.[2] Since Kantian moral theory does not evaluate actions in terms of their consequences, uncertainty might seem like an orthogonal concern. Second, the categorical nature of moral imperatives in the Kantian framework appears to offer a straightforward response to practical uncertainties. Because Kantian theories enjoin the agent to never perform certain types of acts, one might suppose this injunction extends easily to acts that involve uncertainty. If an act that has property x is morally prohibited, then it seems natural to extend this prohibition to those that *might* have property x. We contend, however, that both of these assumptions are misguided.[3]

––––––––––––––––––––

deployed in attempting to address such issues (Steinbock, 1980; Fischer and Ravizza, 1992).

2. For a discussion of the consequentialist commitments of standard decision theory see (Colyvan et al., 2010).

3. Jackson and Smith (2006) have identified a similar problem for absolutist theories, of which they take Kant to instantiate. They focus on absolute prohibitions, such as against killing innocent people, and argue that uncertainty can arise when the status of a person as innocent or not is unclear. They argue that absolutist theories cannot deal with uncertainty without compromising their absolutism. The model we present here is more general in that it not only applies to absolutist theories, but to a wider range of pluralist theories

Kantians have failed to appreciate how uncertainty applies to their own moral standards. Kantian decision makers, we argue, routinely have to make choices under uncertainty. But intuitive ways of extending Kantian decision making to cases of uncertainty are too extreme: they trivialize decision making if they prohibit all acts that might violate Kantian norms or they trivialize the Kantian's moral commitments if they permit any act that does not violate Kantian norms for certain. Nor can familiar techniques from expected utility theory be applied to resolve the problem; they violate the Kantian's moral commitments by forcing a comparison between the value of persons and things.

The fundamental challenge of decision making under uncertainty for Kantian theory, therefore, is to develop a framework that allows agents to meaningfully evaluate acts that risk violating dignity without making invidious comparisons between the value of persons and the value of things. We undertake the first steps in this process by articulating a notion of *morally insignificant risk* within a framework that allows agents to navigate real world decisions involving material benefit and some risk to dignity without violating the Kantian's core commitments.

We begin in Section 2 by identifying key features of Kantian moral theory, and then in Section 3 demonstrating the relevance and pervasiveness of decision making under uncertainty within this theory. In Section 4 we illustrate that there is not a simple, existing solution to the problem of uncertainty about an act's moral status. In Section 5 we develop our own model for Kantian decision making under uncertainty, and in Section 6 we defend this model. In Section 7 we explore a generalization of this model. Section 8 investigates the relationship between these theories and so-called "consistency" principles from decision theory. Section 9 concludes.

––––––––––––––––––––

in which certain trade offs in values may represent a moral breach. Additionally, although we identify a similar problem, our positive approach represents a unique attempt to incorporate uncertainty into a Kantian framework.

Although formal models of decision making have been almost the exclusive province of consequentialist moral theory, this paper demonstrates that the project of formalizing a Kantian decision theory can contribute significantly to this moral tradition and to mathematical models in decision theory. We hope the framework outlined here will provide a foundation for further exploration in this area.

2. Dignity, Price, and "Kantian" Theories

For our present purposes there are three definitive features of "Kantian" moral theories: (1) a claim about the nature and source of the value or status of persons that provides the operational content to (2) a fundamental moral norm and (3) a procedure for applying the norm directly to acts. Our claim is not that these features capture the comprehensive structure of Kant's moral theory. Instead, these are core features that some recognizably "Kantian" theories include. Explicating these features helps to illustrate why these theories have not seriously engaged with the impact of uncertainty on decision making.

The first defining feature of Kantian moral theory is a fundamental distinction between the value of "things" and "persons". We regard this as the most important insight that Kantians should want to preserve. In part, this is because the distinction provides the operational content to the theory's fundamental norm. But also, giving people a special moral status, and claiming that it is a transgression to treat a person like a thing or a tool, is powerfully attractive in its own right. We are interested in Kant because he unpacks and explains this fundamental distinction and places it at the center of a moral system in a way that serves as a common foundation for others to build upon in different ways.

There are three main characteristics Kant uses to distinguish the status of beings with dignity from the worth of things with price. The first we will term "replaceability". In his discussion of the kingdom of ends, Kant says:

In the kingdom of ends everything has either a *price* or a *dig-*

nity. What has a price can be replaced by something *else* as its *equivalent*; what on the other hand is raised above all price and therefore admits of no equivalent has a dignity (4:434-5).

A thing x might be replaceable in at least two senses. It is "token-replaceable" just in case there is another token of the same type that could be substituted for x without loss of value. For example, any new pair of Levi's 501 jeans could be replaced by any other new pair of Levi's 501s in the same size without a loss of value: they perform the same function or have the same use and fill this role in the same manner. A thing x is "equivalent-replaceable" if there is an amount of some other thing thing that can take the place of x without a loss of value. The jeans are equivalent-replaceable in the U.S. by some number of U.S. dollars. Although the dollars do not serve the same function as jeans, they have the same relative value or worth. For Kant, things can be both token-replaceable and equivalent-replaceable and this is the foundation upon which markets and the use of currency are based.

In contrast to things, Kant says that beings with dignity "admit of no equivalent." Presumably, this means one cannot simply replace one person with another without loss of value, even if the two people have the same skills and abilities. In other words, people are not token-replaceable. Similarly, when he argues that those with dignity are "above all price," he presumably means there is no quantity of things that could replace a person: people are not equivalent-replaceable.

The second distinction Kant draws we term "conditionality". Just after the passage quoted above, Kant says:

...that which constitutes the condition under which alone something can be an end in itself has not merely a relative worth, that is, a price, but an inner worth, that is, *dignity* (4:434-5).

The value of things with a price derives from, and is therefore conditional upon and relative to, the ends of agents. Moreover, this conditional relationship to people's ends or purposes helps to explain the replaceability of things. Two identical pairs of jeans are token-replaceable

because they meet the needs of the people who wear them equally well. They are equivalent-replaceable with some particular amount of money or goods of a different type because agents can weigh the relative worth of their various needs and wants and compare the resources required to address them.

By contrast, persons do not derive their value from the way they figure into the projects of other agents. Their value is unconditional and derives purely from their capacity for rational agency or "autonomy" (4:436). As Kant puts it, "...rational beings are called persons because their nature already marks them out as an end in itself, that is, as something that may not be used merely as a means, and hence so far limits all choice (and is an object of respect)" (4:428). A person's value does not spring from their relationship to the goals, desires, or ends of others. Rather, their status as an autonomous agent places a limit on the goals, ends, and desires of rational agents. "Dignity" is the term Kant uses to describe this unconditional worth that is grounded in autonomy.

The third distinction Kant draws we call "comparability". In order for two things to be equivalent-replaceable, it must be possible to determine their relative worth. Things are defined as those objects whose value can be represented in terms of equivalents. However, Kant asserts that dignity and price "cannot be brought into comparison or competition at all without, as it were, assaulting its [dignity's] holiness" (4:435). Here, Kant seems to claim that one cannot compare the value of a person to the value of a thing; by bringing them into "comparison and competition" we devalue or disrespect the status of a being with dignity.

It is not entirely clear what it takes to respect this last requirement. If the act of making *any* comparison between the value of persons and things is disrespectful, then Kant himself would be guilty of this transgression when he asserts that beings with dignity have a value that is "raised above all price." Asserting that the worth of persons is infinitely (or in any other sense) greater than that of things just is to make a comparison between these two forms of value. If there is no sense in which

things with a dignity can be compared to things with a price, dignity cannot be assigned a value higher than price.

If, as seems reasonable, the features of replaceability and conditionality explain the sense in which dignity is more important than price, then the requirement of non-comparability should be understood as consistent with these comparisons. An intuitive response is to say dignity's value is *lexicographically prior* to the value of anything with a price: considerations of price are only relevant when dignity is not meaningfully at stake.[4] This view has two virtues: it preserves the Kantian commitment to the special status of dignity while blocking any comparisons of value that would make dignity equivalent-replaceable with materials goods. As will become clear in the following discussion, a central challenge for Kantian moral theory is to provide a determinate account of this requirement that does not trivialize choice under uncertainty.

The second defining feature of Kantian moral theory is its standard of moral evaluation. Although Kant offers several formulations of this norm—or as Kant calls it, this "categorical imperative"—one formulation is grounded more closely in the distinction between dignity and price. As Kant puts it, the fundamental norm of morality is that rational agents must always regard moral agency "whether in [their] own person or in the person of any other, always at the same time as an end, never merely as a means" (4:429).

Each of the conditions discussed above—replaceability, conditionality, and comparability—helps to flesh out the operational content for the fundamental norm of Kantian ethics. These distinctions clarify what is required to treat someone as an end and never as a mere means. A struggle at the heart of Kantianism is to make operational what is required to treat someone as an end, namely, as a unique and

---

4. The term "lexicographic" evokes the usual way of ordering words alphabetically, which effectively gives earlier letters infinite priority over later letters. For instance, the word "azalea" is ordered ahead of "babe" because "a" comes before "b", and despite the fact that the relative ordering of the second letters is dramatically reversed.

autonomous agent at the center of their own life whose worth is not grounded in their relationship to the goals, desires, or ends of any other being.

The third defining feature of Kantian moral theory is that actions are evaluated *directly* by this fundamental moral norm. Whether an action is right or wrong, permissible or impermissible, is determined by whether it embodies or reflects an attitude of respect for the value of rational agents as ends in themselves.[5] Thus, recognizing the special status of other beings puts a constraint on which actions are permissible. This method of evaluation stands in contrast to *consequentialism*, according to which acts are evaluated solely on the basis of the outcomes they bring about. This is the sense in which Kantian theories are non-consequentialist.[6]

3.   The "Kantian" Evasion of Uncertainty

Having clarified what we mean by "Kantian" moral theory, we can see why proponents of such views might think uncertainty is not a salient issue for their approach to morality. Much of the literature on uncertainty focuses on the indeterminacy of the worldly effects that acts bring about; i.e., the uncertainty inherent in attaining specific outcomes. But uncertainty construed in this way seems a peripheral concern for Kantian theories, which as we have seen are concerned with features of acts themselves and not their consequences.

This distinction in evaluative methods is nicely illustrated by the method proposed by Sen (2004) for integrating Kantian concerns into formal decision theory. Sen describes a *permissibility function* that restricts the set of available acts, M, to only the "permissible" acts. That is, acts in M are first evaluated according to a norm K, and only those that pass this test remain on a "menu" of admissible acts, denoted $K(M)$, from which the agent may permissibly choose. For example, if $K(M)$ contains more than one act, then the agent might then choose the one from $K(M)$ that maximizes expected utility.

To illustrate a permissibility function, Sen describes the following scenario. "Friend" is attending a wedding reception at which there is a buffet that includes a fruit bowl containing mangoes, pears, bananas, and apples. Friend strictly prefers mangoes to pears, pears to bananas, and bananas to apples. But there is a norm of etiquette in this community: one ought not take the last piece of any kind of fruit. If Friend goes through the buffet line and encounters a basket of fruit containing several mangos, then she will choose a mango; on the other hand, if there is only one mango, and more than one pear, then she opts for a pear rather than taking the last mango. Friend uses the norm of etiquette to restrict the menu of admissible actions; "take the last mango" is prohibited from the outset.

Sen's insight is that the Kantian's moral requirement to respect the dignity of agents can be seen as a moral norm that functions to restrict the set of acts open to agents. Consider the following scenario, "Ex-

---

5.   Strict Kantian views hold that this feature is the sole determinant of the moral status of an act. But more permissive Kantian views might hold that this is a central or primary focus for the moral evaluation of an act but that other considerations may be relevant as well.

6.   For our purpose, any theory according to which the moral evaluation of an action is not determined solely by the goodness of the state of affairs that it brings about is non-consequentialist.

plicit Proposition", in which "Politician" offers aspiring "Reporter" the opportunity to break important news about evidence that "Dictator" has weapons of mass destruction. In this scenario, Politician makes explicit that, although the evidence is compelling, his administration knows Dictator does not have weapons of mass destruction. Nonetheless, publishing the evidence will tip public opinion in Politician's favor. Reporter can (a) not publish the evidence, foregoing any benefits of friendship with Politician, (b) publish the evidence, gaining the benefits of continued friendship with Politician, or (c) publish a story that aims to discredit the evidence so that Politician cannot use another media outlet to sway public opinion. We can take (b) to be incompatible with the relevant Kantian norm because it represents complicity with Politician's plan to disrespect others by manipulating their beliefs according to Politician's own goals. Thus, according to Sen, a Kantian Reporter does not attend to the outcomes from (b); in other words, a Kantian Reporter does not first calculate the utility to be gained from having close connections to Politician, and then weigh these benefits against the value of dignity. Instead, Reporter eliminates (b) from consideration at the outset, and decides between the remaining options (a) and (c)—perhaps on the basis of their respective consequences.

Sen's approach has many virtues. It retains the spirit of Kantian non-consequentialism by applying K directly to the acts in M. It treats dignity as being "above all price", in that considerations of dignity are dealt with before any other decision-making procedures come into play. This seems to avoid any kind of comparability between dignity and price that would violate the Kantian prohibition since there is no way for considerations of price to affect which acts are included in $K(M)$.

However, Sen's approach also illustrates the sense in which Kantians have avoided grappling with decision making under uncertainty. Specifically, it ignores the prospect of uncertainty regarding *which acts to remove* from M. It is simply assumed that there is no question as to which acts violate the norm K. If uncertainty arises at all in this setting, it does so only at the stage of decision making after the avail-

able acts have been reduced to those in $K(M)$. This framework thereby hard-codes the assumption that uncertainty only becomes relevant after Kantian considerations have been completely discharged.

Contrary to this assumption, uncertainty can be very relevant to the application of permissibility functions. In particular, it may be far from clear which acts represent affronts to dignity, and therefore which acts should be removed from M. To illustrate this general point, consider the "Uncertain Mango" scenario. A caterer from outside the community strives to provide the most high-end and exclusive service possible at weddings. In part, this involves displaying only a single instance of each available fruit atop its own fancy pillow on the catering line. Everyone at the party knows there is a large basket of mangos, pears, bananas, and apples somewhere out of sight, and as soon as a single fruit is selected it is replaced by another token of that type. At the beginning of the party, the probability that one is choosing the last mango is almost nil. But as the party progresses, the probability increases. How should the agent who wants to conform to the norm of not taking the last fruit of its kind apply the K function to the acts available to her?

A similar uncertainty can creep into Kantian decision making as well. We see two pathways for the introduction of such uncertainty. The first involves uncertainty about our own conduct and motives. For example, Kant says we like to flatter ourselves in thinking we are choosing acts from noble motives, while in reality we cannot be certain that there is not some "covert incentive" motivating our actions:

> In fact, it is absolutely impossible by means of experience to make out with complete certainty a single case in which the maxim of an action otherwise in conformity with duty rested simply on moral grounds and on the representation of one's duty. It is indeed sometimes the case that with the keenest self-examination we find nothing besides the moral ground of duty that could have been powerful enough to move us to this or that good action and to so great a sacrifice; but from this it cannot

be inferred with certainty that no covert impulse of self-love, under the mere pretense of that idea, was not actually the real determining cause of the will; for we like to flatter ourselves by falsely attributing to ourselves a nobler motive, whereas in fact we can never, even by the most strenuous self-examination, get entirely behind our covert incentives, since, when moral worth is at issue, what counts is not actions, which one sees, but those inner principles of actions that one does not see (4:407).

Kant was a prescient forerunner of what is now a long tradition in psychology and philosophy arguing that people do not have privileged access to their own thought processes (Ryle, 1949; Nisbett and Wilson, 1977). If our motivations are not transparent to us, then it is possible we misdiagnose our own reasons for action. In such cases, a Kantian may be forced to evaluate actions without being certain about whether they constitute the use of a person as a means to an end.

For illustration, consider the following scenario which we call "Graduate Advice". A college chemistry professor, "Prof", must advise her student, "Undergrad", on a choice between two different graduate schools. Undergrad has been extremely diligent in assisting Prof in her research. Prof very much enjoys working with Undergrad and would prefer—for selfish reasons—for the student to pursue graduate studies at Prof's institution. However, the student has also been admitted to another institution of similar quality. Prof is a Kantian and knows she should not recommend that Undergrad stay at her institution merely to satisfy her own desire for a good student. Prof thinks carefully about the student's interests, weighing the strengths and weaknesses of each institution, and finally she concludes he should stay at Prof's institution. She believes this is genuinely in the best interest of the student, but she also knows that it is possible she is deceiving herself.

The second way uncertainty creeps into Kantian ethical judgments stems from a much more mundane source. Even if we have privileged access to our own motives, we rarely have such access to the motivations of others. Many social interactions involve compliance with or the facilitation of others' wishes. The Kantian's fundamental moral norm enjoins against using the rational nature of others or of oneself as a mere means. So uncertainty about others' intentions matters when we have to decide whether to comply with, permit, or facilitate that conduct. In some cases, this will involve uncertainty about what one agent is doing to some other agent. But it can also involve uncertainty about how another is treating us. In either case, we might find ourselves in a situation where we are unsure about whether we or someone else is being used as a means or an end. Deciding whether to endorse or to oppose such conduct requires making a decision under uncertainty.

To illustrate this kind of uncertainty consider the "Uncertain Proposition" scenario. Shortly after arriving at a major news outlet, Reporter is invited to a secret meeting with Politician. There, Politician presents compelling evidence that Dictator possesses weapons of mass destruction and offers Reporter the chance to be the first to break the story. In this situation the reporter has a take-it-or-leave-it offer to go on air with the administration's evidence or else to forego the opportunity to be first with a big story. Although the evidence seems compelling to Reporter, he knows this information will sway public opinion in the administration's favor and it is possible that the evidence is intentionally misleading. Turning down the offer will not produce the benefits of friendship with an influential political figure, but Reporter is uncertain whether publishing the evidence represents complicity with an effort to manipulate the public. The uncertainty here stems from asymmetric information—Reporter doesn't know the true motives of Politician.

Reporter wants as many social and material advantages as possible, including the chance to rise in the ranks. But Reporter is also a Kantian and does not want to be complicit in the project of manipulating others for political gain. If Reporter knows that Politician is being duplicitous, Reporter ought not publish the story. But crucially, matters are not so clear.

It should by now be apparent that the Kantian must develop a theory capable of dealing with uncertainty regarding dignity. We turn now to an investigation of how such a theory might be incorporated

into the general framework of Kantian decision making.

4. Finding a Middle Path

There are two extreme strategies that a Kantian might use to handle cases of uncertainty. One strategy holds that whenever there is uncertainty regarding the true nature of the act one is choosing, one is not violating the fundamental Kantian imperative by performing the act. Call this the *permissive strategy*. This strategy seems like an intuitive response to Uncertain Mango. There is a chance that any given piece of fruit is the last of its kind, but uncertainty has a kind of absolving effect on guests—as long as it is not *certain* that the displayed fruit is the last of its kind, it is permissible to take it. Although it may be less intuitive, one might have a similar view of Graduate Advice and Uncertain Proposition.

The obvious problem with this approach is that it is too permissive. Imagine Reporter has a very strong suspicion that Politician's evidence is contrived. Although it is highly likely that Reporter is being asked to be complicit in manipulation, it remains *possible* that this is not the case. Ambiguity, after all, is a *feature* of many social interactions, providing a way to probe for amenability to certain conduct without having to come out and explicitly ask. Being explicit can be risky; by preserving plausible deniability, ambiguity can be socially useful—people with shared interests can find one another and those with divergent interests can discover this without (seriously) damaging their prior relationship. The permissive strategy effectively removes considerations of dignity from the equation altogether. Since most real-world social interactions involve at least some degree of uncertainty, the permissive strategy relegates the Kantian's core ethical requirements to the periphery of life, playing no practical role in the vast majority of decisions.

At the other extreme, one might argue that the categorical nature of the Kantian's moral requirements should be preserved even under uncertainty. This strategy holds that if there is *any* chance an offer represents an affront to dignity then it must be rejected. Call this the *restrictive strategy*. This strategy seems in line with the spirit of the

Kantian moral framework. It retains the lexicographic superiority of considerations of dignity: since dignity is above all price, any act that might compromise its moral sovereignty is unacceptable.

This approach is too restrictive. Applying the restrictive strategy in Graduate Advice yields the result that Prof cannot advise her student to remain at her home institution even if she judges the alternative university to be almost certainly worse, because it remains a possibility that Prof has deceived herself into underestimating the other program for her own ends. Similarly, because the ends of others are rarely as transparent as in the case of Explicit Proposition, the restrictive strategy would prohibit all acts in which there is any positive probability that the agent is complicit a project that is disrespectful of others. Uncertainty thus radically restricts the set of actions that the Kantian can regard as permissible (cf. Jackson and Smith, 2006).

In different ways, the permissive and restrictive strategies trivialize the notion of Kantian decision making under uncertainty. If we regard a more permissive attitude as reasonable in cases like Uncertain Mango and a more restrictive attitude as reasonable in cases like Uncertain Proposition, then we may be motivated to find a middle position in which it is permissible to choose acts that *might* violate dignity provided the risk is *morally insignificant*.

The challenge, then, is to define an appropriate notion of "morally insignificant risk". It is plausible that such a notion should be sensitive to a variety of features of the decision at hand: the severity of threats to dignity, the magnitude of the material gains, or the degree of uncertainty the agent is facing. Classical decision theory is designed to accommodate very similar concerns: balancing possible gains and losses in the face of varying degrees of uncertainty. So it is natural to wonder whether the machinery of expected utility theory (EUT) can be fruitfully imported into the present setting to produce our sought-after "middle ground".

EUT determines the value of an act by reference to the likelihood of that act bringing about certain states of affairs. Actions are distinguished from outcomes, with the former thought of as bringing about

the latter. Each outcome is assigned a numeric utility representing its value, and a given action is evaluated by considering the outcomes it might bring about, weighting each by its probability, and taking the sum of the resulting weighted utility values. This sum is referred to as the *expected utility* of the action. In this way, any action can be compared, numerically, to any other action: those actions with the highest expected utility are identified as *best*.

This template seems to fit the bill of consequentialism to a tee: the value of an action depends entirely on the value of the potential outcomes it produces. As such, it might appear impossible to use the EUT paradigm to model Kantian decision making on pain of violating the non-consequentialism of the Kantian program. However, this issue is not as deep as it may appear. As far as the mathematics of EUT is concerned, outcomes are simply states an agent values and that arise from the choice of particular actions; crucially, they need not be identified with material effects in the world, nor sharply distinguished from the acts that bring them about.[7] In particular, a Kantian who wishes to model acts as having some inherent value can simply take outcomes to encode the relevant properties of the acts they are evaluating. For instance, if lying to a person is a form of disrespectful treatment, one might view the act of "lying to Pam" as bringing about an outcome identified with "disrespecting Pam". Henceforth, we adopt this perspective and assume that outcomes encode affronts to dignity.

One might then try to formalize Kant's claim that dignity is "raised above all price" by assigning infinite negative utility to the loss of dignity, while keeping all material gains and losses finite.[8] This has the

obvious and desirable consequence that no monetary gain outweighs dignity. But in this case, no amount of money can justify accepting *any chance* of a losing one's dignity. To see this, consider an action that guarantees some fixed material gain with utility g, while incurring a small positive probability, say ε, that the agent will violate her own or another's dignity. The expected utility of this action is $g + \varepsilon(-\infty) = -\infty$. Thus, it is regarded as strictly worse than any action that is certain not to violate an agent's dignity, no matter how bad the material consequences may be. Maximizing expected utility in such a setting therefore generates a decision rule that agrees with the restrictive strategy discussed above, and it has the same ridiculous consequences.

Instead one might associate a *finite* utility value with dignity. One way of doing this is unsatisfactory for the Kantian. Suppose the utility of dignity is 100 utiles while the utility of $10 million is something more, say 101 utiles. This treats dignity as worth less than $10 million; in particular, any agent who maximizes expected utility would gladly accept $10 million in exchange for her dignity. This approach therefore violates the following principle:

> **Price-resistance under certainty:** The agent would not exchange her dignity for any material gain.[9]

---

7. Sen's (2009, pp. 215–221) distinction between "culmination" and "comprehensive" outcomes is relevant here. The former represent the consequences of an act in a way that is independent of the processes or actions that brought it about, whereas the latter can include features of these acts or processes. For instance, such features might include the fact that the act in question involved deception or some other form of disrespect.

8. Assigning positive or negative infinite utilities violates the "continuity" axiom common to many formulations of expected utility (von Neumann and Morgenstern, 1953).

9. This is the specification of the Kantian's attitude toward dignity used in the game theoretic treatment of Kantian ethics presented by Athanasiou, London, and Zollman (2015).

A more sophisticated application of EUT would assign dignity a utility that is an upper bound on the utility of material gains. For example, suppose no material gain, no matter how good, is assigned more than 100 utiles, whereas dignity is assigned, say, 120 utiles. In this case it is easy to see that the agent would never exchange any amount of money for her dignity.[10] Nonetheless, this setup seems inconsistent with the Kantian injunction to avoid equating the value of "things" and "persons", for although dignity has no direct monetary equivalent, it is assigned a concrete value that compares it with material goods. This emerges most clearly in the context of trades under uncertainty: while the agent would not accept any monetary payment in exchange for a *certain* loss of dignity, she might well accept a sufficiently large monetary payoff in exchange for, say, an 80% chance of losing her dignity.

In general, EUT renders all outcomes *comparable*: given any two outcomes, either one is strictly preferred to the other, or they are regarded as equivalent.[11] Even if this does not violate the Kantian restriction on comparability *per se*, EUT values outcomes in a way that is inherently relational. An outcome o assigned utility 3 is "worth" three times more than an outcome o' assigned utility 1, in the sense that an expected utility maximizer would view a $1/3$ chance of o as equivalent to receiving o' with certainty. So, while o is more valuable than o', there are situations where receiving o' with certainty is regarding as preferable to receiving o with some small probability. In this respect, the value of o' can be equated with some "amount" (here, probability) of o.

_____

10. Note that this requires the utility of material goods to have an upper bound, which is a restriction on the way that agents can care about material things. One can certainly make this assumption, though such a restriction is not without consequence.

11. This axiom is sometimes called the "ordering axiom". Decision theorists have investigated the consequences of abandoning this axiom and leaving some outcomes fundamentally incommensurable (see, e.g., Seidenfeld, 1988; Levi, 1990). Simply abandoning the ordering axiom would not be sufficient for the Kantian because it would not make dignity strictly more valuable than material goods—it would make them completely incomparable.

If o represents something pertaining to the dignity of a person, while o' represents a material good, this brings dignity into comparison with things in a way that the Kantian wants to reject. An expected utility maximizer can say plainly that the person is worth three times more than the material good. This allows a gamble on the former to be equated with the latter, which seems to violate the Kantian's restrictions on making dignity and price comparable and replaceable. This approach violates the following, informally stated principle:

> **Price-resistance under uncertainty (informal):** The agent does not assign a price to any chance that her dignity will be violated.

Below, we consider two ways of formalizing this principle, a weak version in Section 5 and a strong version in Section 6.

EUT provides a well-understood and mathematically robust operationalization of the notion of a morally insignificant risk. The problem is that it cannot avoid violating the Kantian's foundational moral commitments without trivializing decision making. Kantians need a decision theory that can cope with uncertainty about dignity without attaching to it a price, and EUT is not up to this task.

5. Threshold Significance (Θ_s) Rules

EUT runs afoul of core Kantian commitments because it determines when risks are morally insignificant by counterbalancing degree of likelihood with magnitude of material gain. As a result, the Kantian needs a conception of morally insignificant risk that decouples likelihood from material gain in some meaningful sense. The challenge is to do so without either trivializing all decisions or rendering the Kantian's core values irrelevant to choice.

As a first positive proposal, we present a family of decision rules that define a notion of morally insignificant risk to dignity entirely independently of the material gains at stake in a decision. We call these "threshold significance rules" because each one specifies a particular likelihood threshold below which risks to dignity can be ignored or

tolerated.[12] In this section we show that such rules are consistent with the Kantian's core commitments; in the following section we explore the extent to which they might be made sensitive to additional contextual factors without violating these commitments.

To begin, we must first formalize our focal decision problems. Each of the examples from Section 3 includes a choice between two actions: A and B, where action A provides a material gain relative to B but comes with a chance of damage to dignity, while action B is "safe"— dignity is not at stake. In Graduate Advice, action A is recommending that Undergrad stay at Prof's institution and B is advising Undergrad to go elsewhere. In Uncertain Mango, action A is taking a fruit from the pillow and B is not choosing a fruit at all. In Uncertain Proposition, action A is publishing the evidence of WMD and B is opting not to publish.

None of these decisions have the form of a "moral dilemma" where both options run a risk of treating a person improperly. We have focused instead on more mundane decisions where at least one option (option B) is clearly permissible from a Kantian perspective. In order to formulate a general version of such decision problems, we introduce the notation A_q^g for an act that guarantees the agent a material gain of

utility g,[13] but has probability q of transgressing an important norm, either by using another as a means to an end or by allowing oneself to be used as a means to the ends of another.[14] In the examples above, the decision maker faces a choice between actions A_q^g and B, where $g > 0$, $q > 0$, and B is assumed by contrast to yield no material payoff but also to carry no chance of damage to dignity. (Technically, then, $B = A_0^0$, but we will continue to refer to the "risky" choice and the "safe" choice using A and B, respectively.)

How should a Kantian decide in such contexts? As we have argued, to avoid triviality we want a decision rule that is sensitive to the degree of uncertainty pertaining to loss of dignity. EUT—perhaps the most natural candidate for encoding such a sensitivity—leads to problematic consequences no matter how we define the utility of dignity. The alternative we propose defines a "morally insignificant risk" by appeal to what we call a *significance threshold*, which is just a real number $s \in [0, 1]$. Intuitively, s represents the maximum value of q the decision maker will tolerate—the largest probability that someone's dignity is at stake that does not render the corresponding action inadmissible. As we will see, this simple idea allows us to represent both the permissive and the restrictive strategies and much else besides.

Formally, for each $s \in [0, 1]$, we specify the *threshold significance rule corresponding to s*, or the "Θ_s rule" for short, as follows: given actions A_q^g and B (as above) with $g > 0$, if $q \leq s$, then A_q^g is preferred to B, and if $q > s$, then B is preferred to A_q^g.

---

12. The idea that some values have a threshold is common in non-consequentialist theories. Kagan (1998) presents a nice review of the topic. Jackson and Smith (2006) note that Kantians may want to adopt a threshold view to deal with cases where there is uncertainty about whether an act violates an absolute value. They reject this view because they claim that it creates a dilemma for the absolutist (see Section 7 below). The threshold view is defended against this criticism briefly by Hawley (2008) and in more detail by Aboodi, Borer, and Enoch (2008). Because these authors do not characterize some relevant mathematical properties of the views they propose, it is not clear whether they are discussing the family of rules we describe. This is important because, as we explain below, implementing threshold rules in the context that we have identified poses a special moral challenge since specifying such a threshold can result in a violation of the Kantian's own commitment not to compare the value of things and dignity.

13. We are assuming that the agent is a utilitarian with respect to material gains—that is, there is a single real number that can be assigned to every potential material gain. While not a critical assumption for the argument that follows, it greatly simplifies the discussion.

14. The probability q is either a well-defined objective probability or a point-valued subjective probability that the agent assigns to this property of the action. If one wanted to adopt another theory of uncertainty—e.g., set-valued probabilities—one could still formulate a version of the strategy that we present, though the additional complexities of such systems would have to be dealt with accordingly.

Threshold significance rules can be viewed as identifying a certain range of "low" probability values with 0. The Θ_s rule stipulates that uncertainty as it pertains to dignity can be ignored unless the chance of affront is greater than s. Said differently: all probability values $q \leq s$ are treated as being sufficiently small as to be effectively equal to 0. On this way of thinking, two special cases represent the two most extreme interpretations of "low" probability: Θ_0 identifies nothing with 0 except 0 itself, whereas Θ_1 identifies everything with 0, and so trivializes the role of dignity in decision making.

More precisely, when $s = 0$ we obtain the rule Θ_0, which by definition says A_q^g is preferred to B if and only if $q = 0$. Notice this is exactly the restrictive strategy, which requires that the agent reject precisely those acts that carry a non-zero probability of abrogating the value of dignity. As discussed above, this in turn is equivalent to the output of EUT when loss of dignity is assigned infinite negative utility. In each of these cases, regardless of the magnitude of the material payoff $g > 0$, the agent chooses the "safe" action B over any A_q^g with $q > 0$.

On the other hand, when $s = 1$ we obtain the rule Θ_1, which says that A_q^g is preferred to B if and only if $q \leq 1$, which is always true.[15] Thus, this rule accepts all potential violations of dignity—even certain ones. This nullifies whatever Kantian commitments an agent may have: Kantian values play no role in restricting the set of permissible acts.

Values of s that fall somewhere strictly in between 0 and 1 are our real focus; these have the potential to define a notion of morally insignificant risk that remains sensitive to uncertainty without descending into triviality. To illustrate how such rules function, let us consider their application in two of the examples discussed above. Imagine, for concreteness, that a 20% chance of violating dignity is regarded as sufficiently small that it can be ignored ($s = .2$). In Graduate Advice, Prof must assess the likelihood that her judgment as to the appropriate

graduate school for Undergrad is self-serving rather than based on his best interests. She might do so based on past experience, through evaluating the behavior of others, etc. Regardless, her ultimate decision is based on a comparison between this probability that her judgment is self-serving, q, and the threshold of 20%. If q is greater than 20%, she advises her student to attend the other school. Otherwise, she advises him to remain at her university.

Uncertain Proposition follows the same pattern. Reporter must determine the chance that Politician is presenting intentionally misleading evidence. Here Reporter might rely on prior experience with politicians in general, previous dealings with this particular politician, and any other information at his disposal. If Reporter finds that the probability of deception is higher than the threshold s, then he refuses to publish the evidence. Otherwise, he publishes.

These examples demonstrate how Θ_s rules are flexible in determining the impact of uncertainty on Kantian decision making. We consider the question of where the threshold s comes from in the following section. But first we address an even more basic concern: do Θ_s rules violate the Kantian injunction not to place a price on dignity?

It might appear that they do. Assume $0 < s < 1$. Consider, for the moment, a circumscribed set of potential action A's. Specifically, suppose we hold fixed a material gain $g > 0$ and consider the corresponding family of actions that vary only with respect to the probability of a dignity violation; denote this family by $A^g = \{A_p^g : p \in [0,1]\}$. Now we ask an agent who employs decision rule Θ_s to tell us which members of A^g are regarded as superior to the "safe" alternative B, and which are inferior. The set $S = \{A_p^g \in A^g : p \leq s\}$ contains those actions in A^g that are preferable for our agent to B, while $I = \{A_p^g \in A^g : p > s\}$ contains those actions not preferred to B. An adherent to the expected utility school of decisionmaking might then observe that this is consistent with assigning the price g/s to one's dignity.[16]

---

15. We can approximate the permissive strategy using threshold significance rules by taking the limit as s approaches 1. More formally, we can define a new rule according to which action B is preferred if and only if B is preferred for all threshold significance rules with $s < 1$.

---

16. When $p = s$ the expected utility maximizer is indifferent between A and

To see why, consider an expected utility maximizer who assigns utility g/s to her dignity. Action B always preserves dignity, so yields a payoff of $u(B) = g/s$. Action A_p^g yields an expected payoff of $u(A_p^g) = g + (1-p)g/s$. We can therefore calculate that $u(A_p^g) > u(B)$ if and only if $p < s$. So an expected utility maximizer who assigns utility g/s to her dignity and a Kantian who uses s would produce the same judgments about action B as compared to members of the set A^g. In this limited context, then, it might appear that an agent who employs Θ_s violates the injunction against assigning a price to her dignity because her decision rule is equivalent to one that puts a price on dignity. Although the agent may not conceptualize her own actions this way, she nonetheless acts as though dignity has a determinate price; she may not be aware of it, but her choices reveal the price she assigns to dignity.

This equivalence would seriously undermine the ability of Θ_s rules to represent Kantian values. But the equivalence presented above is illusory; it only holds when we artificially restrict our attention to actions with a common material gain. To see this, consider an expanded family of action A's that includes two distinct material gains: let $A^{g,g'} = \{A_p^h : h \in \{g, g'\} \ \& \ p \in [0,1]\}$, where $g \neq g'$ and $g, g' > 0$. In this case, as before, the actions that are superior to B according to Θ_s are those contained in $S' = \{A_p^g \in A^{g,g'} : p \leq s\}$, while $I' = \{A_p^g \in A^{g,g'} : p > s\}$ contains those actions deemed inferior to B by the rule Θ_s. But there is no longer a single utility value for dignity from which these two sets can be constructed: considering A^g implies the utility of dignity should be g/s, while focusing on $A^{g'}$ results in g'/s as the utility for dignity. Since $g \neq g'$, this is impossible.

We have therefore shown that an agent who uses a Θ_s rule does not, in general, make choices consistent with assigning a determinate

---

B, so to get exactly the same superior/inferior division we must assume that when she is indifferent she chooses action *A*. Since indifference is compatible with either choice, this seems an acceptable assumption.

price to dignity. It is worth restating this principle of Kantian decision making, introduced informally in Section 4, in more formal terms:

> **Weak price-resistance under uncertainty:** There is no value $v \in \mathbb{R}$ such that all the decision maker's choices are consistent with an expected utility maximizer who assigns utility v to dignity.

Θ_s rules satisfy weak price-resistance under uncertainty, but they are not unique in this respect. The restrictive strategy described above satisfies weak price resistance because it behaves as if it assigns utility ∞ to dignity and ∞ is not a member of the set of real numbers, $\mathbb{R}$. (One could easily expand weak price-resistance to eliminate this possibility as well.) For more complicated reasons, the permissive strategy also satisfies weak price resistance because it switches from assigning dignity a price of zero when dealing with cases of uncertainty to assigning it a utility of ∞ in cases of certainty. There are many other rules which satisfy this constraint as well. Some are discussed in Section 7, but before we turn to them we will discuss a number of issues that a Kantian must address if she would utilize such a rule in her ethical theory.

6. Defending Threshold Significance Rules

Threshold significance rules with $0 < s < 1$ have some attractive characteristics for a Kantian. First, unlike lexicographic approaches, they do not prohibit all choices that incur some positive probability of damage to dignity. Second, unlike views that take uncertainty to have an absolving effect, they allow the agent's core Kantian commitments to have concrete influence on decision making. Third, they respect the injunction not to assign a price to the value of persons by decoupling probabilities from the prospect for material gain in a way that is clear and intuitive. They therefore represent an important contribution to a neglected but nevertheless critical project, namely, understanding in mathematically precise terms how it is possible to respect core Kantian commitments in the face of uncertainty.

This approach also has interesting connections to other areas of work on human decision making. In particular, there is psychological

evidence that people often think of probabilities in relatively coarse-grained terms and they may treat a range of very low but positive probabilities as though they are not realistic possibilities (see, e.g., Kunreuther et al., 2001). This suggests that a normative framework that treats some positive moral risks as sufficiently small that they can be permitted or tolerated may have an attractive psychological realism that merits further exploration. Moreover, because boundedly rational agents are often constrained to think of probabilities in fairly coarse-grained terms, the framework for decision making described here has potentially interesting connections to work in bounded rationality as well.

One might be concerned about how agents determine the probability that dignity is at stake (the term q in our illustrative examples). Specifically, one might argue that allowing an agent to rely on a subjective estimate of the probability of dignity violation makes the Kantian theory arbitrary in an unacceptable way, since it is entirely up to the agent to form this estimate.

Any ethical theory in which agents have to make probability assessments faces the challenge of delineating the criteria that should be used to determine whether an agent should be blamed for relying on a particular probability assessment. For example, within EUT an agent could be blamed for relying on an assessment of the probability of particular outcomes that is not adequately grounded in available evidence. The relevant question is how a given framework determines whether an agent should be blamed for relying on a particular probability assessment. In the case of EUT, we might determine whether the agent did due diligence by asking whether the cost of gathering additional information would have been reasonable in light of the stakes at play in a decision.

Since the framework we propose also relies on probability assessments, it also faces this issue. For example, even if Reporter judges the probability that Politician is manipulating the public to be sufficiently small that it can be ignored, it does not follow that Reporter cannot be blamed for publishing the evidence. This depends on whether Re-

porter did due diligence in making their assessment. However, it is less clear how the Kantian should make such assessments since the Kantian must be careful not to violate her own values by allowing the material gains or losses at stake in a decision to determine which risks to dignity she regards as sufficiently small that they can be ignored. This is a topic for future work. For now we simply note that this is a general problem for any Kantian theory that permits decision making under uncertainty—it is not unique to the proposal we offer here. Different Kantian theories may legitimately take different approaches to this issue.

A different but related worry might be raised about the choice of a particular value for s. What is the moral ground for taking a particular probability s as being sufficiently small that it can ignored? Should s be viewed as a subjective threshold determined by each individual, or as an objective feature of proper ethical conduct? The answers to these questions will determine the degree to which the value of s is open to interpersonal evaluation and criticism. This is important because it seems reasonable that an agent could be held morally culpable for being mistaken about, or not having done due diligence with respect to, her choice of s. Without a principled rationale for setting the significance threshold at some particular value, this approach is vulnerable to the charge that it is unacceptably arbitrary (Ellis, 1992; Alexander, 2000; Jackson and Smith, 2006).

Someone who is worried about these concerns might argue that it is arbitrary whether s is set to .3 as opposed to, e.g., .301 or .289, which leaves the Kantian standpoint unmotivated in a certain sense. This problem is not unique to Θ_s rules, however. It applies to any quantitative theory of choice that relies on more than ordinal information. In particular, how finely a utility function is specified in EUT will be subject to a similar charge of arbitrariness. For instance, if the utility of o is 3 and the utility of o' is 1, one might question why o was not instead assigned utility 3.01, or o' utility 0.98. In either case, the richness of quantitative systems renders this sort of "measure arbitrariness" ineliminable.

A more substantive challenge is that the framework presented here requires the agent not only to adopt some particular value for what ought to be regarded as a morally insignificant risk, but to use that same value for all decisions. In the next section we explore to what extent a Kantian can vary the value of s across decisions while remaining true to core Kantian commitments.

While we explore several options, we do not wish to take a principled stand on many of these issues. It is likely that multiple accounts of these philosophical questions are possible and we think it important not to prejudice the discussion. In particular, we are exploring how a family of theories—recognizably Kantian theories—might deal with a common problem. Because these theories differ in their foundational moral commitments, it is possible that they would diverge radically in the way they justify key features of the decision framework presented here. For example, particularists or intuitionists might argue that agents can be faulted for their judgments about what risks can be ignored, but that no general formula for making such assessments in different situations can be stated.[17] Perhaps constructivists or others would try to provide a more systematic account of the considerations that are relevant to assessing how such thresholds are determined (Brennan, 1995).

7. Variable Thresholds

Under what conditions might a Kantian wish to chose different values for s in different contexts? One might vary s for at least three reasons. First, one might vary s based on features of the decision problem pertaining to evidence. For example, the agent might adopt a value for s that depends on how confident she is in her assessment of the probability—for instance, perhaps she has conducted extensive research or she is generally very knowledgable about the situation in question.[18]

Second, one might vary s in order to respond to distinctions in the degree of harm to dignity. Kant appears to treat all violations of the categorical imperative equally. But this is a point where many Kantians have parted ways with their philosophical progenitor. If some violations of dignity are more severe than others, we might think it reasonable for someone in the situation of Prof in Graduate Advice to have a different conception of what counts as a morally insignificant risk to dignity than she would in the situation of Reporter in Uncertain Proposition.

We see no reason to reject off-hand allowing the significance threshold to vary in either of these ways. Particular Kantian theories will need to explore the degree to which this variation is permissible, and we expect that this will differ in alternative Kantian schemes.

The final way that s might vary presents a general and difficult problem: one might vary s with respect to the potential gain offered from assuming a risk to dignity. In the case of certainty, some "moderate" deontologists argue that if enough good is at stake, a constraint can be overridden, making it permissible to act in a way that would otherwise be impermissible.[19] Allowing s to vary with respect to the

---

17. This might follow the form that Ross outlines in his discussion of how conflicts of duty are resolved: "When I am in a situation, as perhaps I always am, in which more than one of these *prima facie* duties is incumbent on me, what I have to do is to study the situation as fully as I can until I form the considered opinion (it is never more) that in the circumstances one of them is more incumbent than any other; then I am bound to think that to do this *prima facie* duty is my duty *sans phrase* in the situation" (1930, p. 19).

18. We thank an anonymous reviewer for this suggestion.
19. Consider the following from Kagan: "Many other people, however, find this absolutist attitude toward the constraint unacceptable. They believe that the constraint against doing harm can itself be outweighed, if enough is at stake. Presumably, killing an innocent person is morally forbidden even if this is the only way to save five, ten, or maybe even a hundred or a thousand people—but

- 15 -

material gain g would represent a natural extension of this view to cases of uncertainty.

But Kantians must be cautious in determining how s could vary with respect to price. Imagine that someone in the situation of Prof tolerates more uncertainty (i.e., use a larger threshold s) when Student is truly exceptional than when Student is merely one among several bright and helpful students. This version of Prof would be more permissive in situations where she stands to gain more and more restrictive in situations that would have only a small impact on her professional success. Does this way of varying the significance threshold violate the Kantian's core commitments?

Kantians are likely to object to the very idea of allowing the significance threshold s to depend on the fiduciary stake, g, precisely because, on this approach, dignity and material gain no longer occupy completely separate, independent domains. Jackson and Smith, for example, argue that any variation with the size of the material gain is inconsistent with Kantianism because it "gives [the material gain] the kind of central role that absolutism precisely rejects" (2006, p. 276).

However, the principle of weak price-resistance does not prevent Prof from varying her threshold as a function of material payoffs. It merely places some (very weak) constraints on how s might be responsive to material payoffs. To see this, we define a generalization of threshold significance rules as follows. Consider a function $\Phi : \mathbb{R} \to [0,1]$; think of Φ as associating with each material payoff g a threshold $\Phi(g)$ to be used for decision making when g is the relevant price. More precisely, define the Φ rule as follows: given actions A_q^g and B (as before) with $g > 0$, choose between them according to the threshold significance rule $\Theta_{\Phi(g)}$. Thus, the Φ rule applies a (potentially different) Θ_s rule to each decision, depending on the utility of the material gain associated with action A.

It is easy to see that some functions produce rules that violate Kant's injunction. Consider, for example, the rule given by $\Phi(x) = x/10$.[20]

This produces a threshold that varies linearly with the material payoff. The set of all actions superior to B is just $S = \{ A_p^g : p \leq g/10 \}$. This is equivalent to an expected utility maximizer who assigns utility 10 to their dignity. To see this, observe that for such an agent, the value of action B is 10—it preserves her dignity with no monetary reward—whereas the value of action A_p^g is $g + (1 - p)10$, which is greater than 10 whenever $g/10 > p$. An agent using this Φ rule would therefore violate weak price-resistance: though they may not realize it, they are acting exactly as though they assign dignity a determinate price.

The function Φ considered above is special: it is not hard to see that a Φ rule generates decisions consistent with an expected utility maximizer who assigns utility $v > 0$ to dignity if and only if $\Phi(x) = x/v$. Is this a problem for functions not of this form? Consider, for example, the function Φ illustrated in Figure 1,[21] which is nonlinear in g. The three points marked a, a', and a'' represent three variations on action A; for example, point a corresponds to something like $A_{.4}^{.25}$, while point a' corresponds to something closer to $A_{.6}^3$. Based on the thresholds determined by Φ, it is easy to see from the graph that a and a'' are both regarded as worse than B, while a' is regarded as better. Yet such preferences are inconsistent with EUT: a' lies half-way on a line from a to a'', so the utility of a' must be average of the utility of a and a''. Thus, if a and a'' are both regarded as inferior to B, an expected utility maximizer must also regard a' as inferior to B.

---

at *some point*, when the amount of good that needs to be done is great enough, the constraint is overridden, and it is morally permissible to act" (1998, p. 79).

20. Here and subsequently, for convenience, we define functions using algebraic expressions that can produce values outside the unit interval; implicitly, all outputs greater than 1 are identified with 1, and all outputs less than 0 are identified with 0. (Formally, all such functions are to be composed with the map $y \mapsto \min\{\max\{y, 0\}, 1\}$.)

21. This example uses an everywhere convex function, but a similar argument would demonstrate the incompatibility of a function that is concave or convex on any interval.
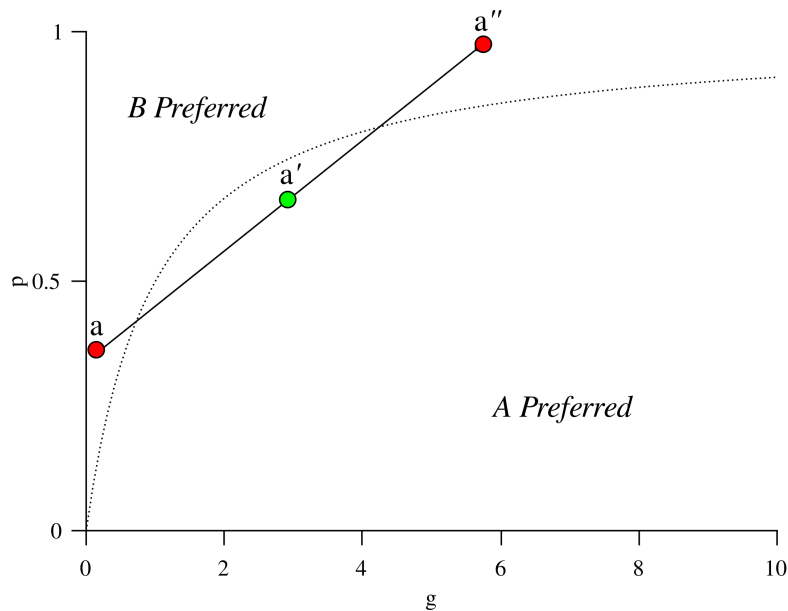
Figure 1: The graph of a nonlinear Φ, $\Phi(x) = 1 - 1/(x+1)$. a and a'' are regarded as inferior to B, while a' is regarded as superior to B.

This basic argument produces a large class of Φ rules that satisfy weak price-resistance under uncertainty; indeed, any Φ not of the form x/v generates such a rule, because there will be at least two points that appear—from the utilitarian perspective—to assign different values to dignity.

A Kantian might well be wary of endorsing weak price-resistance. Weak price-resistance only requires an inconsistency with utilitarian judgments *on some decision*: one can be "almost-utilitarian" and still satisfy it. Viewed from another perspective: it is possible to approximate a function of the form x/v arbitrarily closely while not being strictly of that form, and so satisfy weak price-resistance in point of fact while violating the spirit of the Kantian norm. Thus, if we precisify Kant's in-

junctions using only the principle weak price-resistance, we effectively allow Kantian decision making frameworks that are almost identical to utilitarianism.

More generally, while rules that satisfy weak price-resistance do not *equate* dignity with a utility value, there is nonetheless a sense in which they bring dignity into direct comparison with material goods. Consider, for instance, a pair of actions A_q^g and $A_q^{g'}$ with $g' > g$, and suppose $\Phi(g) < q$ and $\Phi(g') > q$. Then the corresponding Φ rule considers B preferable to A_q^g but $A_q^{g'}$ preferable to B. In other words, though the risk to dignity, q, is the same in both cases, the larger material payoff g' attached to the latter action is enough to justify taking this risk, while the smaller payoff g is not. Dignity, in this case, competes directly with material price: A_q^g is taken to incur a risk significant enough to reject it in favor of B, but once the pot is sweetened from g to g', the very same risk to dignity suddenly becomes acceptable.

These considerations suggest that weak price-resistance might be too weak of a principle for interpreting Kant's injunction. A stronger notion of price-resistance is needed; in order to capture this, we introduce the following strengthened principle:

Strong price-resistance under uncertainty: If q is considered an unacceptable risk to dignity in exchange for some material payoff $g > 0$, then q is considered an unacceptable risk to dignity in exchange for *any* material payoff $g' > 0$.

An expected utility maximizer who assigns a determinate value to dignity violates this principle (for a sufficiently small chance of damage to dignity, there will be material payoffs both greater than and less than the expected value of this damage). Moderate deontologists will likewise violate this injunction, since they allow for violations of dignity when the stakes are sufficiently high.

It follows immediately that strong price-resistance under uncertainty implies weak price-resistance under uncertainty. Moreover, any non-constant function Φ generates a rule that violates strong price-resistance under uncertainty: if $\Phi(g) \neq \Phi(g')$, then any q in between

these two values produces a violation. Thus, the only Φ rules that can satisfy this stronger principle are those where Φ is constant, which are precisely the Θ_s rules.

We now have on the table two variations on the Kantian injunction against putting a price on one's dignity. Weak price-resistance represents, in our mind, the weakest formal principle that captures Kantian core principles. But this may be too weak for many Kantians since decision makers who satisfy this rule can come arbitrarily close to expected utility maximizers who assign a price to their dignity. In contrast, the stronger principle is likely to be endorsed by some Kantians as providing the correct operational content to the restriction on bringing price and dignity into comparison.

There is logical space between the weak and strong principles. For example, in his discussion of assigning a monetary value to one's own life, Bergstrom (1982) proposes a rule (which would translate into a Φ rule in our terminology) that combines some features of the threshold significance rule with the Φ rule. Consider the Φ rule pictured in Figure 2. If a Kantian utilized this rule, she would reject any risk to her dignity larger than s regardless of the potential material gain. But for gambles that incur a risk lower than s, her preference depends on the amount of money on offer. While this rule violates strong price-resistance, it constitutes a potential compromise for those who may view strong price-resistance as too strong, while also viewing weak price-resistance as too weak.

8. Consistency and Value

The decision rules we have introduced face additional challenges that some may find concerning. In particular, there is a family of problems relating to consistency of choice that can only be jointly avoided by expected utility maximizers. Any theory that deviates from EUT will face at least one such problem.

Consider an agent who adopts the $\Theta_{.2}$ rule: she rejects any action that has a larger than .2 probability of violating dignity. Suppose she is asked if, hypothetically, she would take an action C that entails a
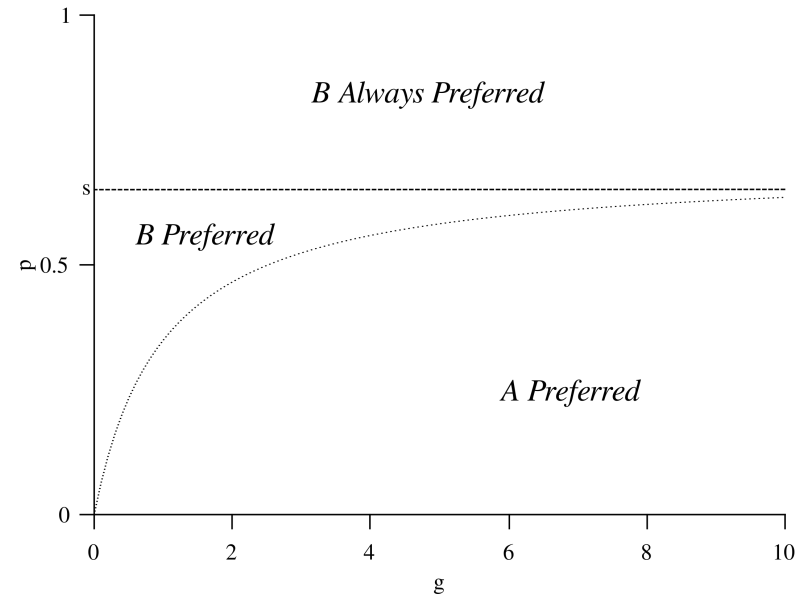

Figure 2: An illustration of a proposal from Bergstrom (1982).

.15 probability of damage to dignity but comes with a financial gain of $100. By assumption, she considers this risk worth taking; if she was offered this option, she would accept it over the status quo. Now someone else asks the same agent about a second hypothetical scenario (action D) which has a .1 probability of damage to dignity and comes with a financial gain of $110. Once again, this action is judged to be acceptable. But suppose these options had instead been offered as a package, a single action incurring a .25 chance of damage to dignity and a monetary payoff of $210. This action she must reject.[22]

---

22. For the purpose of illustration, we assume that the risks to dignity inherent in actions C and D are mutually exclusive—if dignity is violated by C it will not be by D and vice-versa. Had the risks to dignity been probabilistically independent then the probability of harm to dignity would be .235 and the

This rule therefore violates an amalgamation property commonly referred to as the "package principle". That is, "action C is morally acceptable" and "action D is morally acceptable" does not entail "actions C and D together is morally acceptable."[23]

Instead of thinking of this objection as involving a type of "packaging", one can instead view it as raising an issue regarding how decisions are made over time. Consider a Kantian who has adopted the $\Theta_{.2}$ rule, and suppose she is first offered the option to take action C, which she accepts. Then, prior to discovering whether dignity was, in fact, violated, she is offered action D. What does she do? Plausibly, she rejects this option because, when combined with the gamble she has already taken, it yields an unacceptable risk to dignity. Or perhaps, if our Kantian agent knows from the beginning that she would eventually be given the opportunity to accept D, she would refuse action C so she could later take D (after all, D is strictly better than C). In such cases, the agent's behavior seems inconsistent with her stated preferences, since she refuses gambles that, taken independently, are morally acceptable.

A related problem arises for our Φ decision rules. Consider again the Φ rule pictured in Figure 1. Here action a' is morally acceptable

while actions a and a'' are not. Suppose our agent is offered the following "compound" gamble: a fair coin will be flipped. If the coin comes up heads, the agent will be given action a. If the coin is tails, the agent will receive a''. Notice that this compound gamble is just action a'. One can take two morally unacceptable risks and combine them in a certain way to yield a morally acceptable act. This is a different way of "packaging" two actions than that considered above, but again it reveals an odd consequence: it runs afoul of one of the standard axioms of utilitarian decision theory—the independence axiom (von Neumann and Morgenstern, 1953).

As before, this problem creates odd consequences for choice over time. Suppose our agent is offered this compound gamble (action a') first. And suppose she will pay some small fee to acquire it. Then the coin is flipped, and she is told the outcome. She now knows whether she possesses gamble a or a''. Both gambles are bad in her mind; moreover, *regardless of the outcome of the coin toss*, she will now be willing to pay a fee for it to be taken off her hands. So our agent ends up where she started no matter the outcome of the coin flip, paying a fee twice for no return. This is sometimes called the money pump argument for expected utility theory (Davidson et al., 1955).

It is worth emphasizing that these problems are not unique to the framework that we present here: *any* Kantian theory will run afoul of at least some of these problems because only expected utility maximizers can avoid them all.[24] As a result, Kantians face a dilemma: they must either abandon their core commitment to the incomparability of dignity and price or they must accept one or more of the consequences

---

agent would also reject the action.

23. Jackson and Smith (2006) raise this objection in the context of absolutist decision rules where the action under consideration is morally obligatory rather than merely permissible. Aboodi, Borer, and Enoch (2008) argue this problem doesn't constitute an objection to deontological constraints that are grounded in the moral status of persons. The reason is, roughly, that a rule that permits/prohibits act C and also permits/prohibits act D might have nothing to say about the act "C and D" if there is no individual agent whose status is at stake in this compound act. One concern about this response is that it is not sufficiently general—it is tied closely to the specifics of Jackson and Smith's original example. In that example, act C affects one agent while act D affects another. But in cases where C and D affect the same agent, this response is not available. Our analysis is more general. In this section we explain why Kantian theories are vulnerable to this and possibly other problems regarding consistency in choice—namely, the desire to keep dignity and price separate prevents the Kantian from satisfying the axioms of EUT.

24. It is possible that there are alternative approaches to adapting Kantian values to decision making under uncertainty. But to respect the difference in value between persons and things they must diverge from EUT. As a result, they too will be vulnerable to the problems discussed in this section. Different approaches may have more or less attractive resources for managing these problems, but such an analysis requires not just the elaboration of competing frameworks but an exploration of the strategies they might adopt for promoting consistency without violating Kantian values. We thank an anonymous referee for raising this concern.

above. Since the first horn of the dilemma would involve renouncing commitments that help to define Kantian ethics, we expect that Kantians will want to explore the consequences of the second horn.

Kantians are not alone in facing these problems. Many others—often for reasons entirely different from the Kantian—have questioned one or more requirements of EUT and have therefore confronted similar problems. For example, Schick (1986) argues that the package principle is not a constraint of rationality. Several alternative decision theories have been developed that violate the independence axiom (e.g., Seidenfeld, 1988; Buchak, 2013). And many authors, beginning with Strotz (1955), have defended potential solutions to problems that arise with decision-making over time analogous to the two we discuss.

Our point is not to endorse any of these strategies, but simply to note that Kantians need to explore the relative merits of these various strategies if they intend to confront the second horn of this dilemma. Because uncertainty is a pervasive feature of life, finding practical strategies to avoid the prospect of moral inconsistency will have substantial implications for how Kantians can reason, plan, and approach practical problems. For example, Kantians will be most vulnerable to inconsistency if they make decisions that involve some risk to dignity on a case-by-case basis. To minimize this prospect, they may have to think of how any particular decision fits into a larger series or "portfolio" of decisions. We are not aware of any literature that explores the implications of taking such an approach to moral decisions. And while Kantians may seek assistance from the proponents of some of the weaker conceptions of rationality that have had to confront similar problems, the latter group may find new and potentially compelling motivation for their approaches in the concerns of the Kantian. Whether such approaches can succeed remains to be seen. But one should not immediately assume that the dilemma posed here must be resolved to the detriment of the Kantian's core values.

9. Conclusion

It is intuitively and morally compelling that there is a fundamental moral difference between persons and things. So is the correlative idea that a significant portion of ethics involves ensuring that our dealings with persons respect this difference in value. Kantian ethics uniquely places these ideas at the center of a moral system in which the requirements of replaceability, conditionality, and comparability provide intuitive benchmarks for distinguishing the value of agents as ends in themselves from the value of things. Despite these unique virtues, however, the Kantian tradition has not paid significant attention to the difficulties that arise for agents attempting to comply with these requirements under conditions of uncertainty.

Even in cases where the agent faces no moral dilemma, uncertainty can create special problems for Kantians. In Table 1 we summarize the results of the analysis provided here. We have argued that the formal system of expected utility theory (EUT) can be adapted to reflect the Kantian's focus on the characteristics of acts and not simply the states of the world that those acts bring about. In order to accommodate the Kantian's distinction between the value of persons and the value of things under uncertainty, however, EUT has to assign infinite disutility to the prospect of a loss of dignity. As a result, it is unclear how such an agent could act in a world where there is almost always some uncertainty about the implications of morally significant acts for the dignity of some person.

Threshold significance rules fare somewhat better. These rules effectively specify a level of acceptable uncertainty in the context of risk to dignity. A threshold that is greater than 0 and less than 1 allows for non-trivial decision making under uncertainty in a way that arguably respects the Kantian's core values. On the other hand, these rules are vulnerable to certain inconsistencies of choice and therefore require strategies for avoiding or mitigating these inconsistencies.

Table 1: Summary.

	Expected Utility Theory	Θ_s	Φ
Price-resistance under certainty	*Informal*: The utility of dignity is higher than the utility of any material good. *Formal*: $u(\text{dignity}) > u(\text{material good})$.	*Informal*: Any threshold less than 1. *Formal*: $s < 1$.	*Informal*: Any function where for all utilities, the threshold is less than 1. *Formal*: $(\forall g)(\Phi(g) < 1)$.
Weak price-resistance under certainty	*Informal*: Dignity is assigned an infinitely large utility. *Formal*: $u(\text{damage to dignity}) = -\infty$.	*Informal*: Any threshold less than 1. *Formal*: $s < 1$.	*Informal*: Φ is not of the form g/v. *Formal*: $\Phi(g) \neq g/v$.
Strong price-resistance under certainty	Informal: Dignity is assigned an infinitely large utility. *Formal*: $u(\text{damage to dignity}) = -\infty$.	*Informal*: Any threshold less than 1. *Formal*: $s < 1$.	*Informal*: Φ is a threshold rule with a threshold less than 1. *Formal*: Φ is constant and $\Phi < 1$.
Non-triviality	*Informal*: Dignity is *not* assigned an infinitely large utility. *Formal*: $u(\text{damage to dignity}) > -\infty$.	*Informal*: Any threshold greater than 0. *Formal*: $0 < s$.	*Informal*: Φ allows some risk to dignity. *Formal*: $(\exists g)(\Phi(g) > 0)$.
Satisfies all of the above	It is impossible to satisfy all the above condtions.	*Informal*: Any threshold between 0 and 1. *Formal*: $0 < s < 1$.	*Informal*: Φ is a threshold rule with a threshold between 0 and 1. *Formal*: Φ is constant and $0 < \Phi < 1$.

As we have defined it, the threshold s in Θ_s rules remains fixed across all of an agent's choices. Whether these rules can be adapted to allow the decision maker to use a different threshold s in different choice contexts is a topic that Kantians may wish to explore in greater detail. Our analysis of Φ rules demonstrates that if s varies as a function of the material gains at stake, then agents will act in ways that are inconsistent with a strong articulation of Kantian values. However, theories in which the threshold s varies according to features of the decision that are independent of the value of the material gain at stake could be promising.

References

R. Aboodi, A. Borer, and D. Enoch. Deontology, individualism, and uncertainty, a reply to Jackson and Smith. *Journal of Philosophy*, 105 (5):259–272, 2008.

L. Alexander. Deontology at the threshold. *San Diego Law Review*, 37 (893), 2000.

E. Athanasiou, A.J. London, and K.J.S. Zollman. Dignity and the value of rejecting profitable but insulting offers. *Mind*, 124(494):409–448, 2015. doi:10.1093/mind/fzu186.

T. Bergstrom. When is a man's life worth more than his human capital? In M. Jones-Lee, editor, *The Value of Life and Safety*. Amsterdam: North Holland, 1982.

S. Brennan. How is the strength of a right determined? Assessing the harm view. *American Philosophical Quarterly*, 32(4):383–393, 1995.

L. Buchak. *Risk and Rationality*. Oxford: Oxford University Press, 2013.

M. Colyvan, D. Cox, and K. Steele. Modelling the moral dimension of decisions. *Nous*, 44(3):503–529, 2010.

D. Davidson, J.C.C. McKinsey, and P. Suppes. Outlines of a formal theory of value. *Philosophy of Science*, 22(2):194–216, 1955.

A. Ellis. Deontology, incommensurability and the arbitrary. *Philosophy and Phenomenological Research*, 855, 1992.

J.M. Fischer and M. Ravizza. *Ethical Problems and Principles*. Harcourt Brace Jovanovich, 1992.

P. Hawley. Moral absolutism defended. *The Journal of Philosophy*, 105 (5):273–275, 2008.

F. Jackson and M. Smith. Absolutist moral theories and uncertainty. *Journal of Philosophy*, 53(6):267–283, 2006.

S. Kagan. *Normative Ethics*. Boulder, Colorado: Westview Press, 1998.

I. Kant. *Groundwork of the Metaphysics of Morals*. Cambridge: Cambridge University Press, 1997. M. Gregor, translator; C.M. Korsgaard, editor.

H. Kunreuther, N. Novemsky, and D. Kahneman. Making low probabilities useful. *Journal of Risk and Uncertainty*, 23(2):103–120, 2001.

I. Levi. *Hard Choices: Decisionmaking Under Unresolved Conflict*. New York: Cambridge University Press, 1990.

R.B. Marcus. Moral dilemmas and consistency. *The Journal of Philosophy*, 77(3):121–126, 1980.

R. Nisbett and T.D. Wilson. Telling more than we can know: Verbal reports of mental processes. *Psychological Review*, 84(3):231–259, 1977.

W.D. Ross. *The Right and the Good*. Indianapolis: Hackett Publishing Company, 1930.

G. Ryle. *The Concept of Mind*. New York: Barnes and Noble, 1949.

L.J. Savage. *The Foundations of Statistics*. John Wiley & Sons, 1954.

F. Schick. Dutch bookies and money pumps. *Journal of Philosophy*, 83 (2):112–119, 1986.

T. Seidenfeld. Decision theory without "independence" or without "ordering": What is the difference? *Economics and Philosophy*, 4:267–290, 1988.

A. Sen. Maximization and the act of choice. In *Rationality and Freedom*. Belknap Press, 2004.

A. Sen. *The Idea of Justice*. Cambridge, MA: The Belknap Press of Harvard University, 2009.

B. Steinbock. *Killing and Letting Die*. Prentice Hall, 1980.

R.H. Strotz. Myopia and inconsistency in dynamic utility maximization. *Review of Economic Studies*, 23(3):165–180, 1955.

J. von Neumann and O. Morgenstern. *Theory of Games and Economic Behavior*. Princeton: Princeton University Press, 1953.