Deontological decision theory

Barrington, Mitchell

DEONTOLOGICAL DECISION THEORY

Submitted by
Mitchell Barrington BA(Hons) *W.Aust.*

A thesis submitted in total fulfillment of the requirements of the degree of
Master of Philosophy

Dianoia Institute of Philosophy
Faculty of Theology and Philosophy

Australian Catholic University

2022

This thesis contains no material that has been extracted in whole or in part from a thesis that I have submitted towards the award of any other degree or diploma in any other tertiary institution.

No other person's work has been used without due acknowledgment in the main text of the thesis.

All research procedures reported in the thesis received the approval of the relevant Ethics/Safety Committees (where required).

CONTENTS

ABSTRACT

This thesis consists of four papers on ethics and decision theory. Although the papers are closely related, each is written to make a standalone contribution. So, the thesis should not be read as one continuous project.

Ignoring the Improbable. Many problems in decision theory appear to be solvable if agents simply ignore some possibilities. The utility of this approach has given rise to a substantial number of theories endorsing *discounting*: ignoring states whose probability is below a particular threshold. This paper argues that ignoring possibilities—even extremely remote ones—comes at a hefty cost for one's ability to make rational decisions. First, the approach is inescapably partition-sensitive: agents will undertake different acts depending on how the world is described. Second, agents become insensitive to differences in the probability of excluded states; they will be indifferent between taking a small risk and a much smaller risk. Third, agents become insensitive to differences in the value of outcomes in excluded states; they will be indifferent between risking a bad outcome and the same risk of a much worse outcome. And fourth, excluding a state affects the expected value of all acts to which the state is relevant, generating implausible prescriptions for peripheral acts; for instance, agents will not take *any* bet on the excluded state since they have assigned it a probability of zero.

Superiority Discounting Implies the Preposterous Conclusion. Many population axiologies avoid the Repugnant Conclusion (RC) by endorsing Superiority: Some number of great lives is better than any number of mediocre lives. But as Nebel shows, RC follows (given plausible auxiliary assumptions) from the Intrapersonal Repugnant Conclusion (IRC): A guaranteed mediocre life is better than a sufficiently small probability of a great life. This result is concerning because IRC is plausible. Recently, Kosonen has argued that IRC can be true while RC is false if small probabilities are *discounted* to zero. This paper details the unique problems created by combining Superiority with discounting. The resultant view, *Superiority Discounting*, avoids the Repugnant Conclusion only at the cost of the Preposterous Conclusion: Near-certain hell for arbitrarily many people is better than near-certain heaven for arbitrarily many people.

Filtered Maximization. According to moral absolutism, consequentialist considerations may justify sufficiently small risks of violating a duty but not sufficiently large risks. But finding a value function that accommodates these preferences is notoriously difficult: Seemingly, if some amount of consequentialist value outweighs a small risk, then some larger amount should be able to outweigh a larger risk. Critics have taken this difficulty to warrant rejection of absolutism generally; others have attempted to solve the problem by offering decision-theoretic models of absolutist decision making. I outline five desiderata that such a model must satisfy and demonstrate that none of the leading four theories satisfy all five. Then, I present an alternative: Filtered Maximization. This theory models absolutists as assigning "duty value" to outcomes and filtering out acts whose expected duty value is sufficiently low. Then, of those remaining, agents perform the act that maximizes expected value. I illustrate how Filtered Maximization satisfies all five desiderata, then conclude by discussing some implications for existing absolutist theories.

Superiority and Separability. *Superiority* is the view that there exists some pair of valuable objects *x* and *y* such that some quantity of *x* is better than any quantity of *y*; it is very plausible when *x* is an important good and *y* is trivial, such as in the Repugnant Conclusion. This paper shows that (given modest auxiliary assumptions) Superiority is incompatible with *Separability*—the principle that in comparing the value of two outcomes, we may ignore people whose welfare and existence are unaffected.

Ignoring the Improbable

1. Introduction

A number of decision-theoretic proposals instruct agents to ignore sufficiently improbable possibilities—a strategy I will call "discounting".[1] Discounting solves many stubborn problems. Absolutist moral theories are able to give sufficient guidance in contexts of risk if we ignore sufficiently small chances of violating our duty (Kagan 1989: 89-92; Aboodi, Borer, and Enoch 2008; Hawley 2008; Haque 2012; Bjorndahl, London, and Zollman 2017; Lazar 2017; Lee-Stronach 2018; Tarsney 2018). Pascal's Mugging is solved by ignoring the possibility that the mugger is telling the truth (Jordan 1994: 218-19; Chalmers 2017; Schwitzgebel 2017: 273; Monton 2019). The St. Petersburg Paradox (and its variants) are solved by ignoring the possibility that the game goes on a sufficient number of coin flips (Bernoulli 1738; d'Alembert 1761; Buffon 1777; Condorcet 1785; Borel 1962; Jordan 1994: 217-18; Buchak 2013: 73-74; Smith 2014, 2016; Robert 2018; Monton 2019). I will limit my discussion to discounting's use as a strategy for solving these three decision-theoretic problems, but it is worth noting its use in myriad other contexts and for a variety of reasons.[2]

---

[1] In recent years, it has been referred to as "Nicolausian discounting" in reference to Nicolaus Bernoulli's suggestion, in a 1714 letter, that the solution to the St. Petersburg Paradox is to have agents discount small probabilities to zero (Monton 2019: 6). When talking about the strategy itself—irrespective of its purpose—I will simply refer to it as "discounting", theories that utilize it as "discounting theories" and those who endorse it as "discounters".

[2] The suggestion that "*de minimis*" (negligible) risk should be ignored entirely in decision making is ubiquitous in public health, medical, and risk-assessment scholarship. (Peterson (2002) provides a good introduction to *de minimis* principles and their use in these fields.)

It has also been used by philosophers in the literature on collective action. For instance, Miller and Sartorius (1979) argue that free-riding on the contribution of others is often rational and moral, because the probability of any one individual's contribution (e.g., voting) determining whether the (perceived) mutually advantageous outcome is produced (e.g., the best candidate being elected) is so small as to be rationally negligible. In a similar vein, Buchanan (1979) criticises the Marxist claim that the proletariat will revolt due to their collective interests being served by the revolution. Buchanan contends that while each proletarian might be greatly benefitted by the revolution, revolting is not in any individual's self-interest because the probability that *their* revolting will make the revolution successful is rationally negligible.

More recently, Kosonen (2021) has endorsed discounting to attack *ex ante* pareto principles: those holding that if some prospect is better for everyone, then that prospect is simply better. Kosonen argues (in response to Nebel (2019: 320)), that a prospect's risk to each individual might be small

This paper argues that ignoring possibilities comes with a price.[3] In the remainder of Section 1, I will briefly introduce three prominent problems that discounting is invoked to solve. Section 2 explores how discounting generates inconsistent prescriptions depending on how states are partitioned; states may be above the threshold when partitioned coarsely but below the threshold when partitioned finely. In 2.1, I consider the response that the threshold is sensitive to how fine-grained the partition is and argue that this approach does not escape the problem. Sections 3 and 4 demonstrate that when differences between two acts are quarantined to an excluded state, agents are insensitive to these differences. Section 3 contends that discounting makes agents indifferent between acts with a small probability of producing a bad outcome and a much smaller probability of producing the same outcome. In 3.1, I consider whether setting the threshold between these different probabilities will solve the problem, arguing that this approach violates a plausible principle. Section 4 describes how discounting makes agents indifferent between acts with a small chance of producing a bad outcome and acts with the same chance of producing a much worse outcome. Section 5 demonstrates how excluding a state from the decision problem affects the expected value of *all* acts to which the state is relevant; discounting produces a correct prescription for one act only at the cost of an incorrect prescription for another. In 5.1, I consider the response that the threshold should be relative to the act and argue that this approach cannot accommodate some acts.

1.1 *Moral Absolutism and Risk*

Absolutist moral theories are committed to the claim that consequentialist considerations can never justify violating a moral duty. Irrespective of whether it is true, moral absolutism appears to be a perfectly coherent position. Certainly, it is what most ordinary people

---

enough to ignore, while the risk to the group is too great to ignore. Consequently, the prospect might be better for each individual but worse from an impartial perspective.

[3] One note on terminology. I have so far spoken of ignoring possibilities. By a possibility, I mean a state or set of states. For instance, ignoring the possibility that it is raining involves ignoring all the states in which it is raining. And ignoring the possibility that I will get wet if I go outside without an umbrella involves ignoring all the states in which the outcome of going outside is getting wet. As a result, ignoring either of these two possibilities would (presumably) amount to ignore the same set of states. This framework generally conforms to Savage's (1954) theory, but my criticisms are not attached to it. Throughout the paper, I will generally use "possibility" in ordinary parlance and "state" (or "states") when there is a specified decision problem.

would espouse when asked how much pleasure would have to be on the line for killing an innocent person to be morally right. The intuitive thought is that pleasure can be valuable without it being the case that enough of it can justify killing someone. The puzzle for absolutism is that if it is coherent, then it should be possible to model it decision-theoretically. However, this task is notoriously difficult.

The fact that *no* amount of pleasure can outweigh the duty to avoid killing seems to indicate that something strange is going on with the absolutist's value function. At a first pass, we might conclude that the disvalue of violating a duty is *infinitely* greater than the value of upholding consequentialist considerations.[4] After all, if the disvalue of violating the duty were finite, then a sufficient amount of pleasure will outweigh the disvalue of killing, and maximising expected value will lead one to kill for the sake of pleasure. But the infinite-value model has problems. Since any positive value multiplied by infinity amounts to infinity, all acts with *any* risk of violating the duty will have the same, infinitely negative, expected value.[5] The unacceptable result of all acts having the same expected value is that all acts are equally permissible—even those certain to violate a duty.

We can avoid this result by eschewing the infinite-value model. Instead, we might model absolutism with multiple value functions in a lexicographical ordering (e.g., Lee-Stronach 2018: 796-99). The lexicographic model tells agents to maximise primary value (the value of upholding their duties), only looking to secondary value (consequentialist value) to break ties.[6] This approach avoids the first problem but leads to another: agents must *always* perform the act with the smallest risk of violating their duty. They "have to

[4] This modelling assumption is made by: Jackson and Smith (2006, 2016); Colyvan, Cox, and Steele (2010); Huemer (2010); Hayenhjelm and Wolff (2012); Hansson (2013); Bjorndahl, London, and Zollman (2017).

[5] Problems of this form are offered by: McKerlie (1986); Ashford (2003: 298); Jackson and Smith (2006, 2016); Colyvan, Cox, and Steele (2010); Huemer (2010); Fried (2012); Sobel (2012); Isaacs (2014); Holm (2016); Tenenbaum (2017); Alexander (2018).

[6] There are some alternative ways of modelling absolutist theories. Notably, Lazar and Lee-Stronach (2019) make the value of upholding the secondary consideration bounded in such a way that it can never agglomerate to outweigh the value of upholding one's duty; Lee-Stronach (2021) suggests that the value of upholding secondary considerations is contingent on fulfilling one's duty. However, the infinite-value approach is by far the most common, and the lexicographic model is (in my view) the most straightforwardly plausible representation of absolutism.

stay huddled in the corner, not daring to move, trying not to breathe too loudly" out of fear of raising their probability of killing someone (Kagan 1989: 89).

These results are unacceptable because probabilities matter. Absolutists must disregard sufficiently small risks (or small increased risk, as on the lexicographic model) while nevertheless allowing substantial risks to outweigh *any amount* of secondary value.[7]

1.2 *Pascal's Mugging*

Pascal is approached by Mugger, who asks for his wallet, promising to pay him back double its value. Being committed to maximising expected utility, Pascal declines on the grounds that the probability of Mugger telling the truth is less than 0.5, so the deal has negative expected value. Mugger then asks what probability Pascal assigns to the possibility of him telling the truth. Whatever (non-zero) number Pascal responds with, Mugger makes an offer valuable enough to counteract the risk, giving the deal positive expected value. The price they settle on is 1,000 quadrillion happy days of life, which Mugger would no issue procuring if he had the magical powers he claims to have. If Pascal maximises expected value, he will keep giving his wallet away to any crook who makes a large enough promise. Maximising expected value here appears to be irrational (Bostrom 2009).

1.3 *The St. Petersburg Paradox*

How much would you be willing to pay to play the following game? A fair coin is tossed until it comes up heads. You will then be paid $\$2^n$, where n is the number of times the coin was flipped. So, if the coin comes up heads on the first toss, you win \$2; on the second toss, \$4; on the third toss, \$8. Intuitively, you should not pay very much to play the game. While you *could* end up with a substantial amount of money, the probability of this is vanishingly small. Indeed, half the games reward the player with no more than \$2; three quarters reward them with no more than \$4. But if you maximise expected value, you will

[7] Subsequent discussion will use infinite value to model absolutism because it is by far the most popular approach. However, the criticisms apply to any theory that endorses discounting.

be willing to pay *any* finite amount for a single game because its expected value appears to be infinite:[8]

$$½ \cdot 2 + ¼ \cdot 4 + ⅛ \cdot 8 \cdots = 1 + 1 + 1 \cdots = \sum_{n=1}^{\infty} (½) \cdot 2^n = \infty$$

1.4 *Discounting*

The solution is to posit a probabilistic threshold, *t*, governing which possibilities you must attend to, and which are so improbable that you should ignore them. If a state's probability is greater than or equal to *t*, you attend to it as usual. But if its probability is strictly less than *t*, you simply ignore it, treating it as though its probability were zero. (Presumably, after sufficiently improbable states have been excluded, the probabilities of the remaining states are updated in line with Bayesian rules of belief revision.)

Discounting allows absolutists to uphold their duties under risk by ignoring small chances of violating their duties while allowing substantial risks to outweigh any other considerations. It enables agents to hold onto their wallets in the face of philosophically literate muggers since they will not attend to the possibility of them telling the truth. And it caps the amount agents will be willing to pay to play the St. Petersburg game since the remote possibilities in which they win large sums of money will not contribute to the game's expected value. In what follows, I will focus on the problem for moral absolutism, but each problem can be extended to all discounting theories.

[8] The conventional response to the problem is to posit that an agent's utility function is bounded, preventing the expected value of the game from reaching infinity (Arrow 1971; Bassett 1987; Samuelson 1977; McClennen 1994). However, Monton (2019: 2-3) offers a convincing case where this response does not work: if the game is truncated at 999 tosses and you are gambling with your 1,000 remaining days of life (the value of which is not bounded, he argues), then the expected value of the game will be 999 life-days. But gambling 999 life-days has a 7/8 chance of leaving you with less than 10 days to live. We should not, it seems, maximise expected value here.

2. Partition Sensitivity

Discounting generates partition-sensitive prescriptions; when states are described coarsely, the theory prescribes one act, but when they are described finely, it prescribes another. Consider the following case:

> Boris and Doris are two combatant military commanders with absolutist moral duties to avoid killing civilians. Both of their respective countries have recently developed an identical missile, and each commander would like to test it. On the one hand, testing the missile would be very beneficial. However, each missile has a small chance of malfunctioning and propelling itself into a nearby building, killing its occupants. Boris's and Doris's cases are identical in all but one respect: If Boris's missile malfunctions, it will strike a building containing 100 civilians, while if Doris's malfunctions, it will hit one of two buildings: one containing 99 civilians and one containing 101 civilians.[9]

Boris's decision problem contains two states. In Safe, the missile works as intended, and he benefits from having tested it. In ¬Safe, it malfunctions and kills 100 people. The probability of ¬Safe is 0.004, and *t* = 0.003.

TABLE 1

	Safe (0.996)	¬Safe (0.004)	Expected value
Fire	10	$-\infty$	$-\infty$
¬Fire	0	0	0

Since ¬Safe is above the threshold, Boris cannot ignore it, and firing the missile is forbidden.

Doris must construct her decision problem differently because there are three possible outcomes of firing the missile in her case. In Safe, the test is conducted safely; in $\neg Safe_{99}$, the missile malfunctions and kills 99 people; in $\neg Safe_{101}$, the missile malfunctions

[9] To be clear, this case is not, technically, a partition problem. Partition problems occur when one world is described in two different ways; this case concerns two different worlds that, due to their differences, should be described differently. Boris's partition is suitable for his decision and Doris's is suitable for hers. This case represents a greater problem than standard partition problems because it cannot be avoided by privileging one partition over another.

and kills 101 people. She assigns the same 0.996 probability to Safe but splits the complement evenly among $\neg S_{99}$ and $\neg S_{101}$.

TABLE 2

	Safe (0.996)	$\neg$Safe$_{99}$ (0.002)	$\neg$Safe$_{101}$ (0.002)	Expected value
Fire	10	$-\infty$	$-\infty$	10
¬Fire	0	0	0	0

Since the probabilities of $\neg$Safe$_{99}$ and $\neg$Safe$_{101}$ are each 0.002, and t = 0.003, she ignores these states and redistributes her credence into Safe. As a result, Fire > ¬Fire. Discounting gives the bizarre result that testing the missile is forbidden for Boris but permitted for Doris.

In addition to the partition problem (that conflicting prescriptions are generated in relevantly similar cases), there is a further problem: discounters may ignore arbitrarily large probabilities. While the previous case partitioned the probability of violating the duty into two states, there is no limit on the number of states into which this probability can be split. So, no matter how risky an act is, there is some number of states n large enough that $1/n < t$; if the probability of violating the duty is split equally among n states, each state's probability will be below the threshold. As a result, the (arbitrarily probable) possibility of violating the duty will be ignored and the (arbitrarily risky) act permitted.[10]

2.1 *Partition-sensitive Thresholds*

One solution to the partition problem is to relativise the threshold to the partition: when states are partitioned coarsely, the threshold is correspondingly higher than when they are

[10] There is an additional problem for theories that discount *relatively* improbable states. Lee-Stronach's threshold, for instance, "evaluates whether a state is sufficiently *more probable than its most probable alternative*" (2018: 801, my emphasis). To see the issue, suppose that Safe, where your duty is upheld, has a probability of 0.998, while ¬Safe, where your duty is violated, has a probability of 0.002; and t = 0.003. If Safe is split into two equally probable states, then ¬Safe will become twice as probable, *relative to the most probable alternative*. While the *relative* threshold will remain unchanged, the *absolute* threshold will halve from 0.003 to 0.0015, and the previously permitted act will now be forbidden. Again, with no upper bound on the number of states able to feature in a decision problem, the threshold can be driven to an arbitrarily low number, forbidding acts with an arbitrarily small probability of violating the duty.

partitioned finely. For instance, perhaps Doris's threshold—where the probability of violating the duty is split among two states—should be half that of Boris's. As a result, both Boris and Doris would be forbidden from firing the missile. But this solution is inadequate.

Consider Morris, a third military commander whose decision is the same as Boris's, but whose missile is safer: the probability of Morris's missile malfunctioning (¬Safe) is slightly less than t ($t - \varepsilon$, where ε is an arbitrarily small number).

TABLE 3

	Safe (0.997 + ε)	¬Safe (0.003 − ε)	Expected value
Fire	10	-∞	10
¬Fire	0	0	0

Morris excludes ¬Safe from the decision problem and is permitted to test the missile.

Now, pick some n large enough that $1/n < t$, and suppose that if the missile does not malfunction, Morris will be rewarded with one of n (differently valued) prizes. The partition will now consist of n equally probable states ($Safe_1$, $Safe_2$ . . . $Safe_n$), corresponding to the possible outcomes of the lottery.

TABLE 4

	$Safe_1$ (0.997/n)	$Safe_2$ (0.997/n)	$Safe_n$ (0.997/n)	¬Safe (0.003 − ε)	Expected value
Fire	10 + prize	10 + prize	10 + prize	-∞	0
¬Fire	0	0	0	0	0

Since the probability of each remaining state is less than $1/n$—and by stipulation, $1/n$ is below the threshold—the discounter will ignore all states in the decision problem.

The discounter's solution is to lower t to below $1/n$, allowing states ($Safe_1$ . . . $Safe_n$) to enter the equation. However, since ¬Safe is lurking marginally below t, lowering the threshold will cause ¬Safe to enter the equation, making the act forbidden. The result is that Fire is permitted *unless* it comes with the lottery—even though the lottery merely sweetens the deal! As there is no way to have *t* below states ($\text{Safe}_1 \ldots \text{Safe}_n$) but above ¬Safe, the discounter's strategy fails to avoid the partition problem.[11]

Partition problems appear to be a relatively deep feature of discounting theories. Orthodox decision theory avoids them because states' probabilities will have the same sum, regardless of how they are partitioned. But when we discount, if some probability is split among sufficiently improbable states, they will count for nothing. Nevertheless, even if partition sensitivity is written off as merely a framing problem, more-troubling problems remain.

3. Probability Insensitivity

The threshold is too blunt an instrument. In excluding a problematic state from the decision problem, important features of that state go with it. In the first place, excluding a state prevents agents from deliberating on its probability. Consider a case:

> You are a clinician who is preparing a vaccine dose for your next patient. Unfortunately, a small portion of the vaccines are defective, and injecting someone with a defective vaccine will kill them. Despite your absolutist duty to avoid killing your patients, you can ignore risks less than one in 500,000. And luckily, the probability of each vaccine dose being defective is well below this mark, at one in 100 million. However, you recently learned that mixing a certain chemical into a vaccine dose has two interesting effects. First, it turns the clear liquid bright pink. And second, it makes the dose one hundred times more deadly. While waiting for your patient, you try mixing the chemical into a dose to watch it change colour. And as your patient enters the room, you wonder if it would be wrong to inject her with it.

Obviously, you should not inject your patient with the pink vaccine: you have a perfectly good alternative that is one hundred times less likely to kill her. However, even though the probability of Pink being defective is one hundred times greater than that of the original vaccine, both states are nevertheless under the 1/500k threshold. So, you

[11] I am thankful to Dmitri Gallow for articulating this response.

ignore these states and reason that your patient will be safely vaccinated, whichever vaccine you pick.

TABLE 5

	Neither Defective (1 – 1/1m – 1/100m)	Pink Defective (1/1m)	Original Defective (1/100m)	Expected value
Pink	10	-∞	10	10
Original	10	10	-∞	10
None	0	0	0	0

As a result, both acts have the same expected value; each is equally permissible. But this result is wrong. The only difference between the two is that Pink is one hundred times deadlier than Original. There is no reason to impose the additional risk upon your patient, and no theory should treat this difference as entirely irrelevant to your decision.

3.1 *Threshold Manipulation*

Let us consider another response from the discounter, who objects to the fact that we were in charge of setting the threshold. "You tell *me* the relevant probabilities", she insists, "and *I'll* tell you the value of *t*". Her strategy is immediately apparent: wherever we set the probabilities, she will set *t* between the two, forcing the agent to attend to one possibility while ignoring the other. As a result, the theory will forbid one act and permit the other, thus eliminating the possibility of a tie. So, in a choice between a 1/1m and a 1/100m risk, perhaps t = 1/50m.

TABLE 6

	Neither Defective (1 – 1/1m – 1/100m)	Pink Defective (1/1m)	Original Defective (1/100m)	Expected value
Pink	10	-∞	10	-∞
Original	10	10	-∞	10
None	0	0	0	0

As the state where Pink is defective is now above the threshold, the theory gives the correct prescription: you ought to inject Original.

If we lower the probability of Pink Defective to 1/60m (and thus below the threshold), then the discounter simply shifts *t* down to 1/70m—once again avoiding indifference. As long as there is *any* gap between the two probabilities (which there must be for the theory to be insensitive to a difference in probabilities), she will be able to set the threshold between them.

The discounter's view—that the threshold is always set in the most convenient position for her theory—certainly sounds desperate. Waiting to see the available acts before placing the threshold between the best and second-best acts, irrespective of their risk, seems to raise the worry that the threshold is arbitrary. If the threshold were sensitive to the available acts, we would be unable to infer from our judgments about one case where the threshold would be for another case with different options, when the cases are otherwise identical.[12] Nevertheless, we can say more about this strategy than merely that it is implausible. Consider the following popular principle:

Irrelevant Alternatives One's preference ordering between acts does not change if new acts are added to the decision problem.[13]

Irrelevant Alternatives gets at the following idea: if you prefer an apple to an orange, adding the option of a banana should not make you suddenly prefer an orange to an apple.

Consider the previous case, where the value of *t* is 1/50m, so you ignore Original Defective but attend to Pink Defective.

TABLE 7

	Neither Defective (1 – 1/1m – 1/100m)	Pink Defective (1/1m)	Original Defective (1/100m)	Expected value
Pink	10	-∞	10	-∞
Original	10	10	-∞	10
None	0	0	0	0

[12] The formulation of this concern is influenced by Lazar's (2017: 592) criticism of an unrelated theory.

[13] This articulation is from Peterson (2017: 59).

In this case, you express the following preference ordering: Original > None > Pink. However, now suppose that, as you decide to inject Original, you are told that new vaccine stock has just arrived. The arrival of these vaccines has been highly anticipated because they have been transported in ultra-low-temperature freezers, which reduces the rate of defectiveness tenfold. As you open the freezer, you can verify that they have been kept at -80° because they have taken on a bluish hue.

Naturally, the blue vaccine is now the best option. To extract this prescription, the discounter moves *t* from 1/50m to somewhere between 1/100m and 1/1b, so you will ignore Blue Defective but not ignore Original Defective.

TABLE 8

	Neither Defective (1 – 1/1m – 1/100m – 1/1b)	Pink Defective (1/1m)	Original Defective (1/100m)	Blue Defective (1/1b)	Expected value
Pink	10	-∞	10	10	-∞
Original	10	10	-∞	10	-∞
Blue	10	10	10	-∞	10
None	0	0	0	0	0

However, your preference ordering is now: Blue > None > Original ~ Pink. While you preferred Original to Pink in the original problem, you are now indifferent between them. And while you preferred Original to None in the original problem, the addition of Blue has *reversed* this ordering: you now prefer None to Original. Unfortunately for the discounter, there is no way to manipulate the threshold without opening the door for violations of Irrelevant Alternatives. But even if we are not concerned by the discounter's violation of Irrelevant Alternatives, a related problem cannot be solved by threshold manipulation.

4. Value Insensitivity

Ignoring a possibility also prevents us from deliberating on its value. Consider a case:

> Deep brain stimulation has been approved as a treatment for severe, medication-resistant depression. You are a surgeon who has been tasked with implanting the

neurostimulator in the patient in front of you. However, two companies make the required neurostimulator, and you must choose which product to use: NS1 or NS2. The effectiveness of each neurostimulator largely depends on the patient's genes, which you are unable to determine. In S_1, the patient's genes are receptive to NS1: her depression would be cured by NS1 and unaffected by NS2. In S_2, her genes are receptive to NS2: her depression would be cured by NS2 and unaffected by NS1. S_1 and S_2 each have a probability of 0.499. The remaining 0.002 is assigned to S_3, in which the neurostimulator will malfunction, paralysing the patient from the waist down. Although you have a duty to avoid harming your patient, you decide that the risk is small enough to ignore. There is one difference between the two devices, however: if S_3 obtains, using NS1 will paralyse the patient and fail to cure her depression, while using NS2 will paralyse the patient but nevertheless cure her depression.

TABLE 9

	S_1 (0.499)	S_2 (0.499)	S_3 (0.002)	Expected value
NS1	10	0	$-\infty$	5
NS2	0	10	$-\infty$	5
None	0	0	0	0

Since you are only looking at S_1 and S_2, and in these states, each act has a 0.5 probability of curing the patient with no side effects and a 0.5 probability of being ineffective, NS1 and NS2 have the same expected value. However, you should not be indifferent between the two options. NS2 guarantees that, even if the patient is paralysed, her depression will be cured, while NS1 does not. Clearly, you should use NS2. And the discounter cannot achieve this result by setting the threshold between the two states' probabilities because there is only one state being ignored; the difference between the acts is merely their outcomes in this state.[14] Nor can she resort (as some have suggested)

[14] Ironically enough, the explicit purpose of arranging considerations lexicographically—so that one consideration takes absolute precedence over the other, as in absolutist moral theories—is to consult secondary considerations "only when necessary to break ties" (Rawls 1971: 42, note 23). But if the value of upholding secondary considerations is contained within excluded states, then it will fail to break ties. (You had one job!)

to adding a dominance principle, stipulating that agents always prefer dominating acts (Hájek 2014: 557; Monton 2019: 20). Dominance reasoning does not help in the case presented because neither act is dominating the other: NS1 is better in S_1, and NS2 is better in S_2.

Discounting makes agents indifferent between acts whose differences are quarantined to excluded states. But this indifference becomes even more implausible when the difference between two acts is the degree to which each would result in a duty violation. For instance, the probability of hitting someone with your car on your way to work today is (presumably) small enough to ignore. However, if your car contained a device that would set off the world's nuclear warheads—destroying all life on Earth—in the event you hit someone today, you should probably take the day off work. There should be some probability of hitting someone, *p*, such that a *p* chance of hitting someone with the normal car is permissible, but a *p* chance of hitting someone with the nuclear car is not. But to the discounter, if *p* is above the threshold, then both acts are forbidden, and if *p* is below the threshold, then both acts are *equally* permissible. For no value of *p* will the discounter drive the normal car without being equally willing to drive the nuclear car.[15]

5. Peripheral Acts

The justification for discounting is that it generates a plausible expected value for acts that orthodox expected value theory gets wrong. However, states can matter differently to different acts, and the effects of ignoring a state will not be isolated to a single act.

> You are watching TV late at night with your friend Jimmy when he suggests playing trumpet in the basement instead. You would like to, but you promised not to play trumpet while your housemate, Leonard, is home, and you have a duty to keep that promise. The problem is that you do not know if he is home. Leonard is the fire department chief, so he only sleeps at home one (randomly determined) night per

[15] Perhaps cases like these suggest that the threshold should be act-relative, so an act with a particular probability of killing one person might be permitted while another act with the same probability of killing many people is forbidden. I will demonstrate why act-relative thresholds do not work in 5.1.

calendar year. And since today is January 1st, each of the next 365 nights has a 1/365 chance of being his night off.

If you are permitted to ignore probabilities less than 1/350, then you will be permitted to ignore the state where Leonard is home and enjoy your trumpeting.

TABLE 10

	Home (1/365)	¬Home (364/365)	Expected value
Trumpet	-∞	10	10
None	0	0	0

Before you can give Jimmy an answer, he offers you a wager: if Leonard is home, he will pay you $1 billion; if he is not home, you must pay him a penny. Jimmy is an eccentric billionaire who constantly gives absurd amounts of money away, so you know he is good for the money. In fact, you have been waiting for him to make you such an offer. And fortunately, his offer is extremely favourable: a ticket to a 365-ticket lottery that pays out $1 billion is well worth the price of 1¢. However, discounting theories do not get this result.[16]

TABLE 11

	Home (1/365)	¬Home (364/365)	Expected value
Trumpet	-∞	10	10
Bet	1b	-0.01	-0.01
None	0	0	0

Since you have discounted the probability of Leonard being home to zero, you see the bet as offering a 100% chance of losing 1¢ with a 0% chance of winning $1 billion. According to discounting theories, this favourable bet is tantamount to throwing your money away.

[16] For simplicity, the following table represents the value of money linearly, with a dollar being equal to one unit of value. As will become clear, this representation makes no difference to the substance of the problem.

5.1 *Act-Relative Thresholds*

Once a bet against the excluded state has been added into the decision problem, we want to shift *t* below that state, so we are sensitive to the value of the bet. Unfortunately, as we saw in Section 3.1, shifting *t* when an option is added leads to violations of Irrelevant Alternatives. Perhaps, then, the discounter would employ a different strategy: relativising the threshold to acts. On this view, the value of Trumpet's outcome in Home would not contribute to its expected value, while *Bet's* outcome in Home would contribute to its expected value. Unfortunately, there are acts that this approach cannot accommodate. For instance, suppose we introduce some hybrid act that results in both trumpet-playing *and* taking the bet. We know that *t* is above Home for Trumpet and below Home for Bet, but where is *t* for the hybrid act? There is no possible value of *t* that generates the correct prescription that you should perform the act: if *t* is above Home, then you will ignore the possibility of winning the bet; if *t* is below (or equal to) Home, then you will attend to the possibility of winning the $1 billion, but *also* attend to the possibility of waking Leonard.

Suppose you go to play trumpet with Jimmy, having determined that you can ignore the possibility of Leonard being home. He offers you a choice of trumpet—Gold or Silver—over which you have no preference. However, he explains that if you play Gold, it will be understood that you have also accepted the bet, but if you play Silver, it will be understood that you have declined the bet.

Gold is preferable to Silver: both acts are identical except that Gold comes with a favourable bet. So, the discounter must set *t* somewhere that makes discounting prescribe Gold. Setting *t* above Home will yield the same result as before: you will ignore the possibility that Leonard is home and see the bet as giving away 1¢ for no reward. So, the discounter will want to set *t* below Home. As a result, you will attend to the possibility of Leonard being home, allowing you to enter the $1b into your expected-value calculation. However, since playing Gold not only contains the bet but *also* results in trumpet-playing—and trumpet-playing in Home would result in a violation of your duty—the infinite disvalue of waking Leonard will *also* enter the equation.[17]

---

[17] It is worth noting that this result is not a peculiarity of the infinite-value model. On the lexicographic model, lowering the threshold below Home will lower Gold's primary value and raise

TABLE 12

	Home (1/365)	¬Home (364/365)	Expected value
Gold	-∞	10	-∞
Silver	-∞	10	10
None	0	0	0

Now the $1 billion is in the equation, but the duty violation's infinite disvalue has swamped the outcome's value. This result is even worse than the original one: now you not only prefer Silver to Gold, but you *also* prefer None to Gold. Discounting results in trumpet-playing being permitted, but trumpet-playing *with* a favourable bet being forbidden. Wherever we set the threshold, the discounter will always avoid the bet.

6. Conclusion

The challenges for discounting appear to be quite serious. Partition sensitivity seems to be a deep feature of discounting because arbitrarily probable states may always be split into enough states that each state's probability is below t. Insensitivity to both the probability and the value of risks also appears inescapable because excluded states do not contribute to the expected value of acts. And finally, since a state can be relevant to multiple acts, we cannot avoid influencing the expected value of peripheral acts when excluding states from the decision problem.

Solving these problems while retaining the spirit of discounting is a challenging task. But perhaps we can say a few things about what this kind of proposal would look like. Partition sensitivity (problem 1) can be avoided by considering the probability of some act violating a duty rather than the probability of states *in which* the act would violate the duty. The move away from ignoring states will also prevent agents from ignoring the value of these states' outcomes (problem 3). However, simply discounting this probability to zero will leave problems 2 and 4. A more promising approach would be to ignore acts whose probability of violating a duty is too great, irrespective of consequentialist considerations. As a result, agents will never violate a duty for the sake of consequentialist considerations,

its secondary value. And since you will maximise primary value, only looking to secondary value to break ties, you will prefer Silver.

eliminating the need for infinite-values or a lexicographical ordering. Excluding sufficiently risky acts (instead of sufficiently improbable states) will not affect the expected value of peripheral acts (problem 4). Finally, they will pick the act, among those remaining, that maximises expected value. Agents will not be indifferent between (otherwise equivalent) acts with different probabilities of violating the duty (problem 2) because these differences will eventually manifest as differences in expected value. While many features of this model remain undefined, I will have more to say on this matter in future work.

References

Aboodi, R., Borer, A., and Enoch, D. 2008. 'Deontology, Individualism, and Uncertainty: A Reply to Jackson and Smith.' *The Journal of Philosophy,* 105: 259-72.

Alexander, L. 2018. 'The Need to Attend to Probabilities—for Purposes of Self-Defense and Other Preemptive Actions.' *San Diego Law Review,* 55: 223-29.

Arrow, K. J. 1971. *Essays in the Theory of Risk-Bearing*. Amsterdam: North-Holland Pub. Co.

Ashford, E. 2003. 'The Demandingness of Scanlon's Contractualism.' *Ethics,* 113: 273-302.

Bassett, G. W., Jr. 1987. 'The St. Petersburg Paradox and Bounded Utility.' *History of Political Economy,* 19: 517-23.

Bernoulli, D. 1738. 'Exposition of a New Theory on the Measurement of Risk.' *Econometrica,* 22: 23-36.

Bjorndahl, A., London, A. J., and Zollman, K. J. S. 2017. 'Kantian Decision Making under Uncertainty: Dignity, Price, and Consistency.' *Philosophers' Imprint,* 17.

Borel, E. 1962. *Probabilities and Life*. New York: Dover.

Bostrom, N. 2009. 'Pascal's Mugging.' *Analysis,* 69: 443-45.

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

Buchanan, A. 1979. 'Revolutionary Motivation and Rationality.' *Philosophy & Public Affairs,* 9: 59-82.

Buffon, G.-L. L. d. 1777. Essai D'arithmétique Morale. In *Supplement a I'histoire Naturelle*. Paris: L'Imprimerie.

Chalmers, A. 2017. An Offer You Can't (Rationally) Refuse: Systematically Exploiting Utility-Maximisers with Malicious Gambles, University of Sydney.

Colyvan, M., Cox, D., and Steele, K. S. 2010. 'Modelling the Moral Dimension of Decisions.' *Noûs,* 44: 503-29.

Condorcet, J.-A.-N. d. C. 1785. *Essai Sur L'application De L'analyse à La Probabilité Des DéCisions Rendues à La Pluralité Des Voix*. Paris: Imprimerie royale.

d'Alembert, J. L. R. 1761. *Opuscules MathéMatiques, Ou MéMoires Sur DifféRens Sujets De GéOméTrie, De MéChanique, D'optique, D'astronomie &C*. Paris: David.

Fried, B. H. 2012. 'What Does Matter? The Case for Killing the Trolley Problem (or Letting It Die).' *The Philosophical Quarterly,* 62: 505-29.

Hájek, A. 2014. 'Unexpected Expectations.' *Mind,* 123: 533-67.

Hansson, S. O. 2013. *The Ethics of Risk: Ethical Analysis in an Uncertain World*: Palgrave Macmillan.

Haque, A. A. 2012. 'Killing in the Fog of War.' *Southern California Law Review,* 86: 63-116.

Hawley, P. 2008. 'Moral Absolutism Defended.' *Journal of Philosophy,* 105: 273-75.

Hayenhjelm, M. and Wolff, J. 2012. 'The Moral Problem of Risk Impositions: A Survey of the Literature.' *European Journal of Philosophy,* 20: e26-e51.

Holm, S. 2016. 'A Right against Risk-Imposition and the Problem of Paralysis.' *Ethical Theory and Moral Practice,* 19: 917-30.

Huemer, M. 2010. 'Lexical Priority and the Problem of Risk.' *Pacific Philosophical Quarterly,* 91: 332-51.

Isaacs, Y. 2014. 'Duty and Knowledge.' *Philosophical Perspectives,* 28: 95-110.

Jackson, F. and Smith, M. 2006. 'Absolutist Moral Theories and Uncertainty.' *Journal of Philosophy,* 103: 267-83.

Jackson, F. and Smith, M. 2016. The Implementation Problem for Deontology. In *Weighing Reasons*. New York: Oxford University Press.

Jordan, J. 1994. 'The St. Petersburg Paradox and Pascal's Wager.' *Philosophia,* 23: 207-22.

Kagan, S. 1989. *The Limits of Morality*: Oxford University Press.

Kosonen, P. 2021. 'Discounting Small Probabilities Solves the Intrapersonal Addition Paradox.' *Ethics,* 132: 204-17.

Lazar, S. 2017. 'Deontological Decision Theory and Agent-Centered Options.' *Ethics,* 127: 579-609.

Lazar, S. 2017. 'Risky Killing.' *Journal of Moral Philosophy,* 16: 1-26.

Lazar, S. and Lee-Stronach, C. 2019. 'Axiological Absolutism and Risk.' *Noûs,* 53: 97-113.

Lee-Stronach, C. 2018. 'Moral Priorities under Risk.' *Canadian Journal of Philosophy,* 48: 793-811.

Lee-Stronach, C. 2021. 'Morality, Uncertainty.' *Philosophical Quarterly,* 71: 334-58.

McClennen, E. F. 1994. Pascal's Wager and Finite Decision Theory. In *Gambling on God: Essays on Pascal's Wager*, ed. J. Jordan, 115-38. Boston: Rowman & Littlefield.

McKerlie, D. 1986. 'Rights and Risk.' *Canadian Journal of Philosophy,* 16: 239-51.

Miller, F. and Sartorius, R. 1979. 'Population Policy and Public Goods.' *Philosophy and Public Affairs,* 8: 148-74.

Monton, B. 2019. 'How to Avoid Maximizing Expected Utility.' *Philosophers' Imprint,* 19: 1-25.

Nebel, J. M. 2019. 'An Intrapersonal Addition Paradox.' *Ethics,* 129: 309-43.

Peterson, M. 2002. 'What Is a De Minimis Risk?' *Risk Management,* 4: 47-55.

Peterson, M. 2017. *An Introduction to Decision Theory*. 2 ed, *Cambridge Introductions to Philosophy*. Cambridge: Cambridge University Press.

Rawls, J. 1971. *A Theory of Justice*. Cambridge, MA: Harvard University Press.

Robert, D. 2018. 'Expected Comparative Utility Theory: A New Theory of Rational Choice.' *Philosophical Forum,* 49: 19-37.

Samuelson, P. A. 1977. 'St. Petersburg Paradoxes: Defanged, Dissected, and Historically Described.' *Journal of Economic Literature,* 15: 24-55.

Savage, L. J. 1954. *The Foundations of Statistics*. Vol. 11: Wiley Publications in Statistics.

Schwitzgebel, E. 2017. '1% Skepticism.' *Noûs,* 51: 271-90.

Smith, N. J. J. 2014. 'Is Evaluative Compositionality a Requirement of Rationality?' *Mind,* 123: 457-502.

Smith, N. J. J. 2016. 'Infinite Decisions and Rationally Negligible Probabilities.' *Mind,* 125: 1199-212.

Sobel, D. 2012. 'Backing Away from Libertarian Self-Ownership.' *Ethics,* 123: 32-60.

Tarsney, C. 2018. 'Moral Uncertainty for Deontologists.' *Ethical Theory and Moral Practice,* 21: 505-20.

Tenenbaum, S. 2017. 'Action, Deontology, and Risk: Against the Multiplicative Model.' *Ethics,* 127: 674-707.

Superiority Discounting Implies the Preposterous Conclusion

1. The Repugnant Conclusion

The Repugnant Conclusion (RC) is that for any number of people living great lives, it would be better to have a sufficiently greater number of people with lives that are barely worth living (Parfit 1986). Nebel (2019) shows that the Repugnant Conclusion follows from the Intrapersonal Repugnant Conclusion (IRC), alongside plausible auxiliary assumptions. This result is surprising because RC is intuitively false, but IRC is not.

The Intrapersonal Repugnant Conclusion is that for each person that might exist, it is better for them to have a guaranteed mediocre life than a sufficiently small chance of a great life and otherwise nothing. Suppose prospect Z guarantees every person who might exist a mediocre life, while prospect A guarantees those in a small subset great lives and the others nothing. According to IRC, if each person's probability of being in the lucky subset is small enough, they are better off with Z’s guaranteed mediocre life than A’s gamble on a great life.

The crucial premise in the move from IRC to RC is:

> **Weak Pareto for Equal Risk**: For any egalitarian prospects X and Y, if X is better than Y for each person who might exist in either prospect, then X is better than Y (Nebel 2019: 320).[18]

Weak Pareto gets at the intuitive idea that if some prospect is better for everyone, then it is simply better. And since prospect Z (where everyone is guaranteed mediocre lives) is better than A (where each member of a small subset receives a great life) for every person who might exist, it follows via Weak Pareto that Z is better than A. And since Z is better than A, the certain outcome of Z (a greater number of people living mediocre lives) must

[18] The principle specifically covers egalitarian prospects (those in which everyone who exists is equally well off and every person who might exist has an equal probability of existing) because prospects that are better for everyone might still be worse if they increase inequality.

be better than the certain outcome of A (a smaller number of people living great lives).[19] And this is the Repugnant Conclusion.[20]

2. Superiority Discounting

We can avoid the Repugnant Conclusion by endorsing the *superiority* of great lives over mediocre lives:[21]

> **Superiority**: There exists some pair of valuable objects *x* and *y* such that some quantity of *x* is better than any quantity of *y*.[22]

In population ethics, this position is most notably adopted by perfectionists, according to whom, "even if some change brings a great net benefit to those who are affected, it is a change for the worse if it involves the loss of one of the best things in life" (Parfit 1986: 163). Accordingly, a population consisting of any number of great lives (those containing the best things in life) is better than a population consisting of any number of mediocre lives (those without the best things in life); RC is false.

But to resist RC in light of Nebel's argument, advocates of Superiority must reject either IRC or Weak Pareto. Nebel supposes that a perfectionist might reject IRC by claiming that "even if some prospect would, in expectation, bring a net benefit to a person, it is worse for her if it lowers her probability of enjoying the best things in life" (2019: 324). This perfectionist would not accept IRC because *any* chance of a great life is better than none. Ultimately, Nebel rejects this kind of perfectionism because it results in an "absurdly reckless" decision theory that instructs agents to "prefer prospects that will almost

[19] Nebel covers this inference with the uncontroversial principle, Certainty Equivalence, which states that the certain outcome of prospect X is better than the certain outcome of Y if and only if X is better than Y (2019: 322).

[20] To be precise, the Repugnant Conclusion is that Z is better than another prospect A*. A* is identical to A except that it guarantees a specified subset of people a great life, rather than giving every person a small chance of getting a great life. Since A and A* are equally good, and Z is better than A, Z is also better than A*.

[21] Superiority appears in many areas of value theory. Prominent endorsements of Superiority include Hutcheson (1755: 118); Ross (1930: 150); Glover (1977: 710); Edwards (1979: 69-72); Griffin (1986: 85-86); Crisp (1992: 151); Lemos (1993); Mill (1998: 56); Skorupski (1999: 94-101); Brentano (2009: 106).

[22] This is *Weak Superiority* (Arrhenius and Rabinowicz 2015). *Strong Superiority* is the view that *any* quantity of *x* is better than any quantity of *y*.

certainly be worse for us in pursuit of arbitrarily small chances of enjoying the best things in life" (2019: 324).

Kosonen (2021) suggests Superiority advocates instead deny Weak Pareto by adopting *discounting*:

Discounting: Agents should discount small probabilities to zero.[23]

This *Superiority Discounter* (SD) will not chase arbitrarily small probabilities at any cost because they will ignore sufficiently small probabilities entirely. SD will endorse IRC because they will ignore the (sufficiently small) probability of any individual getting a great life under A, and judge that it is better for them to receive a guaranteed mediocre life under Z than nothing. However, even though Z is better than A for every individual, SD will not conclude that Z is (impartially) better than A because the probability that *someone* will acquire a great life is too large to ignore. (Indeed, it is certain.) And as per Superiority, a smaller number of people living great lives is better than a larger number of people living mediocre lives. So, Z is better for every person who might exist, but A is nevertheless better than Z; Weak Pareto is false.

3. The Preposterous Conclusion

While discounting steers superiority theories clear of the Repugnant Conclusion, it guides them to a more uncomfortable result. The Preposterous Conclusion is that, for an arbitrarily great number of people, near-certain hell is better than near-certain heaven (where hell is an arbitrarily horrible life and heaven is an arbitrarily great life). In the remainder of this paper, I will demonstrate how Superiority Discounting implies the Preposterous Conclusion. (In what follows, I use "prefer" to mean *judges to be better for the subject in question*.)

[23] Proposals that endorse discounting include: Bernoulli (1738); d'Alembert (1761); Borel (1962); Buffon (1777); Condorcet (1785); Kagan (1989: 89-92); Jordan (1994); Aboodi, Borer, and Enoch (2008); Hawley (2008); Haque (2012); Buchak (2013); Smith (2014, 2016); Bjorndahl, London, and Zollman (2017); Chalmers (2017); Lazar (2017); Schwitzgebel (2017); Lee-Stronach (2018); Robert (2018); Tarsney (2018); Monton (2019).

I will begin with Kosonen's paradigm case, in which SD prefers a non-negligible chance of living a great life and otherwise nothing (prospect A) to a guaranteed mediocre life (prospect Z). Then, I will alter the decision problem in ways that sweeten Z and dampen A but do not affect this preference ordering. Each alteration will make SD's continued preference for A more preposterous. The end result will be the Preposterous Conclusion.

3.1 *The Paradigm Case*

The central commitment of Superiority Discounting is:

> *Risky Non-Repugnance*: q chance (or greater) of obtaining at least one life at a high welfare level a is better than certainty of obtaining any number of lives at a low welfare level z, where q is the smallest probability that should not be discounted down to zero (Kosonen 2021: 212).

Supposing that the value of q is one in one million, this comparison is represented in Table 13. Prospect A offers a one-in-one-million chance of a great life and otherwise nothing, while Z guarantees a mediocre life.

TABLE 13

	State 1 (1 – 1/1m)	State 2 (1/1m)
A		a
Z	z	z

Since both states' probabilities are at least equal to one in one million, neither is discounted. And since A's probability of providing a great life is higher than Z's, SD will prefer A.

3.2 *Discounting Insurance*

In Table 14, we will shift some probability from State 1 into some new state, State 3, whose probability will be just low enough to ignore: 1/1m – ε (where ε is an arbitrarily small number). In State 3, A will result in nonexistence, while Z will result in a great life. So, now both prospects have a small chance of producing a great life. However, Z offers the insurance of at least a mediocre life at the cost of an arbitrarily small decrease in the probability of securing a great life.

TABLE 14

	State 1 (1 – 1/1m – (1/1m – ε))	State 2 (1/1m)	State 3 (1/1m – ε)
A		*a*	
Z	*z*	*z*	***a***

SD prefers Z to A because they will ignore State 3 and prefer A due to its higher probability of producing a great life.[24] But this preference seems wrong: Z's guarantee of a mediocre life seems well worth the price of an arbitrarily small decrease in the probability of getting a great life. It is not that SD has become absurdly reckless again: They are not chasing an arbitrarily small probability at any cost; they are chasing an arbitrarily small *increase* in probability at any cost. We invoked discounting to escape the former only to end up with the latter, and it is not clear that this result is any better.[25]

[24] We could instead cash out the discounting strategy as applying to *differences* in prospects' probabilities of producing a great life, rather than the absolute probabilities of states. So, for any prospects X and Y, if X's probability of producing a great life is sufficiently less than Y's, then Y is better than X. Unfortunately, this proposal leads to transitivity violations: If A's probability (of producing a great life) is not sufficiently greater than B's, B might be better than A; if B's probability is not sufficiently greater than C's, then C might be better than B. Nevertheless, A's probability might be sufficiently greater than C's such that A *must* be better than C. So, even though C > B and B > A, A > C.

[25] SD might respond that the strategy of inserting a state *slightly* below the threshold will not work if the threshold is vague, because whether its probability is below the threshold will be indeterminate. However (assuming truths of classical logic are determinate, and determinacy is closed under logical entailment) this will not pose a problem for the argument. SD is right that for no pair of numbers an arbitrarily small distance apart *n* and *n* – ε is it the case that determinately, *n* is not discounted but *n* – ε is. Nevertheless, it *is* determinate that *there exists* some pair of numbers an arbitrarily small distance apart *n* and *n* – ε such that *n* is not discounted but *n* – ε is. And so determinately, a state whose probability is an arbitrarily small amount less than the threshold (wherever it is) will not affect SD's preferences. (By analogy, approaches to vagueness that endorse classical logic, including epsitemicism and supervaluationism, maintain that determinately, there is a pair of numbers *n* and *n* + 1 such that with *n* hairs on one's head one is bald, but with *n* + 1 hairs one is not bald, while conceding that for no pair of numbers *n* and *n* + 1 is it the case that determinately, with *n* hairs one is bald, but with *n* + 1 hairs one is not bald.) This point applies *mutatis mutandis* to the threshold separating superior objects from inferior objects, discussed in 3.3.

3.3 *Superiority Insurance*

The discounting insurance problem arose because discounting makes agents insensitive to insufficiently probable states. But a commitment to Superiority also harbors an insurance problem because it makes agents insensitive to insufficiently *valuable* outcomes. Consequently, SD will be unwilling to buy insurance on their gamble on a great life at an arbitrarily small price to the value of that life.

SD maximizes their probability of living a great life. So, except in the event of a tie, they will be insensitive to all outcomes in which they do not acquire a great life—even if the life they acquire is arbitrarily close to the greatness threshold.[26] As a result, increasing Z's insurance payout to any value below the threshold will not alter SD's preference. For illustration, suppose that great lives are those at or above welfare level 80. In Table 15, we have increased Z's insurance payout to a life at welfare level 80 – ε. (Call lives that are slightly below the greatness threshold *very good* lives.)

TABLE 15

	State 1 (1 – 1/1m – (1/1m – ε))	State 2 (1/1m)	State 3 (1/1m – ε)
A		80	
Z	**80 – ε**	**80 – ε**	80

SD continues to prefer A to Z, despite the high probability of a very good life under Z because the probability of acquiring a *great* life remains higher under A. But now the preference of A looks very strange indeed. Given that A and Z have an approximately equal probability of producing the same great life, they should, on this basis alone, be approximately equally good. Add the fact that Z also *guarantees* a life valued approximately

[26] SD might suppose that there is no *absolute* threshold: For all welfare levels *x* and *y*, if *x* is sufficiently (e.g., 25 points) greater than *y*, then lives at *x* are superior to lives at *y*. But this, too, violates transitivity. For any number of lives at welfare level 80, there is some number if lives at 60 that would be better; and for any number of lives at 60, there is some number of lives at 40 that would be better. By transitivity, for any number of lives at 80, there is some number of lives at 40 that would be better. But this contradicts the stipulation that lives at 80 are superior to lives at 40 (in virtue of the greater-than-25-point difference in welfare levels).

equally to the one A takes a one-in-a-million gamble on, and Z should be much better (in the order of one million times better, if such quantifications make sense) than A.

3.4 *Extreme Values in Excluded States*

In 3.2, we introduced State 3 to establish the possibility of Z producing a great life. Since SD will ignore it due to its improbability, we can make the outcomes in this state arbitrarily extreme without affecting their preference. Suppose we raise the welfare level of the life Z would produce in State 3 from 80 to 100 and lower that of A's life to -100. (Call lives at welfare level 100 *heaven* and those at welfare level -100 *hell*).

In Table 16, A still offers a one-in-a-million chance of a life at welfare level 80, but it now offers approximately the same chance of hell (and otherwise nothing). Z again guarantees a very good life, but it now offers a small chance of heaven.

TABLE 16

	State 1 (1 – 1/1m – (1/1m – ε))	State 2 (1/1m)	State 3 (1/1m – ε)
A		80	**-100**
Z	80 – ε	80 – ε	**100**

SD still prefers A because they will continue to ignore State 3 due to its improbability. But intuitively, A is now worse than certain nonexistence: The small probability of a great life does not seem worth the risk of an approximately equal probability of hell. On the other hand, Z *guarantees* a very good life alongside a small chance of heaven; it is clearly (and considerably) better than certain nonexistence. Nevertheless, SD prefers A not only to certain nonexistence, but also to Z.

3.5 *Ignoring Negative Value*

Since SD maximizes the probability of acquiring a great life, their preferences are only responsive to outcomes in which they acquire a great life (except in the event of a tie). In 3.2, this feature allowed us to increase Z's insurance payout to a welfare level of just under

80 without impacting their preference. But there is another problem we can use this feature to exploit. SD does not merely ignore lives marginally below the greatness threshold; they also ignore the possibility of procuring a life that is worse than nonexistence.

In Table 17, A produces a life at -100 in State 1. As in 3.3, this problem is not the result of discounting insufficiently probable states (indeed, State 1 is the *most* probable state) but of ignoring insufficiently valuable outcomes. SD's focus on maximizing their probability of acquiring a great life makes them insensitive to this extremely bad, extremely probable outcome.

TABLE 17

	State 1 (1 – 1/1m – (1/1m – ε)	State 2 (1/1m)	State 3 (1/1m – ε)
A	**-100**	80	-100
Z	80 – ε	80 – ε	100

SD continues to prefer A because, in maximizing their probability of living a great life, they ignore the high probability of hell under A. Unsurprisingly, ignoring very bad, very probable outcomes has unfortunate results. A now offers a one-in-a-million chance of *not* ending up in hell; if they avoid hell, they will get a barely great life. Z continues to guarantee a very good life alongside a chance of heaven approximately equal to A's chance of avoiding hell. Nevertheless, SD still prefers to chase the small chance of a great life, now risking near-certain hell in its pursuit.[27]

There seems to be an intuitive fix here, whereby avoiding a *horrible* life (at a welfare level of less than or equal to -80) is also assigned superiority over non-great lives. On this version, SD will maximize their probability of living a great life *and* minimize their probability of living a horrible life. Nevertheless, the problem can be reintroduced by

[27] Parfit appeared to be aware of a related unfortunate result for perfectionism: If enjoying the best things in life takes lexical superiority over all other experiences, then perfectionists will not be very concerned with alleviating suffering. He eventually decided that the matter was extraneous since the Repugnant Conclusion does not involve suffering (Parfit 1986: 163-64).

assigning A's life in State 1 a welfare level of -80 + ε, which barely avoids the *horrible* range. The problem remains slightly watered down: SD is no longer ignoring *hell*, but merely a very bad life.

Raising the threshold above -80 would dilute the problem further but exacerbate another problem: SD will begin to care too much about disvaluable lives. To illustrate, suppose we raise the threshold all the way to 0 to ensure SD is not insensitive to *any* disvaluable lives. Now, SD will treat avoiding *any* disvaluable life as superior to very good lives; they will refuse to take a one-in-a-million risk of a life that is barely not worth living for the complement probability (near-certainty) of a very good life. So, introducing a range of bad lives that are not inferior to great lives will not escape the problem.

3.6 *Accumulation of Ignored States*

As we saw in 3.2, we can create value that flies under SD's radar by adding a state whose probability is below the threshold. But why stop at one? With each new state in the decision problem, the total probability of the discounted states will agglomerate, but SD will continue to ignore them due to their small individual probabilities. For example, suppose we add one million states to the decision problem, each almost identical to State 3. (To avoid licensing SD to lump all these new states into one, much more probable state,[28] we might assign each of Z's outcomes in these new states a *unique* value approximately equal to 100 and each of A's outcomes a unique value approximately equal to -100. For simplicity, this complication is ignored in Table 18.)

TABLE 18

	State 1 (1 – 1/1m – ((1/1m – ε) × 1m))	State 2 (1/1m)	State 3 (1/1m – ε)	. . .	State 1,000,002 (1/1m – ε)
A	-100	80	-100	**. . .**	**-100**
Z	80 – ε	80 – ε	100	**. . .**	**100**

[28] While this would bring into focus the problem of partition variance, our task here is to establish the Preposterous Conclusion.

SD still prefers A because they will ignore the new, insufficiently probable states.[29] Now, A would almost certainly result in hell, while Z would almost certainly result in heaven. For instance, if ε = one in one billion, then the probability of heaven on Z is 0.999, and the probability of hell on A is 0.999999.[30] Both prospects' remaining probabilities are assigned to State 2, in which they would perform approximately equally. (A would produce a life at 80, while Z would produce a life at 80 – ε.)

We have one more step to get to the Preposterous Conclusion, but we have arrived at what we might call:

> **The Intrapersonal Preposterous Conclusion**: Near-certain hell is better than near-certain heaven (even when both alternative outcomes are approximately equal in value).

3.7 *Populations*

So far, we have been judging what is better for a single person that might exist. The problems we have identified in sections 3.1–3.6 are magnified when we look at the value of *populations*. We established the Intrapersonal Preposterous Conclusion by making SD ignore every possible outcome except A's outcome in State 2. Since these outcomes are being ignored due to their lack of great lives or their corresponding states' improbability, increasing the number of people in these outcomes will not affect SD's preference. If it is

[29] SD can avoid this result by relativizing the threshold to the most probable state (e.g., Lee-Stronach 2018: 801). So, as State 1's probability falls (as it is redistributed), the threshold falls accordingly, and each new state will be above the new threshold. However, we can reintroduce the problem by dividing the probability of each new state by the amount required to stay under the threshold, then multiplying the number of new states by the same number. This group of problematic states will have the same total probability (since the decrease in each state's probability is offset by the addition of the new states), but each individual state will be below the threshold.

[30] It is worth noting that an approach concerned with the probability of a great life on some prospect (rather than the probability of states) would avoid these issues: The probability of Z producing a great life is 0.999, and thus presumably over the threshold (even though the probability of every state *in which* Z produces a great life is miniscule). It is unclear how such an approach would prevent insufficiently probable acquisitions of great lives from contributing to the value of the prospect (since without this feature, agents will chase arbitrarily small probabilities), but perhaps the details could be filled out plausibly. Nevertheless, careful constructions of discounting theories invariably discount states' probabilities. Kosonen does not explicitly make this distinction, but in more precise moments appears to follow suit (e.g., Kosonen 2021: 212, note 35).

preposterous to prefer A to Z when there is one person in the equation, it should be more preposterous to maintain this preference when there are more people in the equation. In Table 19, each outcome (other than A's in State 2) contains lives at the same welfare level as in Table 18, but the number of such lives is arbitrarily great (*n*). We also increase A's outcome in State 2 to ten billion: the number of great lives that is better than any number of mediocre lives. (This step is unnecessary if we are dealing with a *Strong Superiority* theory.)

TABLE 19

	State 1 (1 – 1/1m – ((1/1m – ε) × 1m))	State 2 (1/1m)	State 3 (1/1m – ε)	. . .	State 1,000,002 (1/1m – ε)
A	*n* × -100	10b × 80	*n* × -100	. . .	*n* × -100
Z	*n* × (80 – ε)	*n* × (80 – ε)	*n* × 100	. . .	*n* × 100

In preferring A to Z, SD endorses the Preposterous Conclusion:

> **The Preposterous Conclusion**: Near-certain hell for *n* people (and otherwise a small fraction living barely great lives) is better than near-certain heaven (and otherwise very good lives) for *n* people.

Just as it is difficult to imagine a worse result for a theory of well-being than the Intrapersonal Preposterous Conclusion, it is difficult to imagine a worse result for a population axiology than the Preposterous Conclusion.

4. Conclusion

Axiologies can avoid the Repugnant Conclusion by endorsing the superiority of great lives over mediocre lives. They can deny Weak Pareto by adopting discounting, allowing them to embrace the Intrapersonal Repugnant Conclusion without thereby accepting the Repugnant Conclusion. But Superiority Discounting avoids the Repugnant Conclusion only at the expense of the Preposterous Conclusion. The Superiority Discounter might patch their theory to reduce the impact of some of these issues, the result of which would be a moderately preposterous conclusion. But giving up all the features that can be exploited would simply be to give up Superiority Discounting.

References

Aboodi, R., A. Borer, and D. Enoch. 2008. Deontology, Individualism, and Uncertainty: A Reply to Jackson and Smith, *The Journal of Philosophy*, 105: 259-72.

Arrhenius, G., and W. Rabinowicz. 2015. Value Superiority. In Iwao Hirose and Jonas Olson (eds.), *The Oxford Handbook of Value Theory* (Oxford University Press), pp. 225-248.

Bernoulli, Daniel. 1738. Exposition of a New Theory on the Measurement of Risk, *Econometrica*, 22: 23-36.

Bjorndahl, A., A. J. London, and K. J. S. Zollman. 2017. Kantian Decision Making under Uncertainty: Dignity, Price, and Consistency, *Philosophers' Imprint*, 17.

Borel, Emile. 1962. *Probabilities and Life* (New York: Dover).

Brentano, Franz. 2009. *The Origin of Our Knowledge of Right and Wrong* (London: Taylor & Francis Group).

Buchak, Lara. 2013. *Risk and Rationality* (Oxford University Press).

Buffon, Georges-Louis Leclerc de. 1777. Essai D'arithmétique Morale. In *Supplement a l'histoire Naturelle* (Paris: L'Imprimerie).

Chalmers, Adam. 2017. An Offer You Can't (Rationally) Refuse: Systematically Exploiting Utility-Maximisers with Malicious Gambles, University of Sydney.

Condorcet, Jean-Antoine-Nicolas de Caritat. 1785. *Essai Sur L'application De L'analyse à La Probabilité Des DéCisions Rendues à La Pluralité Des Voix* (Paris: Imprimerie royale).

Crisp, Roger. 1992. Utilitarianism and the Life of Virtue, *The Philosophical Quarterly*, 42: 139-60.

d'Alembert, Jean Le Rond. 1761. *Opuscules MathéMatiques, Ou MéMoires Sur DifféRens Sujets De GéOméTrie, De MéChanique, D'optique, D'astronomie &C* (Paris: David).

Edwards, Ren B. 1979. *Pleasures and Pains: A Theory of Qualitative Hedonism* (Ithica, NY: Cornell University Press).

Glover, Jonathan. 1977. *Causing Death and Saving Lives* (New York: Penguin).

Griffin, James. 1986. *Well-Being: Its Meaning, Measurement, and Moral Importance* (Clarendon Press).

Haque, Adil Ahmed. 2012. Killing in the Fog of War, *Southern California Law Review*, 86: 63-116.

Hawley, Patrick. 2008. Moral Absolutism Defended, *Journal of Philosophy*, 105: 273-75.

Hutcheson, Francis. 1755. *A System of Moral Philosophy* (A.M. Kelley).

Jordan, Jeff. 1994. The St. Petersburg Paradox and Pascal's Wager, *Philosophia*, 23: 207-22.

Kagan, Shelly. 1989. *The Limits of Morality* (Oxford University Press).

Kosonen, Petra. 2021. Discounting Small Probabilities Solves the Intrapersonal Addition Paradox, *Ethics*, 132: 204-17.

Lazar, Seth. 2017. Risky Killing, *Journal of Moral Philosophy*, 16: 1-26.

Lee-Stronach, Chad. 2018. Moral Priorities under Risk, *Canadian Journal of Philosophy*, 48: 793-811.

Lemos, Noah M. 1993. Higher Goods and the Myth of Tithonus, *Journal of Philosophy*, 90: 482-96.

Mill, J. S, and R. Crisp. 1998. Utilitarianism (Oxford: Oxford University Press).

Monton, Bradley. 2019. How to Avoid Maximizing Expected Utility, *Philosophers' Imprint*, 19: 1-25.

Nebel, Jacob M. 2019. An Intrapersonal Addition Paradox, *Ethics*, 129: 309-43.

Parfit, Derek. 1986. Overpopulation and the Quality of Life. In Peter Singer (ed.), *Applied Ethics* (Oxford University Press), pp. 145-164.

Robert, David. 2018. Expected Comparative Utility Theory: A New Theory of Rational Choice, *Philosophical Forum*, 49: 19-37.

Ross, W. D. 1930. *The Right and the Good. Some Problems in Ethics* (Clarendon Press).

Schwitzgebel, Eric. 2017. 1% Skepticism, *Noûs*, 51: 271-90.

Skorupski, John. 1999. *Ethical Explorations* (Oxford University Press).

Smith, Nicholas J. J. 2014. Is Evaluative Compositionality a Requirement of Rationality?, *Mind*, 123: 457-502.

——. 2016. Infinite Decisions and Rationally Negligible Probabilities, *Mind*, 125: 1199-212.

Tarsney, Christian. 2018. Moral Uncertainty for Deontologists, *Ethical Theory and Moral Practice*, 21: 505-20.

Filtered Maximization

1. Absolutism

Deontology is the view that moral duties constrain which acts we may perform. For instance, agents with a duty to avoid killing are prohibited from killing, even if doing so would save multiple lives. A moderate deontological theory allows these constraints to be permissibly violated if the consequences are extreme enough: Perhaps, for example, preventing the extinction of humanity would justify a killing.

Moral absolutism is the most extreme brand of deontology, according to which (in a context of certainty) consequentialist considerations *never* justify a duty violation. However, in contexts of risk, where each act has some *probability* of violating a duty, it is unclear what absolutism recommends. Intuitively, agents should be permitted to take sufficiently small risks of violating their duties for the sake of consequentialist considerations but forbidden from taking sufficiently large risks, irrespective of the consequences. However, it is difficult to imagine how some number of good consequences could justify the small risk, yet no *larger* number of good consequences could ever justify the larger risk; there does not appear to be any straightforward value function consistent with these preferences. Some commentators have taken this difficulty to show that absolutism is unable to offer sufficient guidance in the real world (e.g., McKerlie 1986; Ashford 2003: 298; Jackson and Smith 2006, 2016; Colyvan, Cox, and Steele 2010; Huemer 2010; Fried 2012; Sobel 2012; Isaacs 2014; Holm 2016; Tenenbaum 2017; Alexander 2018). Others have attempted to extend absolutism to contexts of risk by modelling how absolutists ought to make decisions under risk (e.g., Kagan 1989: 89-92; Aboodi, Borer, and Enoch 2008; Hawley 2008; Haque 2012; Bjorndahl, London, and Zollman 2017; Lazar 2017; Lee-Stronach 2018; Tarsney 2018; Lazar and Lee-Stronach 2019; Lee-Stronach 2021).

This paper consists of three parts. In the first part, I consider four theories that have been offered in the literature as purported models of absolutism and show that each misses some crucial aspect of absolutist decision making. The second part offers a positive proposal, which I call Filtered Maximization. The name refers to the two distinct aspects of the decision procedure: filtration and maximization. Agents *filter* out impermissible acts on the basis of their insufficient expected “duty value”, irrespective of consequentialist value (where duty value measures the degree to which, in some outcome, an agent upholds their duties). Then, they *maximize* expected value among the remaining acts. In the third part, I discuss some implications of the view. I look at how it offers an accurate and fruitful model for classic absolutist theories, suggests solutions to other decision-theoretic problems, and how the theory relates to preference axioms.

2. Existing Theories

Having given only a vague characterization of what our theory will look like, it will be instructive to consider what a correct model of absolutism *must* look like. This section will introduce five desiderata—each accompanied by a case illustrating the desideratum’s importance—that our theory must satisfy. As each new desideratum is introduced, I will refine and replace a toy model derived from existing theories in the literature. Eventually, I will conclude that none of the theories are adequate.

In brief, the correct theory should prohibit agents from performing acts with a sufficiently great probability of violating a duty for the sake of consequentialist considerations (2.1) yet allow sufficiently small risks to be justified by the consequences (2.2). It should require agents to prefer, all else equal, less-severe violations of their duties (2.3) and smaller risks of such violations (2.4). And it should instruct agents to uphold higher-ranking duties when they clash with lower-ranking duties (2.5).

2.1 *Unacceptable Risk*

In addition to prohibiting acts that are *certain* to violate an absolutist duty, our theory should also forbid acts whose risk of doing so is unacceptably high.

> *Unsafe Car*: Gary is late for work, but he can only get there by driving a very unsafe car. Gary must decide whether to take the risk, given his absolutist moral duty to avoid killing people.

Even though driving Unsafe Car is not *certain* to kill someone, its sufficiently high probability of doing so should prohibit Gary from driving it to work, irrespective of the benefits of going to work. The case suggests a straightforward desideratum:

1. **Unacceptable Risk**: Sufficiently large risks of violating a duty cannot be justified by consequentialist considerations.

An intuitive way of making our absolutist *always* prefer to uphold their duties in the face of arbitrarily valuable consequences is by introducing infinite values. The following theory takes this approach:

> **Infinite Maximization**: Agents assign infinitely negative value to outcomes in which they violate an absolutist duty, then maximize expected value.

Infinite Maximization prescribes agents to never violate a duty for the sake of consequentialist considerations because no amount of consequentialist value will be greater than the (infinite) disvalue of a duty violation. Indeed, this is the most common way of modelling absolutism (e.g., Jackson and Smith 2006, 2016; Colyvan, Cox, and Steele 2010; Huemer 2010; Hayenhjelm and Wolff 2012; Hansson 2013; Bjorndahl, London, and Zollman 2017).

2.2 *Acceptable Risk*

Unfortunately, Infinite Maximization has well-documented problems. Most concerning is that it is not sensitive to an act's risk of violating a duty: Since any number multiplied by infinity amounts to infinity, any act with a positive probability of violating a duty will have infinitely negative expected value; agents will be unable to distinguish between acceptably small risks and unacceptably large risks of violating their duties (McKerlie 1986; Ashford 2003: 298; Jackson and Smith 2006, 2016; Colyvan, Cox, and Steele 2010; Huemer 2010; Fried 2012; Sobel 2012; Isaacs 2014; Holm 2016; Tenenbaum 2017; Alexander 2018; Lee-Stronach 2018: 796-99; 2021: 335).

> *Safe Car*: In addition to Unsafe Car, Gary has the option of driving a well-maintained car with no safety issues.

Again, even though driving Safe Car *could* kill someone,[31] if this risk is small enough, Gary should be permitted to take it—even though the risk is greater than that of Stay Home. However, Infinite Maximization does not deliver this result. If the outcome in which Gary kills someone with Safe Car is assigned infinite disvalue, it will swamp the equation, and the expected value of driving Safe Car will be infinitely negative. The case appears to suggest a second desideratum:

> **Acceptable Risk**: Sufficiently small risks of violating a duty may be justified by consequentialist considerations.

Fortunately for the absolutist, a simple fix to Infinite Maximization will yield the right result for *Safe Car*:

Discounted Infinite Maximization: Agents assign infinite disvalue to outcomes in which they violate a duty, *discount states whose probability is sufficiently low to zero*, then maximize expected value.

Our new theory instructs agents to ignore sufficiently improbable states, such as those in which a relatively safe act would violate a duty. For instance, if the state where driving Safe Car would kill someone is sufficiently improbable, Gary will simply ignore it and act as though the car were certain to transport him to work safely.[32] As a result, while driving Unsafe Car has infinitely negative expected value (since the state where it kills someone is

[31] On some views of probability, knowledge implies probability 1 (e.g., Williamson 2000; Clarke 2013; Greco 2015). And in *Safe Car*, one could argue that Gary knows he would not kill someone on his way to the store. As a result, the probability of killing someone would be zero. To accommodate views like these, we can make the case explicitly probabilistic. For instance, suppose Gary knew that there is a 1/*n* chance that someone has secretly switched Safe Car with Unsafe Car. For sufficiently great values of *n*, driving the car should be justified.

[32] It is worth noting that whether some state is discounted may depend on how we partition states. If states are partitioned coarsely, some state's probability might be above the threshold; but if this state were described as two fine-grained states—each with half the probability of the course-grained one—then they may both be below the threshold. As a result, the theory's prescriptions may be determined by how the world is described. These partition problems can be avoided if the discounter can specify a privileged partition—one that is, in some sense, *correct*—so the prescriptions arising from *that* partition are the theory's true ones. But a further problem remains: There will be scenarios in which two cases, due to morally irrelevant details, must be partitioned differently. As a result, the theory would give two inconsistent prescriptions in two relevantly similar cases.

too probable to ignore), and Stay Home has neutral expected value, Safe Car will have positive expected value (due to its benefit of getting Gary to work).

2.3 *Violation Variance*

Discounted Infinite Maximization improves on Infinite Maximization, but it still has problems. One problem is that agents who assign infinite disvalue to outcomes in which they violate their duties will not be sensitive to the *severity* of violations.

> *Nuclear Car*: Gary has a third car in his garage in addition to Safe Car and Unsafe Car. But this car is peculiar: Its battery is being used to power the world's nuclear weapons safety system. If this car is crashed, the world's nuclear warheads would be launched, and Earth would soon become uninhabitable.

Suppose the probability of crashing Nuclear Car is the same as the probability of Safe Car killing someone; call it p. Driving either car has a p probability of killing someone, but in such a state, Safe Car would kill one person while Nuclear Car would kill billions. Seemingly, Gary should assign greater disvalue to the outcome in which he kills billions of people than the one where he kills one. But on Discounted Infinite Maximization, both outcomes are infinitely disvaluable. As a result, whatever the value of p, Gary will be indifferent between the two cars. But intuitively, irrespective of the value of p, Gary should strongly prefer to drive Safe Car than Nuclear Car. The problem is that this theory does not satisfy the following desideratum:

2. **Violation Variance**: Duty violations are treated in accordance with their severity.

Both Infinite Maximization and Discounted Infinite Maximization fail to uphold Violation Variance because every outcome containing a duty violation, irrespective of its severity, is assigned the same, infinitely negative value.

An alternative approach, taken by Lee-Stronach (2018), is to use a multi-dimensional value function that recognizes different kinds of value. Agents do not simply maximize expected value but instead maximize expected *lexicographic* value: They perform the act that maximizes (expected) satisfaction of primary considerations (call this primary value) only looking to the satisfaction of secondary considerations (secondary value) in the event of a tie. We model an agent's absolutist moral duties as their primary considerations and consequentialist considerations as secondary considerations. This approach upholds Violation Variance because more-severe duty violations will generate more negative primary value: Since the outcome where Nuclear Car is crashed (and the world becomes uninhabitable) involves more killings, its primary disvalue will be correspondingly greater than that of killing one person by driving Safe Car.

Our new model continues to require agents to discount small probabilities to zero. Otherwise, they will always take the smallest risk of violating their duties. For instance, Gary will refuse even to drive Safe Car, since Stay Home has a lower probability of killing someone and so greater primary value. And since secondary value is only consulted as a tiebreaker, Gary will keep choosing to stay home indefinitely, presumably until he starves to death, because he will refuse to raise the probability of killing someone by *any* amount. However, if Gary discounts small probabilities, he will ignore the possibility of either act killing someone, and Safe Car's secondary value will break the tie (since it would get him to work, whereas Stay Home would not). Nevertheless, Gary will *not* be willing to drive Unsafe Car because its probability of killing someone is too great to ignore. So, this new model gets the right result in each of the three cases we have looked at so far.

> **Discounted Lexicographic Maximization**: Ignore sufficiently improbable states, then maximize expected *lexicographic* value.

2.4 *Risk Variance*

Even though Discounted Lexicographic Maximization avoids assigning violations of different severity the same value, it faces a problem when translating this value into *expected* value. Specifically, agents who discount will be insensitive to differences between two acts when these differences are quarantined to low-probability states.

> *Autonomous Car*: Gary's fourth option is to take a self-driving car, which is much safer than any human-driven car. Indeed, the car's technology is so advanced that the probability of killing someone by driving Autonomous Car is approximately zero. Other than its reduced risk of killing, Autonomous Car is relevantly identical to Safe Car.

While Safe Car *was* Gary's best option, the introduction of Autonomous Car—which reduces the risk of killing at no cost—makes Safe Car clearly inferior. But on Discounted Lexicographic Maximization, Gary will be indifferent between the two. Since Gary ignores the states where either act would kill someone, he will treat both as certain to uphold his duty to avoid killing. And since both cars would perform identically in the remaining state—where either car would safely transport him to work— he will be indifferent between them. But given that Autonomous Car is considerably safer than Safe Car at no cost, Gary should not be indifferent between them.

3. **Risk Variance**: Duty violations are treated in accordance with their probabilities.

Risk Variance (4) is related to Violation Variance (3). However, where Violation Variance concerns the value assigned to outcomes (which must be in line with the severity of violations in those outcomes), Risk Variance concerns the *expected* value assigned to *acts* (which must be in line with the probability of a violation). As we saw, Discounted Lexicographic Maximization satisfies Violation Variance because more-severe violations result in more primary disvalue. But it does not satisfy Risk Variance because agents will discount sufficiently improbable states to zero, and differences in the probability of those states (or the value of the corresponding outcomes) will fail to translate into differences in *expected* value.[33]

Fortunately, there are theories of absolutism that do not rely on discounting. An approach taken by Lazar and Lee-Stronach (2019) posits a bound on consequentialist value so that the total value realized by producing consequentialist goods never exceeds that of a sufficiently probable duty violation. Suppose the value of a token duty violation is -10 million, while the value of any number of consequentialist goods is bounded at 1,000. When the value of some collection of consequentialist goods is a substantial distance from the bound, each additional token will increase its value relatively linearly. But as the bound approaches, the increase in value from each token will diminish marginally such that the

[33] This problem extends to all discounting theories, including those invoked to solve decision-theoretic problems such as the St. Petersburg Paradox (Bernoulli 1738; d'Alembert 1761; Buffon 1777; Condorcet 1785; Borel 1962; Jordan 1994: 217-18; Buchak 2013: 73-74; Smith 2014, 2016; Robert 2018; Monton 2019) and Pascal's Mugging (Jordan 1994: 218-19; Chalmers 2017; Schwitzgebel 2017: 273; Monton 2019).

total never exceeds 1,000. As a result, no amount of consequentialist value will ever exceed 1/10,000 of the disvalue of a single duty violation.

> **Bounded Maximization**: Assign a bound to consequentialist value equal to the value of the maximum permissible risk of violating a duty, then maximize expected value.

Bounded Maximization satisfies all four desiderata we have looked at so far. First, when the probability of killing someone is greater than 1/10,000, agents will refuse to take the risk for the sake of consequentialist considerations; Gary will uphold Unacceptable Risk by refusing to drive Unsafe Car. Second, when the probability of killing someone is less than 1/10,000, agents may take the risk for the sake of (sufficiently weighty) consequentialist considerations; Gary upholds Acceptable Risk by driving Safe Car. Third, since more-severe duty violations will create more disvalue, agents will, all else equal, minimize the severity of possible violations; Gary upholds Violation Variance by preferring Safe Car to Nuclear Car. And fourth, since agents will not discount small probabilities, they will avoid becoming insensitive to differences in the probability of excluded states (or the value of the corresponding outcomes); Gary upholds Risk Variance by taking Autonomous Car over Safe Car.

2.5 *Precedence*

While Bounded Maximization fixes the problems with the three preceding models, it is not the theory we are looking for. One problem concerns its ability to represent multiple layers of duties.[34] It should be possible for some absolutist duties to take precedence over others. For instance, perhaps agents have a primary duty to avoid killing, which takes precedence over a secondary duty to save lives, which takes precedence over consequentialist considerations. As a result, they would not kill for the sake of saving any number of lives and would save a life at the expense of any amount of consequentialist value. Unfortunately, even though this theory appears perfectly coherent, the Bounded Maximizer has trouble representing it.

[34] Both infinite-value models suffer from this problem, too.

Presumably, the Bounded Maximizer will model this arrangement by positing multiple bounds: a lower one for consequentialist value and a higher one for lower-ranking duty value. Suppose the value of killing someone is -10 million, and the value of saving a life is 1 million. If the upper bound of the value of saving lives is 5 million, then agents may perform an act whose probability of killing is less than 0.5, providing it is expected to save a sufficiently great number of lives. However, if an act's probability of killing is greater than 0.5, then no number of lives saved will outweigh the negative value of the possible killing. In the same vein, if the upper bound of consequentialist value is 1,000, then agents will always take greater than a 0.01 chance of saving a life, irrespective of the amount of consequentialist value they forego. These results should be broadly consistent with the absolutist's preferences. But consider the following case:[35]

> *Carbon Car*: In addition to Unsafe Car, Safe Car, Nuclear Car, and Autonomous Car, Gary has a fifth option. This car has revolutionary carbon-capture technology that substantially reduces the amount of carbon in the atmosphere when driven. Indeed, a single trip to work and back in this car is expected to save one life. However, driving Carbon Car would cost Gary $1 more than Safe Car. There are no other relevant differences.

Here, Gary faces a choice between saving a life (by driving Carbon Car) and an extra $1 (from driving Safe Car). Suppose the value of $1 is one unit of value. Clearly, Gary should drive Carbon Car. And Bounded Maximization delivers this result: Not only is $1 *one-millionth* the value of a life-saving, but *no* amount of money is worth even 1% of a life-saving. But suppose we add some details to the case:

> Gary is the president of a large charity, which is in the final stages of developing an aid plan to save 5 million lives. With Gary's approval being the only outstanding matter, simply showing up to work would save 5 million lives—whichever car he takes. Nevertheless, driving Carbon Car would save an additional life.

[35] This case is adapted from a case in Hawthorne Isaacs, and Littlejohn (unpublished).

The altered case is not relevantly different from the original: Gary should drive Carbon Car. The fact that both acts are expected to save 5 million lives does not make $1 more valuable than saving an additional life. But on Bounded Maximization, it does.

Since the value of saving lives is bounded at 5 million, the value of saving 5 million lives must be extremely close to the upper bound. As a result, the value of each additional life saved after this point will be minuscule. Indeed, after the first life is saved (whose value is 1 million), the value of each of the next 4,999,999 lives saved must be, on average, less than (approximately) 0.8 to stay below the bound. The increase in value that results from saving the 5-million-and-1st life (by driving Carbon Car) must therefore be worth considerably less than 0.8.[36] Consequently, the value of this extra life will be less than that of a single dollar. So, Bounded Maximization gives the bizarre prescription that Gary should choose the extra dollar over saving a life. The problem with Bounded Maximization is that it does not respect our final desideratum:

4. **Precedence**: Higher-ranking duties take precedence over lower-ranking duties.

2.6 *The Takeaway*

Table 20 provides a summary of how each of the four theories performs with respect to the five desiderata:

[36] Since the value of additional lives saved diminishes marginally, the value of the 50,001st life will be significantly less than the average of lives 2 through 50,000.

TABLE 20

	Unacceptable Risk (Unsafe Car)	Acceptable Risk (Safe Car)	Violation Variance (Nuclear Car)	Risk Variance (Autonomous Car)	Precedence (Carbon Car)
Infinite Maximization	✓	×	×	×	×
Discounted Infinite Maximization	✓	✓	×	×	×
Discounted Lexicographic Maximization	✓	✓	✓	×	✓
Bounded Maximization	✓	✓	✓	✓	×

It seems that none of these theories are adequate. But we now have an idea of what we are looking for. The correct theory will instruct agents to never perform acts with a sufficiently great probability of violating a duty, irrespective of the consequences (as in *Unsafe Car*) but allow sufficiently small risks to be justified by the consequences (as in *Safe Car*). It will tell agents to minimize the severity of duty violations (as in *Nuclear Car*) without allowing them to ignore small probabilities (as in *Autonomous Car*). And it will instruct agents to uphold higher-ranking duties when they clash with lower-ranking duties (as in *Carbon Car*).

3. Filtered Maximization

In slogan form, deontologists believe that the right is prior to the good. Absolutists endorse the most extreme version of this principle, on which rightness is never compromised for goodness. The idea is that our duties constrain how our acts may produce goodness, but goodness is never a constraint on our duties. Filtered Maximization—the theory I will defend—takes this principle seriously. The name refers to the two distinct processes by which absolutists make decisions. First, agents filter out acts that do not sufficiently uphold their duties. Then, having fulfilled their commitment to the right, they are permitted to maximize the good by bringing about the best possible state of affairs.

To determine which acts uphold an agent's duties sufficiently, we must introduce a value function corresponding to the "duty value" of outcomes. An outcome's duty value is determined by the degree to which the agent upholds or violates their duties in that outcome. For instance, the negative duty value of some outcome x_1 in which one person would be killed would be half that of another outcome x_2 in which two people are killed (assuming the value of killings is linear).

We can speak loosely of "the value of a killing" as the difference between the value of the status quo and the value of the status quo with an additional killing (stripped of any consequentialist value that might go along with the killing). If the value of one killing is -10 million, we will assign x_1 a duty value of -10 million and x_2 a duty value of -20 million. Of course, it might be that some killings are greater duty violations than others, in which case, they will be assigned greater disvalue. For instance, the severity of a violation of the duty to avoid killing might track the degree to which the victim's life is shortened. For simplicity, these complications will be put aside in this paper.

But we need this duty value function to capture the outcome's performance with respect to multiple duties since agents might have more than one absolutist duty. Suppose agents not only have a duty not to kill but also have an equal-ranking duty to keep promises. Suppose the value of breaking a promise is -10,000. In that case, the duty value of an outcome in which one promise is broken (and no one is killed) will be -10,000, the duty value of an outcome with two broken promises will be -20,000, and the duty value of an outcome in which two people are killed and three promises are broken will be -20.03 million.

Finally, in the same way that we derive an act's expected value by weighting its outcomes' values by their respective probabilities, we can derive its expected *duty* value by weighting its outcomes' duty values by their respective probabilities. So, an act's expected duty value will be the probability-weighted average of its possible outcomes' duty value.

We now have the resources to guide agents in simple cases. Absolutists filter out acts whose expected duty value is not sufficiently great. Suppose the threshold (governing the minimum acceptable amount of expected duty value) is set at -1,000. Agents will not perform any act whose probability of killing someone is greater than 1/10,000, whose probability of killing two people is greater than 1/5,000, whose probability of breaking a promise is greater than 0.1, or any other combination of killings and broken promises whose probability-weighted average duty value is less than -1,000. After acts with insufficient duty value are filtered out, agents maximize expected value among those remaining.

4. Satisfying the Desiderata

In this section, we will test Filtered Maximization against the five desiderata presented in Section 2.

4.1 *Unacceptable Risk*

The first desideratum ensures that our theory is sufficiently restrictive: It is not only *certain* violations that are forbidden but also sufficiently *probable* ones.

(1) **Unacceptable Risk**: Sufficiently large risks of violating a duty cannot be justified by consequentialist considerations.

As we saw in *Unsafe Car*, our theory must forbid Gary from driving the dangerous car. On Filtered Maximization, Gary first assesses the duty value of each of his two options: Unsafe Car and Stay Home. Suppose the value of killing someone is -10 million, and the threshold is set at -1,000; acts with greater than a 1/10,000 probability of killing someone are absolutely prohibited. Presumably, Stay Home will easily clear this mark since the probability of killing someone by staying home will be approximately zero. However, the probability of killing someone by driving Unsafe Car is greater than 1/10,000, so its expected duty value will be less than -1,000. Gary will exclude driving Unsafe Car from his options and perform the only remaining act: Stay Home.

4.2 *Acceptable Risk*

Our second desideratum ensures our theory is permissive enough: We do not need to be *certain* that performing some act would uphold our absolutist commitments, but we must be sufficiently confident.

(2) **Acceptable Risk**: Sufficiently small risks of violating a duty may be justified by consequentialist considerations.

In *Safe Car*, we added a third option to the decision problem: Safe Car. So, Gary first determines whether Safe Car's expected duty value is great enough to join Stay Home as a viable option or small enough to be filtered out alongside Unsafe Car. Suppose the probability of Safe Car killing someone on Gary's drive to work is 1 in 1 million. (And ignore, for simplicity, the possibility of killing multiple people.) The duty value of driving Safe Car will then be -10, which is well above the -1,000 threshold.

Since Gary has two acts that survive filtering, he must perform the one that maximizes expected value. There are two relevant differences between the acts: While Stay Home has a smaller probability of killing someone, Safe Car allows Gary to get to work. Providing the added value of getting to work is greater than -10, Gary ought to drive Safe Car, despite its increased risk of violating his duty.

4.3 *Violation Variance*

The third desideratum requires our theory to be sensitive to the severity of possible violations.

(3) **Violation Variance**: Duty violations are treated in accordance with their severity.

In *Nuclear Car*, Gary has a third car to choose from, which has the same probability of killing someone as Safe Car, but which threatens to kill many more people. Gary's first task is to determine whether the option of driving Nuclear Car should be filtered out. We know that the probability of crashing Nuclear Car is 1 in 1 million (since it is equal to Safe Car's risk of killing) and that in such an event, 8 billion people will be killed. Since the value of each killing is -10 million, the duty value of crashing the car will amount to 8 billion × -10 million. Weighted by its (1 in 10 million) probability, the expected value of Nuclear Car is -8 billion—well below the -1,000 threshold. So, Gary should filter this act alongside Unsafe Car, then maximize expected value by driving Safe Car.

4.4 *Risk Variance*

The fourth desideratum requires our theory to remain sensitive to differences between low-probability states.

(4) **Risk Variance**: Duty violations are treated in accordance with their probabilities.

In *Autonomous Car*, Gary must choose between driving Autonomous Car and Safe Car. As we saw, the risk of killing someone in either car is low enough to be justified by consequentialist considerations. However, since Autonomous Car's risk is virtually nil—call it 1 in 10 billion—with no cost, Gary should clearly drive Autonomous Car. On Filtered Maximization, both options exceed the -1,000 threshold because Safe Car's expected duty value is -10, while Autonomous Car's is -0.01. Then, when maximizing expected value, Gary's choice is simple: Both acts are identical, except that Autonomous Car has a lower probability of producing the same disvaluable outcome (in which he kills someone). So, Gary maximizes expected value by taking Autonomous Car.

5. Refinements

Filtered Maximization, thus articulated, comfortably upholds the first four desiderata. However, upholding the last one requires filling in some details. In cases where an agent has multiple layers of duties, the filtration needs to be replicated for each tier (with the number of tiers depending on substantive moral facts). Considerations on the lowest-ranking tier will be consequentialist considerations; those on every higher-ranking tier will be absolutist duties, with higher-ranking duties taking precedence over lower-ranking duties. Each tier of considerations will have a corresponding value function: The primary duty value function will correspond to primary duty value, which will measure the degree to which some outcome upholds the considerations on the highest tier; the secondary duty value function will correspond to secondary duty value, which measures the degree to which some outcome upholds considerations on the top *two* tiers; the ternary duty value function will correspond to ternary duty value, which measures satisfaction of considerations on the top *three* tiers, and so on. The lowest-ranking value function will be the agent's ordinary value function, identifying the total amount of value—both deontological and consequentialist—in any given outcome.

As before, agents filter out acts that do not have sufficiently great expected primary duty value. However, they will then repeat this process for secondary duty value, ternary duty value, quaternary duty value (and so on) until they have run the procedure for every duty tier in their deliberations. Then, they will maximize overall expected value—weighing up the value of possible duty violations against consequentialist value—among the remaining acts.

5.1 *Cumulative Filtration*

It bears emphasis that our value functions are *cumulative*: Secondary duty value does not measure an outcome's performance merely for secondary duties, but for primary *and* secondary duties; ternary duty value measures satisfaction of the *top three* tiers, rather than merely the third; and quaternary value to the *top four* tiers, rather than just the fourth. This feature is important for two reasons. First, it is what makes higher-ranking duties more powerful than lower-ranking ones. Primary duties get a chance to exclude acts irrespective of lower-ranking duties. Then, they contribute to every subsequent filtration and the eventual maximization. Secondary duties miss the first filtration but participate in every subsequent one and the eventual maximization. Ternary duties miss the first two but contribute to subsequent filtrations and maximization, and so on. And finally, consequentialist considerations do not play any role in filtration, merely contributing to maximization. In effect, every role played by some duty is *also* played by higher-ranking duties.

Without cumulative filtration, lower-ranking duties would be just as constraining as higher-ranking ones; each would have the same opportunity to exclude acts that take too great a risk of violating them. But on Filtered Maximization, the fact that lower-ranking value *contains* higher-ranking value means that, for instance, higher-ranking duties can compromise lower-ranking duties, but not vice versa: Increased satisfaction of primary duties will increase an outcome's secondary duty value, but increased satisfaction of secondary duties will have no effect on its primary duty value.

The second benefit of cumulative value functions is that they prevent agents from discriminating between two acts because of a small difference in their satisfaction of *secondary* duties in cases where there is a larger difference in their satisfaction of *primary* duties. For instance, suppose a_1 has a *barely* permissible risk of violating the primary duty, while a_2 has no risk at all. Then, if a_1 again has a *barely* permissible risk of violating the secondary duty, while a_2 has a *barely impermissible* risk of violating the secondary duty, a_2 will be filtered out while a_1 will survive: The agent will prefer a small increase in satisfaction of the secondary duty at the expense of a *larger* increase in satisfaction of the primary duty, even though primary duties are supposed to take precedence over secondary duties. But when secondary duty value measures the satisfaction of primary *and* secondary duties, the difference in primary value manifests as a difference in secondary value, too.[37]

5.2 *The Decision Procedure*

We can now state the final decision procedure precisely: Agents should perform some act iff that act survives filtration and maximizes value. An act survives filtration iff, for every duty value function and corresponding threshold, its expected duty value is equal to or greater than the threshold. An act maximizes value iff its expected value is at least as great as that of every other act that survives filtration. Performing an act that survives filtration is permissible; performing the one that maximizes value is supererogatory.

5.3 *Precedence*

Our fifth and final desideratum requires our theory to accommodate multiple layers of duties, where some have priority over others.

(5) **Precedence**: Higher-ranking duties take precedence over lower-ranking duties.

[37] It should be noted that, on the view I have presented, equal-ranking duties can trade-off against each other: Agents might be permitted to perform a killing *if* they have an equal-ranking duty to save lives, and doing so would save a sufficiently large number of lives. So, it is not strictly speaking true to say that, on this view of absolutism, agents may *never* permissibly violate their duties; it is merely the case that consequentialist considerations can never justify duty violations. Absolutists opting for the stronger variant, on which duty violations are never permitted, can model their theories with a unique value function for each particular duty, rather than for each tier of duties. That way, permissible acts must have a sufficiently great amount of duty value *of every kind*. Of course, each duty's corresponding value function must remain cumulative in the sense described above: It must measure not only satisfaction of that duty, but also satisfaction of any higher-ranking duties.

Recall that Gary faced a choice between driving Carbon Car, which would save a life, and driving Safe Car, which would save a dollar. Since his duty to save lives takes precedence over his consideration to save money, it was obvious that Gary ought to save the life. However, when we added in the detail that driving *either* car would save at least 5 million lives—but Carbon Car would save one additional life—Bounded Maximization was unable to instruct Gary to save the life due to the diminishing marginal gains from saving lives.

On Filtered Maximization, Gary first assesses each option's expected primary duty value. Again, suppose the value of killing someone is -10 million, and the threshold is set at -1,000. Since Safe Car and Carbon Car each has a 1 in 10 million probability of killing someone, each has an expected primary duty value of -1. So, both acts survive the first filter.

Now, suppose again that the value of saving a life is 1 million. Since Safe Car is expected to save 5 million lives, each valued at 1 million, its expected secondary value will be 5 trillion. (We need to subtract one unit of value to account for its primary duty value, but since the number "5 trillion" is easier to discuss than "4,999,999,999,999", put this aside for now.) Since Carbon Car is expected to save 5 million and 1 lives, its expected value is 5 trillion and 1 million. Wherever the threshold is set, Gary will drive Carbon Car. If the threshold is greater than 5 trillion, Safe Car will be filtered out, and Gary will drive Carbon Car. But if it is less than (or equal to) 5 trillion, both acts will survive the second filter and proceed to the maximization stage. When maximizing expected value, Carbon Car's 1 million units of extra value (from its expectation of saving an additional life) will be greater than Safe Car's one unit of extra value (from its saving $1). So, Filtered Maximization delivers the result that Gary should save the extra life by driving Carbon Car.

6. Discussion

At this point, I have completed the tasks set out in Section 1. In Section 2, I laid out five desiderata for a risk-apt theory of moral absolutism, showing that none of the four discussed theories can satisfy all five. In Section 3, I introduced a new theory: Filtered Maximization. Then, in Sections 4 and 5, I illustrated how Filtered Maximization satisfies the five desiderata. In this section, I draw some connections to classic absolutist theories and illustrate how Filtered Maximization is a useful way of framing these views. Then, in Section 7, I discuss how Filtered Maximization relates to standard decision-theoretic axioms.

Filtered Maximization harbors very little by way of commitment to substantive moral facts. The decision procedure accurately describes consequentialist deliberation since, in the absence of absolutist duties, agents will skip filtration and proceed directly to maximization. The specific decisions the consequentialist will make depend on how they distribute value. (E.g., utilitarians will maximize wellbeing (or preference satisfaction), egoists will maximize their wellbeing, and objective list theorists will maximize a broader class of value.)

Our theory also accurately describes straightforward absolutist theories, on which one layer of duties constrains our acts, but no duties constrain each other. Such views commonly arise from views positing a set of fundamental rights possessed by persons, which constrain how they may be treated. This kind of view is famously expressed in Nozick (1974: ix), who viewed, for instance, appropriating someone's (legitimately acquired) property as absolutely forbidden, irrespective of the possible consequences of doing so. On Filtered Maximization, we model Nozick's theory with two tiers of considerations: The first tier contains the duty prohibiting appropriating others' property, with the second tier containing consequentialist considerations.

Compare Nozick's theory to Rawls's (1971), who argues for a lexicographical ordering of several principles of justice. Rawls believed that securing an expansive set of basic liberties for all members of society (the Liberty Principle) takes absolute precedence over arranging social and economic institutions to the greatest benefit of the least advantaged members of society (the Difference Principle). The Liberty Principle is placed on a tier above the Difference Principle on our model. However, unlike Nozick's theory, Rawls's must be modelled with more than two tiers to prevent our assistance to the least advantage being compromised by other considerations, such as assistance to the upper class: a very anti-Rawlsian result! Instead, we need one tier for the Liberty Principle, a second tier for the Difference Principle, and at least one more tier for the consideration to help the upper class.

Our framework's flexibility also gives us options for how constraining we want to make different absolutist duties. There are two mechanisms by which we can alter a duty's strength. First, as we have seen, some duties can take precedence over others: If the duty to avoid killing is ranked above the duty to avoid breaking promises, then agents may compromise on the latter for the sake of the former, but not vice versa. Second, we can place two duties on the same tier but make one stronger than the other via the agent's duty value function. For instance, if the value of a killing is -10 million and the value of a broken promise is -10, agents will be far more sensitive to possible killings than broken promises (both in their filtering and maximizing), even though both duties occupy the same tier.

Our historical catalogue of absolutists includes figures who might have benefitted from this flexibility. For instance, one of Kant's most enduring criticisms concerns how constraining his moral duties are. Famously, Kant believed it was always wrong to lie—even to a would-be murderer who asks the whereabouts of your friend. In our framework, the duty to avoid lying might be an absolutist moral duty, a violation of which is never justified by consequentialist considerations. Yet, this duty might be subordinate to other duties (either by precedence, strength, or both), such as the duty to avoid abetting would-be murderers. For instance, suppose the duty value of abetting the murderer by giving away your friend's whereabouts is -100,000, and that of lying is -100. While you would not lie for any consequentialist considerations, you might lie to prevent a 1/1,000 probability of your friend being murdered.

In addition to shedding light on historical absolutist theories, Filtered Maximization has interesting applications to various areas. First, there is an outstanding question of what agents are to do when all their available acts are filtered out; we need a non-ideal decision procedure to guide agents in cases where they cannot help but violate a duty (or an explanation of why these cases cannot arise). Second, the fact that many purported duties are vague introduces challenges to any deontological theory: To assess the probability of a duty violation, the theory must know which outcomes *are* violations. (For instance, some acts that would shorten someone's life by a trivial amount are not killings, but if they shortened it by a sufficiently great amount, *ceteris paribus*, they would be a killing.) Fortunately, the theory's utilization of "duty value"—which can measure violations gradationally—makes it particularly apt to deal with this problem. Finally, it is worth flagging that discounting has also been endorsed to solve decision-theoretic problems, like the St. Petersburg Paradox and Pascal's Mugging. Indeed, the approach here suggests what a successful solution to these problems might look like: Agents filter out acts that do not have a sufficiently high probability of securing some amount of money (e.g., the act of paying a sufficiently great amount of money to play the St. Petersburg Game), before picking the remaining one that maximizes expected value (i.e., the lowest available price that has not been filtered). I hope to say more about the theory's extension to these problems in future work.

7. Filtered Maximization and Decision-theoretic Principles

In this section, I will discuss three principles that Filtered Maximization violates. The first is Independence, a common preference axiom; the second is Dominance, a popular principle implied by most axiomatizations; the third is Forbidden Randomization, which I introduce. With each violation, we will seek to answer two questions:

(1) Is the violation reasonable by the absolutist's lights? And if not,

(2) Can absolutists avoid the violation?

The first determines whether the absolutist should be concerned about the violation or whether it is just a feature of absolutism. If we determine there *is* cause for concern, we should ask whether absolutists could conceivably avoid the violation. If not, then the case will be a problem for absolutism. But if they can, then the case may merely be a problem for Filtered Maximization (as a purported model of absolutism), not necessarily for absolutism itself.

7.1 *Independence*

Decision-theoretic axiomatizations provide a set of axioms that guarantee that, if an agent conforms to those axioms, there exists some probability and value function such that the agent can be described as maximizing expected value. Notice that, on Filtered Maximization, the absolutist cannot be described as if they are maximizing expected value:

There is no pair of value and probability functions consistent with the filtration process.[38] By *modus tollens*, then, Filtered Maximizers violate at least one axiom in any such axiomatization.

Determining which axioms are violated by Filtered Maximizers will tell us something about absolutism, Filtered Maximization (as a model of absolutism), and the axiom itself. If it violates a plausible axiom, we may have a strong reason to reject either absolutism or Filtered Maximization. But the more controversial the axiom, the greater reason absolutists may have to reject that axiom.

Our discussion will proceed in terms of von Neumann and Morgenstern's (1944) axiomatization, which posits the following axioms: Completeness, Transitivity, Continuity, and Independence. Filtered Maximization appears to uphold the first three. Completeness is upheld because the filtration process splits the available acts into an upper and a lower tier; all acts that survive filtration are better than those that do not.[39] Within tiers, acts are ordered by their overall expected value.[40] Transitivity is upheld because there is no possible scenario in which our absolutist would have a cycle in their preference ordering. And Continuity will be upheld because agents will be willing to take *some* but not *any* risk of violating their duties. However, as I will show, Independence is violated.

Independence states the following (where ApB means A with probability p and otherwise B):

Independence: A $\succ$ B if and only if ApC $\succ$ BpC

Independence implies that we can determine which act is better without looking at the states in which both acts agree (that is, produce the same outcome). It is to von Neumann and Morgenstern's theory what the Sure Thing Principle is to Savage's (1954) theory.

[38] Modelling absolutism as maximization of expected value requires preventing consequentialist value from ever exceeding the disvalue of a duty violation. We can do this in one of two ways: Bound the value of consequentialist outcomes or treating value as a vector quantity and maximize expected lexicographical value. As we have seen, both approaches violate at least one desideratum.

[39] When multiple tiers are in play, acts are ordered by the stage at which they were filtered out.

[40] Actually, I have not committed to a particular way of ordering filtered acts. We might instead order them lexicographically (that is, based on duty value and only looking to consequentialist value to break ties); or perhaps their ordering would require another layer of filtration. The point is merely that we have no reason to think that we could not state a complete ordering of acts from the resources provided by Filtered Maximization.

Perhaps fortunately for the absolutist, both axioms (Independence and Sure Thing) are the most controversial of each theory. However, this controversy is primarily because they imply Allais' paradox. And since there is no obvious relationship between absolutism and Allais' paradox, it is unlikely that absolutists can reject Independence on the same grounds.

First, suppose you must choose between A and B in Table 21. Let *Good* refer to an outcome containing some quantity of consequentialist value, *Kill* refer to a duty violation, and *Bad* refer to some quantity of negative consequentialist value.

TABLE 21

	State 1 (0.94)	State 2 (0.04)	State 3 (0.02)
A	Good	Bad	Kill
B	Good	Kill	Kill

Suppose the negative duty value of a killing is -100, and t = -5. So, acts whose probability of killing one person is higher than 0.05 are filtered out. Since A has a 0.02 probability of killing someone, it is permissible; since B has a 0.06 probability of killing someone, it is impermissible. So, according to Filtered Maximization, you ought to A.

Independence tells us that we can make this determination—that you ought to A—without looking at State 3 since both acts agree in this state. However, without State 3, the decision problem looks as follows:

TABLE 22

	State 1 (0.94/0.98)	State 2 (0.04/0.98)
A	Good	Bad
B	Good	Kill

In this decision problem, the probability of A killing someone is zero, and that of B killing someone is just over 0.04; both probabilities are sufficiently low that neither act is filtered out, and the decision will come down to which act has the greatest expected value. If the negative value of Bad is sufficiently great, then it will outweigh that of Kill, and we ought to B. So, even though the acts agree in State 3, excluding this state will change the theory's prescription.

(We can see precisely why this case violates Independence if we let A refer to Good with probability 0.94/0.98 and otherwise Kill, B refer to Good with probability 0.94/0.98 and otherwise Bad, and C refer to Kill in the following:

Independence: A ≻ B iff ApC ≻ BpC

According to Independence, A is better than B if and only if A*p*C is better than B*p*C. Independence is violated because A is better than B (as demonstrated in Table 22), but A*p*C is worse than B*p*C (as demonstrated in Table 21).)

I am inclined to think that the violation of Independence is not problematic. We know that absolutists must permit sufficiently small risks of violating their duties but have the absolute constraints kick in for sufficiently large risks. So, when excluded states determine whether the size of the risk is permissible or impermissible, it seems obvious that we cannot determine the right course of action without looking at those states. As a result, by the lights of the absolutist, this violation should not be particularly worrying.

7.2 *Dominance*

The Dominance principle instructs agents not to perform acts that would produce a worse outcome than some other act in every possible state. Or, more precisely:

Dominance: If A's outcome is at least as good as B's outcome in every state and better in at least one state, then A ≻ B.

Suppose you aim a gun with 100 chambers at Amber; you know that precisely one chamber is loaded. If Chamber 1 is loaded, pulling the trigger (*Pull*) will result in you killing her, and not pulling the trigger (¬*Pull*) will result in many people dying—but you would not have killed them. If any other chamber is loaded, Pulling will give Amber a small fright, and ¬Pulling will do nothing.[41]

[41] This case is thanks to Zachary Goodsell.

TABLE 23

	Chamber 1 (0.01)	Other Chamber (0.99)
Pull	Kill	Scare
¬Pull	Lots of Deaths	Nothing

Given each act has a sufficiently low number of expected killings, they both survive filtration. Then, you perform the act whose expected value is greatest. If the number of deaths in Lots of Deaths is sufficiently high, then its negative value will outweigh that of Kill, so you ought to Pull. However, ¬Pull dominates Pull: If you were certain that the actual state is Other Chamber, then you ought to ¬Pull (since it is better to avoid scaring Amber, all else equal); if you were certain that the actual state is Chamber 1, then you ought to ¬Pull (since no consequentialist goods can justify a certain killing).

Is Filtered Maximization faithfully representing the preferences of the absolutist? I believe so. It is difficult to see how any theory of absolutism will avoid dominance violations. The absolutist must allow small risks of killing to be traded off against consequentialist value. (And they cannot ignore this consequentialist value just because it occupies low-probability outcomes, or they will face the same problems as Discounted Infinite Maximization and Discounted Lexicographic Maximization.) In non-dominance cases, these features do not seem problematic. For instance, suppose that Lots of Deaths would occur if you ¬Pull and Chamber 67 is loaded.

TABLE 24

	Chamber 1 (0.01)	Chamber 67 (0.01)	Other Chamber (0.98)
Pull	Kill	Scare	Scare
¬Pull	Nothing	Lots of Deaths	Nothing

In this case, Pulling does not seem to be problematic. After all, the probability of killing is low enough that consequentialist considerations may justify the risk. And if the number of deaths in Lots of Deaths is sufficiently great, then clearly, we ought to risk the killing to avoid Lots of Deaths. Nevertheless, the only difference between Table 24 and Table 23 (in which Dominance is violated) is that Lots of Deaths would occur in Chamber 67 instead of Chamber 1. It is difficult to see on what grounds the absolutist could prefer

¬Pull in Table 23 but Pull in Table 24: Surely whether it is permissible to Pull does not depend on whether Lots of Deaths would occur if Chamber 1 or Chamber 67 is loaded, given they have the same probability. So, it seems that absolutists are destined to violate Dominance.

So, is this a problem for the absolutist? I think there are reasonably strong arguments for both sides. On the one hand, it seems absurd that the absolutist would ¬Pull if they were certain of the state of the world, no matter which state they were in, but Pull if they were uncertain. It is as though they *know* that if they asked God what to do, then God would tell them to ¬Pull no matter what the world is like, but nevertheless, they ought to Pull.

On the other hand, perhaps treating this case as a dominance violation is not the right approach by the absolutist's lights. Recall that our dominance principle tells us that A is better than B if *A's outcome is better in every state*. But absolutists deny that they ought to maximize (expected) value, so *betterness* here cannot simply refer to the value of the outcomes concerned. Instead, we presumably must spell out *better* in terms of the absolutist's preferences among acts that would certainly produce these outcomes. So, it is merely the case that *if* I were certain of Chamber 1, then I would ¬Pull, and *if* I were certain of Other Chamber, then I would ¬Pull.

It is not clear that violating *this* notion of Dominance is so problematic for the absolutist. After all, if I were certain of Chamber 1, I would be prohibited from Pulling, and if I were certain of Other Chamber, there would be no incentive to Pull. But given that I am uncertain about the state, this prohibition does not apply to me, and there *is* an incentive to Pull: avoiding Lots of Deaths. And if I do not have a prohibition against Pulling, and there is a *strong enough* incentive to Pull, then why shouldn't I?

7.3 *Forbidden Randomization*

The third problem arises when agents randomize over forbidden acts. Randomizing means allowing the state of the world to determine which act you perform: For instance, in the choice between A and B, you might decide to do A if a coin toss comes up heads and B if it comes up tails. Suppose you point a 100-chamber gun at Amber, and all but one chamber is loaded.

TABLE 25

	Chamber 1 (0.01)	Other Chamber (0.99)
Pull	Nothing	Kill
¬Pull	Lots of Deaths	Nothing

According to the absolutist, Pull is forbidden, irrespective of the number of deaths it could avoid.

Now, pick some *n* great enough that taking a 0.99×0.5^n risk of killing someone may be justified by consequentialist considerations, and suppose you decide to toss *n* coins and Pull if and only if they all come up heads. Call this act *Flip*.

TABLE 26

	Chamber 1 & All Heads	Chamber 1 & Not All Heads	Other Chamber & All Heads	Other Chamber & Not All Heads
Pull	Nothing	Nothing	Kill	Kill
¬Pull	Lots of Deaths	Lots of Deaths	Nothing	Nothing
Flip	Nothing	Lots of Deaths	Kill	Nothing

By stipulation, *n* is sufficiently great that the probability of all *n* coins coming up heads is low enough that consequentialist considerations may justify Flipping. So, if enough people would die in Lots of Deaths, you ought to Flip.

Note that this is not a dominance problem: Flip is not being dominated by either Pull or ¬Pull. (It is better than Pull in the fourth state and better than ¬Pull the first state.) However, given that pulling the trigger will near-certainly kill someone, it seems blatantly against the spirit of absolutism to do it—and a large number of coin tosses coming up heads would not make it permissible. Flipping violates the following principle:

Forbidden Randomization: Randomizing over forbidden acts is forbidden.

Forbidden Randomization tells us that if A is forbidden and B is permitted, you are not allowed to toss a coin to determine whether you will A or B; you should just B!

Again, this feature does not appear to be peculiar to Filtered Maximization. It is difficult to see how an absolutist is to avoid violations of Forbidden Randomization, given that consequentialist considerations can justify small risks, and large risks can be converted into small risks by randomizing. However, whereas the first case (in 7.1) does not seem to pose a problem for the absolutist, and it was at least arguable whether the second case (7.2) does, it seems difficult to justify violating Forbidden Randomization.

References

Aboodi, R., Borer, A., and Enoch, D. 2008. 'Deontology, Individualism, and Uncertainty: A Reply to Jackson and Smith.' *The Journal of Philosophy,* 105: 259-72.

Alexander, L. 2018. 'The Need to Attend to Probabilities—for Purposes of Self-Defense and Other Preemptive Actions.' *San Diego Law Review,* 55: 223-29.

Ashford, E. 2003. 'The Demandingness of Scanlon's Contractualism.' *Ethics,* 113: 273-302.

Bernoulli, D. 1738. 'Exposition of a New Theory on the Measurement of Risk.' *Econometrica,* 22: 23-36.

Bjorndahl, A., London, A. J., and Zollman, K. J. S. 2017. 'Kantian Decision Making under Uncertainty: Dignity, Price, and Consistency.' *Philosophers' Imprint,* 17.

Borel, E. 1962. *Probabilities and Life*. New York: Dover.

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

Buffon, G.-L. L. d. 1777. Essai D'arithmétique Morale. In *Supplement a l'histoire Naturelle*. Paris: L'Imprimerie.

Chalmers, A. 2017. An Offer You Can't (Rationally) Refuse: Systematically Exploiting Utility-Maximisers with Malicious Gambles, University of Sydney.

Clarke, R. 2013. 'Belief Is Credence One (in Context).' *Philosophers' Imprint,* 13: 1-18.

Colyvan, M., Cox, D., and Steele, K. S. 2010. 'Modelling the Moral Dimension of Decisions.' *Noûs,* 44: 503-29.

Condorcet, J.-A.-N. d. C. 1785. *Essai Sur L'application De L'analyse à La Probabilité Des DéCisions Rendues à La Pluralité Des Voix*. Paris: Imprimerie royale.

d'Alembert, J. L. R. 1761. *Opuscules MathéMatiques, Ou MéMoires Sur DifféRens Sujets De GéOméTrie, De MéChanique, D'optique, D'astronomie &C*. Paris: David.

Fried, B. H. 2012. 'What Does Matter? The Case for Killing the Trolley Problem (or Letting It Die).' *The Philosophical Quarterly,* 62: 505-29.

Greco, D. 2015. 'How I Learned to Stop Worrying and Love Probability 1.' *Philosophical Perspectives,* 29: 179-201.

Hansson, S. O. 2013. *The Ethics of Risk: Ethical Analysis in an Uncertain World*: Palgrave Macmillan.

Haque, A. A. 2012. 'Killing in the Fog of War.' *Southern California Law Review,* 86: 63-116.

Hawley, P. 2008. 'Moral Absolutism Defended.' *Journal of Philosophy,* 105: 273-75.

Hayenhjelm, M. and Wolff, J. 2012. 'The Moral Problem of Risk Impositions: A Survey of the Literature.' *European Journal of Philosophy,* 20: e26-e51.

Holm, S. 2016. 'A Right against Risk-Imposition and the Problem of Paralysis.' *Ethical Theory and Moral Practice,* 19: 917-30.

Huemer, M. 2010. 'Lexical Priority and the Problem of Risk.' *Pacific Philosophical Quarterly,* 91: 332-51.

Isaacs, Y. 2014. 'Duty and Knowledge.' *Philosophical Perspectives,* 28: 95-110.

Jackson, F. and Smith, M. 2006. 'Absolutist Moral Theories and Uncertainty.' *Journal of Philosophy,* 103: 267-83.

Jackson, F. and Smith, M. 2016. The Implementation Problem for Deontology. In *Weighing Reasons*. New York: Oxford University Press.

Jordan, J. 1994. 'The St. Petersburg Paradox and Pascal's Wager.' *Philosophia,* 23: 207-22.

Kagan, S. 1989. *The Limits of Morality*: Oxford University Press.

Lazar, S. 2017. 'Risky Killing.' *Journal of Moral Philosophy,* 16: 1-26.

Lazar, S. and Lee-Stronach, C. 2019. 'Axiological Absolutism and Risk.' *Noûs,* 53: 97-113.

Lee-Stronach, C. 2018. 'Moral Priorities under Risk.' *Canadian Journal of Philosophy,* 48: 793-811.

Lee-Stronach, C. 2021. 'Morality, Uncertainty.' *Philosophical Quarterly,* 71: 334-58.

McKerlie, D. 1986. 'Rights and Risk.' *Canadian Journal of Philosophy,* 16: 239-51.

Monton, B. 2019. 'How to Avoid Maximizing Expected Utility.' *Philosophers' Imprint,* 19: 1-25.

Nozick, R. 1974. *Anarchy, State, and Utopia*. Vol. 7: Basic Books.

Rawls, J. 1971. *A Theory of Justice*. Cambridge, MA: Harvard University Press.

Robert, D. 2018. 'Expected Comparative Utility Theory: A New Theory of Rational Choice.' *Philosophical Forum,* 49: 19-37.

Savage, L. J. 1954. *The Foundations of Statistics*. Vol. 11: Wiley Publications in Statistics.

Schwitzgebel, E. 2017. '1% Skepticism.' *Noûs,* 51: 271-90.

Smith, N. J. J. 2014. 'Is Evaluative Compositionality a Requirement of Rationality?' *Mind,* 123: 457-502.

Smith, N. J. J. 2016. 'Infinite Decisions and Rationally Negligible Probabilities.' *Mind,* 125: 1199-212.

Sobel, D. 2012. 'Backing Away from Libertarian Self-Ownership.' *Ethics,* 123: 32-60.

Tarsney, C. 2018. 'Moral Uncertainty for Deontologists.' *Ethical Theory and Moral Practice,* 21: 505-20.

Tenenbaum, S. 2017. 'Action, Deontology, and Risk: Against the Multiplicative Model.' *Ethics,* 127: 674-707.

Von Neumann, J. and Morgenstern, O. 1944. *Theory of Games and Economic Behavior*: Princeton University Press.

Williamson, T. 2000. *Knowledge and Its Limits*. Oxford: Oxford University Press.

Superiority and Separability

1. Introduction

Superiority and Separability are both very intuitive axiological principles. In this paper, I show that they are incompatible. In the remainder of this section, I introduce these two principles and the argument's three structural assumptions: Continuity, Weak Agglomeration, and Transitivity. (Let *CAT* refer to these three principles.) Section 2 presents the argument for the incompatibility of Superiority and Separability (given CAT). Section 3 discusses how analogous arguments can be formulated for well-being and normative ethics. Section 4 discusses the argument's implications for views that purport to uphold both Superiority and Separability. In Section 5, I consider whether we can plausibly deny Weak Agglomeration by discounting small probabilities. I argue that the combination of Superiority and discounting creates uniquely challenging problems.

1.1 *Superiority*

According to

> **Superiority**: There exists some pair of valuable objects x and y such that some quantity of x is better than any quantity of y.[42]

Superiority is particularly attractive in cases where x is an important good and y is trivial; we do not want to say that any quantity of some trivial good is better than a large quantity of some important good. The paradigm such case is Parfit's Repugnant Conclusion:

> For any possible population of at least ten billion people, all with a very high quality of life, there must be some much larger imaginable population whose existence, if other things are equal, would be better, even though its members have lives that are barely worth living (Parfit 1984: 388).

[42] This is *Weak* Superiority (Arrhenius and Rabinowicz 2015). Weak Superiority is the view that *any* quantity of x is better than any quantity of y. I use Weak Superiority because problems for the weaker articulation will necessarily be problems for the stronger one. The distinction will not matter for our purposes.

But the Repugnant Conclusion does not seem to merely concern the way we aggregate individual well-being into the value of a population since an analogue of the Repugnant Conclusion arises in a single-person case:

> Suppose that I can choose between two futures. I could live for another 100 years, all of an extremely high quality. Call this the *Century of Ecstasy*. I could instead live for ever, with a life that would always be barely worth living. Though there would be nothing bad in this life, the only good things would be muzak and potatoes. Call this the Drab Eternity (Parfit 1986: 161).[43]

Accepting Superiority allows us to avoid (both versions of) the Repugnant Conclusion: If great lives (those containing goods that contribute to a high welfare level) are superior to mediocre lives (those containing only those goods that contribute to a low-positive welfare level), then no number of mediocre lives will be better than ten billion great lives, and no mediocre life of any length would be better than a great life of 100 years.

1.2 *Separability*

Unfortunately, as I will argue, Superiority violates Separability—the intuitive idea that in comparing two outcomes, we can ignore any people whose welfare and existence are unaffected (e.g., those in the distant past or on distant planets) (Thomas 2022). Or, more precisely:

> **Separability**: For any populations X, Y, and Z, X is at least as good as Y just in case adding X to Z would be at least as good as adding Y to Z—i.e., a population composed of X and Z would be at least as good as a population composed of Y and Z (Nebel 2022: 210)

Separability is often defended with *Egyptology Arguments*: When determining whether to have a baby, I do not need to investigate the welfare level of the ancient Egyptians. The thought is that the marginal value of my child's existence to this population could not possibly depend on the value of other, unaffected populations.

[43] A similar case is offered by McTaggart (1927: 42-53).

1.3 *Continuity*

Continuity is a preference axiom in many theories, such as von Neumann and Morgenstern's (1944) and Jeffrey's (1965). Axiomatizations that do not endorse Continuity contain another principle that serves the same purpose (e.g., Non-atomicity in Savage (1954)). Let A*p*B refer to a lottery in which one wins A with probability *p* and B with probability 1 – *p*. For some probabilities *p* and *q* strictly between 0 and 1:

> **Continuity**: If A ≻ B ≻ C, there exists some *p* and *q* such that A*p*C ≻ B ≻ A*q*C.[44]

Continuity states that if A is better than B, and B is better than C, then a gamble with a sufficiently *high* probability of A and otherwise C is *better* than B, and a gamble with a sufficiently *low* probability of A and otherwise C is *worse* than B. So, if we prefer an apple to a banana, then whether we should give up a banana for a gamble on an apple depends on the gamble's odds.

Continuity requires rational agents to take *some* risk but not *any* risk. An agent who is willing to take any risk is *Reckless*:

> **Recklessness**: Certainty of B is worse than any (nonzero) probability of A and otherwise C.

An agent who is not willing to take some risk is S*tubborn*:

> **Stubbornness**: Certainty of B is better than any (non-maximal) probability of A and otherwise C.

Distinguishing between Continuity's two requirements—to avoid Recklessness and Stubbornness—will be helpful since anti-Recklessness and anti-Stubbornness will play distinct roles in the argument.

Representation theorems require Continuity (or a functionally equivalent axiom) because if an agent is Reckless or Stubborn, they will behave the same, irrespective of the probability of getting the better outcome: Reckless agents will always take the gamble, while Stubborn agents will never. However, the best reason to accept Continuity is that Recklessness and Stubbornness are implausibly extreme positions. Once we appreciate

[44] This articulation is found in Peterson (2017: 106).

that probabilities can be arbitrarily small, it seems crazy to accept or decline a gamble that would improve or worsen our life before we even find out the odds.

1.4 *Weak Agglomeration*

The second assumption tells us that if a prospect A_1 is better than an alternative B_1 irrespective of whether we have A_2 or B_2, and A_2 is better than B_2 irrespective of whether we have A_1 or B_1, then the composite prospect A_1A_2 is better than B_1B_2.[45] Or, more generally:

> **Weak Agglomeration**: For any pair of composite prospects $A_1 \ldots A_n$ and $B_1 \ldots B_n$, if $A_1 \succ B_1$ irrespective of whether $A_2 \ldots A_n$ or $B_2 \ldots B_n$, and … and $A_n \succ B_n$ irrespective of whether $A_1 \ldots A_{n-1}$ or $B_1 \ldots B_{n-1}$, then $A_1 \ldots A_n \succ B_1 \ldots B_n$.[46]

So, if it is better for Monday to be sunny than rainy, irrespective of whether each day Tuesday through Sunday is sunny or rainy; it is better for Tuesday to be sunny (than rainy), irrespective of whether the other six days are sunny or rainy; … and it is better for Sunday to be sunny (than rainy), irrespective of whether each day Monday through Saturday is sunny or rainy, then it is better for every day this week to be sunny than for every day to be rainy.

Weak Agglomeration is a variant of Jackson's (1985: 178) *Agglomeration*, which is stronger and more controversial in the absence of the *irrespective of* qualifier. According to Jackson's version, if A_1 is better than B_1, and A_2 is better than B_2, then A_1A_2 is better than B_1B_2.[47] This stronger Agglomeration principle is rejected by *actualists* (e.g., Jackson and Pargetter 1986) due to cases like the following:)

---

[45] A_1 is better than B_1 *irrespective of* whether A_2 or B_2 iff the following conjunction holds:

1) If A_2, then $A_1 \succ B_1$, and
2) If B_2, then $A_1 \succ B_1$.

[46] This articulation is analogous to Hare's (2016: 460) articulation of a principle by the same name. Whereas Hare's principle concerns which *action* one *ought* to perform, mine concerns which *prospect* is *better*. Since actions are prospects (which we have control over) and *ought* picks out the *best* action in a class of alternatives, the translation to prospects and betterness should be unproblematic.

[47] Again, Jackson's discussion uses *ought to be*, rather than *better*. Since "what ought to be is not what is better, but what is best" (Jackson 1975: 177), there should be no trouble cashing out *ought to be* in terms of *betterness* relations.

> Attila and Genghis are driving their chariots towards each other. If neither swerves, there will be a collision; if both swerve, there will be a worse collision … but if one swerves and the other does not, there will be no collision. Moreover if one swerves, the other will not because neither wants a collision. Unfortunately, it is also true to an even greater extent that neither wants to be 'chicken'; as a result what actually happens is that neither swerves and there is a collision. It ought to be that Attila swerves, for then there would be no collision. … Equally it ought to be that Genghis swerves. But it ought not to be that both swerve, for then we get a worse collision (Jackson 1985: 189).

However, even actualists would not reject Weak Agglomeration since the *irrespective of* qualification deals with these cases: It is not the case that Attila ought to swerve *irrespective of* whether Genghis does because, if Genghis swerves, then Attila swerving would cause a collision. All the better for actualists (that they would not deny Weak Agglomeration), in my view, because it is difficult to imagine how Weak Agglomeration could be false.

1.5 *Transitivity*

The final of the CAT principles is transitivity:

> **Transitivity**: If $A \succ B \succ C$, then $A \succ C$.

Transitivity should be familiar and intuitive. It tells us that if A is better than B and B is better than C, then A is better than C.[48]

2. The Argument

In this section, I show that given CAT, Separability implies the Repugnant Conclusion. The argument proceeds by extracting three results. The first result, *Anti-recklessness*, follows

[48] Some deny transitivity (e.g., Temkin 1987; Temkin 2011; Persson 2004; Rachels 2004). However, this denial is motivated by very specific cases, such as when each of a large number of changes improves things, but the changes *collectively* worsen things (or so these authors argue). It is worth noting that the use of transitivity in the argument here does not resemble such cases, so perhaps we could get by with a weaker version that would be acceptable to these philosophers.

from Continuity, Separability, and Weak Agglomeration. The second result, *Anti-stubbornness*, follows from Separability and Continuity. The third result, *Anti-superiority*, follows from *Anti-recklessness* and *Anti-stubbornness* via Transitivity.[49]

Suppose we find a planet ($Planet_1$) with features that make it hospitable to life. It is one of two types. *Safe* planets certainly sustain 11 billion mediocre lives, while *Longshot* planets have a low probability *p* of sustaining 11 billion great lives and otherwise contain no lives. By Continuity, we know that for a sufficiently small *p*, it is better that $Planet_1$ is Safe. Call this probability $_{\text{LOW}}$, and suppose that Longshot planets have a $_{\text{LOW}}$ probability of containing 11 billion great lives. Now suppose we find a second planet ($Planet_2$) that is also either Safe or Longshot. By Separability, we know that the value of lives on $Planet_2$ does not depend on the lives on $Planet_1$. So, irrespective of whether $Planet_1$ is Safe or Longshot, it is better that $Planet_2$ is Safe. And for a third planet, it is better that $Planet_3$ is Safe, irrespective of $Planet_1$ and $Planet_2$. Iterating this reasoning for any number *n* planets gives us the result that, for each of *n* planets ($Planet_1$, $Planet_2$, …, $Planet_n$), it is better for that planet to be Safe than Longshot, irrespective of the other planets. And by Weak Agglomeration, we can conclude that for any *n*, it is better that $Planet_1$…$Planet_n$ are Safe than that $Planet_1$…$Planet_n$ are Longshot.

> **Anti-recklessness**: There is some (nonzero) *p* such that, for any (nonzero) *n*, it is better to have *n* lots of 11 billion mediocre lives than *n* independent gambles at *p* on 11 billion great lives and otherwise none.

Now, suppose we were to find a collection of *n* planets, and either all *n* planets are Longshot (and so each has an independent $_{\text{LOW}}$ probability of containing 11 billion great lives), or one planet contains 10 billion great lives and the others contain no lives. Since 11 billion great lives are better than 10 billion great lives, Continuity tells us there is some (sufficiently high) probability of 11 billion lives and otherwise none that is better than certainty of 10 billion lives. And since the probability of at least one planet containing 11 billion great lives increases as *n* increases—and arbitrarily large values of *n* bring this probability arbitrarily close to 1—there is some *n* such that it is better for all *n* planets to

[49] I am thankful to Remco Heesen for helping me work out the formal details of this argument.

be Longshot than for one to contain 10 billion great lives for certain.[50] Call an *n* that witnesses this observation $_{\text{BIG}}$.

> **Anti-stubbornness**: For any (nonzero) *p*, there is some *n* such that it is better to have *n* independent gambles at *p* on 11 billion great lives and otherwise none than 10 billion great lives for certain.

From Anti-recklessness, we know that $_{\text{BIG}}$ number of Safe planets (each containing 11 billion mediocre lives) is better than $_{\text{BIG}}$ number of Longshot planets (each with $_{\text{LOW}}$ of containing 11 billion great lives and otherwise none). From Anti-stubbornness, we know that $_{\text{BIG}}$ number of Longshot planets (each with $_{\text{LOW}}$ of containing 11 billion great lives) is better than certainty of one planet containing 10 billion great lives and the others containing none. By transitivity, it is better to have $_{\text{BIG}}$ number of Safe planets (each containing 11 billion mediocre lives) than one containing 10 billion great lives and the others containing none.

> **Anti-superiority**: There is some *n* such that it is better to have *n* lots of 11 billion mediocre lives than 10 billion great lives.

So, there is some number of mediocre lives ($_{\text{BIG}}$ × 11 billion) that would be better than 10 billion great lives. This is the Repugnant Conclusion.

3. Beyond Populations

While we have spoken exclusively about the Repugnant Conclusion and Population Superiority so far, the result here generalizes beyond population ethics. In this section, I extend the argument to well-being and normative ethics.

[50] Separability is playing a role in the proof of Anti-stubbornness too. I avoid invoking it in the main text to avoid getting into the weeds, since it merely rules out an implausible axiology according to which 11 billion great lives would be good but 22 billion great lives would be disastrous. On this view, no number *n* gambles on 11 billion great lives would be better than 10 billion great lives for certain: If *n* is small, then the low probability of getting an extra billion would not be worth the risk of getting none; if *n* is large, then the high probability of getting 11 billion would not be worth the risk of getting 22 billion. Of course, this axiology would violate Separability because the value of the second lot of great lives depends on the existence of the first.

3.1 *Well-being*

In Section 1, we saw Parfit's single-life analogue of the Repugnant Conclusion, in which a life of 100 years at a high welfare level seems better than a life of any length at a low positive welfare level. But many more results in the context of well-being have the same structure. Perhaps you follow Ross (1930: 150) in believing that "no amount of pleasure is better than any amount of virtue" or Mill (1998) in believing that any quantity of higher pleasures ("pleasures of the intellect, of the feelings and imagination, and of the moral sentiments") is better than any quantity of lower pleasures ("those of mere sensation").

The Repugnant Result we will use for our illustration is Nozick's experience machine case:

> Suppose there were an experience machine that would give you any experience you desired. Superduper neuropsychologists could stimulate your brain so that you would think and feel you were writing a great novel, or making a friend, or reading an interesting book. All the time you would be floating in a tank, with electrodes attached to your brain. Should you plug into this machine for life, preprogramming your life's experiences? (Nozick 1974: 42).

According to Nozick, no artificial life—in which we merely have the *experience* of falling in love, accomplishing our goals, cultivating fulfilling relationships, etc.—is better than a life in the real world where we *actually* do these things. Call this view—that some quantity of real experiences (say, 50 years' worth) is better than any quantity of artificial experiences—Well-being Superiority.

Perhaps Separability is less obviously true in this case: Wouldn't the value of artificial experiences diminish marginally as we get bored of them? Perhaps so, but it is difficult to see why the value of experiences would diminish for creatures who do not get bored of—or even remember—their past experiences.[51] And if Separability is at least true of *these* creatures' experiences, then a CAT Argument will imply that a life in the experience

[51] A similar point is made by Nebel (2022: 204) concerning McTaggart's (1927: 452-53) view that an excellent life that lasts for a million years is better than an oyster-like life (containing "very little excess of pleasure over pain") of any length.

machine—in which you do not retain memories or get bored—of a sufficient length is better than life in the real world.

Suppose you are looking after Sleeping Beauty, who will be unconscious until a prince awakens her, at which point she will live for 51 years in the real world before she dies. You are given the option of placing Sleeping Beauty inside the experience machine for a day. Since time is dilated in the experience machine, a single day in the machine will allow her to live a full, wonderful, artificial life. However, if the prince arrives while Sleeping Beauty is inside the machine, he will not awaken her, and she will lose the opportunity to live in the real world.

Continuity tells us that there is a sufficiently small probability *p* of the prince arriving today such that it is better for Sleeping Beauty to spend the day in the experience machine. Call this probability $_{\text{LOW}}$. Suppose you are given the same choice on Day 2: Leaving her in the bed provides an independent $_{\text{LOW}}$ probability of 51 years in the real world, and placing her in the experience machine guarantees an artificial life. By Separability, we know that the value of the artificial life inside the experience machine on Day 2 does not depend on whether Sleeping Beauty lived an artificial life on Day 1. So, irrespective of whether she spends Day 1 in bed or the machine, it is better for her to spend Day 2 in the machine. And on Day 3, it is better for her to spend the day in the machine irrespective of whether she spent Day 1 and Day 2 in the machine. Iterating this reasoning for any number *n* days gives us the result that, for each of *n* days (Day 1, …, Day *n*), it is better for Sleeping Beauty to spend that day inside the experience machine irrespective of what she does on the other days. And by Weak Agglomeration, we can conclude that it is better for Sleeping Beauty to spend all *n* days in the experience machine.

> **Anti-recklessness**: There is some (nonzero) *p* such that, for any (nonzero) *n*, it is better to have *n* days of artificial life than *n* independent gambles at *p* on 51 years in the real world and otherwise nothing.

Suppose we remove the option of placing Sleeping Beauty in the experience machine. You have the resources to keep Sleeping Beauty alive for some number *n* days. Naturally, the probability that the prince will arrive before she dies depends on the value of *n*. Now, suppose you have the option of calling the prince and demanding that he come today. Making the call would certainly result in the prince arriving today, but it would divert some power from Sleeping Beauty's life-support machine, which will cost her a year of life in the real world: Rather than living for 51 years after being woken, she would only live for 50. Whether it is better for Sleeping Beauty for you to call the prince depends on the value of *n*. If we can only keep her alive for a few days, then the probability that the prince will arrive before she dies is very low, and it is better for her that we secure her 50 years of life. But if we can keep her alive for sufficiently many days, then the probability of him arriving in her lifetime is high enough that we should not call the prince. Since it is better to live 51 years than 50 years—and for arbitrarily great values of *n*, the probability of the prince arriving in her lifetime is arbitrarily close to 1—Continuity tells us there is some *n* such that it is better for Sleeping Beauty that we do not call the prince. Call an *n* that witnesses this observation $_{\text{BIG}}$.

> **Anti-stubbornness**: For any (nonzero) *p*, there is some *n* such that it is better to have *n* independent gambles at *p* on 51 years of real life and otherwise none than 50 years of real life for certain.

From Anti-recklessness, we know that it is better for Sleeping Beauty to spend $_{\text{BIG}}$ number of days in the experience machine than $_{\text{BIG}}$ number of days in bed (each with a $_{\text{LOW}}$ probability of providing 51 years in the real world and otherwise nothing). From Anti-stubbornness, we know that having $_{\text{BIG}}$ days in bed is better than certainty of 50 years in the real world. By transitivity, it is better for Sleeping Beauty to spend $_{\text{BIG}}$ days in the experience machine than 50 days in the real world.

> **Anti-superiority**: There is some *n* such that it is better to have *n* days of artificial life than 50 years of real life.

So, there is some amount of time in the experience machine ($_{\text{BIG}}$ days) that would be better than 50 years in the real world. Well-being Superiority is false.

3.2 *Ethics*

Similarly structured results arise in the context of ethics. Most obviously, well-being Repugnant Results will straightforwardly translate into ethical Repugnant Results for any normative theory on which it is permissible, all else equal, to improve people's lives: If allowing them to live a significantly longer life at a low positive welfare level is improving their life, then it is permissible to deny them the Century of Ecstasy for the sake of securing them Drab Eternity. For instance, Mill's (1998) utilitarian theory is usually interpreted as positing a lexical priority of higher pleasures over lower pleasures: Agents should maximize higher pleasures, then look to lower pleasures in the event of a tie.

But some kinds of moral theory posit a Superiority distinct from considerations of well-being. For instance, absolutist deontological theories submit that no quantity of (at least some kinds of) consequentialist considerations can justify violating a duty. Perhaps, for instance, you think that a judge should not find an innocent person guilty, no matter how many people would be pleased with such a verdict, or that you should not torture an infant, no matter how many people would enjoy watching. The case we will use motivates a *contractualist* deontology—whereby significant harms to one person cannot be justified by trivial benefits to others:

> Suppose that Jones has suffered an accident in the transmitter room of a television station. Electrical equipment has fallen on his arm, and we cannot rescue him without turning off the transmitter for fifteen minutes. A World Cup match is in progress, watched by many people, and it will not be over for an hour. Jones's injury will not get any worse if we wait, but his hand has been mashed and he is receiving extremely painful electrical shocks. Should we rescue him now or wait until the match is over? Does the right thing to do depend on how many people are watching—whether it is one million or five million or a hundred million? (Scanlon 1998: 235).

Intuitively, even though each additional person watching the World Cup gives us an extra reason to continue the transmission, no number of viewers can justify allowing Jones to suffer. (Call this Ethical Superiority.) Again, Separability seems plausible here: We should be able to determine the value of one person's enjoying the World Cup broadcast without looking at how many people are also enjoying it.

(Note a slight shift: We are no longer talking about which state of affairs would be better, but which act is better. The shift should not matter since deontologists should be able to make sense of the notion that some acts are better than others.)

Suppose one million people are watching the World Cup broadcast. Cutting the transmission would prevent them from watching but would give you a chance of saving Jones. Saving Jones would require sending a medic to pull the equipment off him. But unfortunately, no medic has the keycard to Jones's room. So, cutting the transmission would save Jones if and only if the door to the transmitter room door's keycard system is malfunctioning. By Continuity, we know there is some *p* such that if a million people are watching the broadcast and there is a *p* probability of the keycard system malfunctioning, then we ought to continue the transmission. Call this probability $_{\text{LOW}}$. Now, suppose you learn that another employee, Smith, is trapped in a different transmitter room in a relevantly identical situation to Jones. By Separability, we know that since one million people are watching this second transmission and the probability that Smith's door is malfunctioning is $_{\text{LOW}}$, we ought to continue the transmission, irrespective of whether we continue Jones's. Iterating this reasoning for *n* transmissions and *n* corresponding employees gives us the result that, for each of *n* transmissions, it is better to continue that transmission, irrespective of whether we continue the others. And by Weak Agglomeration, it is better to continue all *n* transmissions than cut all *n* transmissions.

> **Anti-recklessness**: There is some (nonzero) probability *p* such that, for any (nonzero) *n*, it is better to allow *n* million people to enjoy the World Cup than take *n* independent gambles at *p* on saving an employee.

Now, suppose all *n* transmissions have already been cut; you must choose how to go about saving the employees. If you send one medic to each of the *n* transmitter rooms, then each employee will be saved if and only if their room's keycard system is broken. Alternatively, you could send all *n* medics to a single transmitter room to break down the door and certainly save that employee but no others. Of course, if you send one medic to each room, the probability of saving *multiple* employees increases as *n* increases; for arbitrarily great values of *n*, this probability is arbitrarily close to 1. And since it is better to save multiple employees than to merely save one,[52] we know via Continuity that there is some *n* such that you ought to send one medic to each room, rather than sending them all to one. Call an *n* that witnesses this observation $_{\text{BIG}}$.

> **Anti-stubbornness**: For any (nonzero) *p*, there is some *n* such that it would be better to have *n* independent gambles at *p* on saving an employee than to save one for certain.

From Anti-recklessness, we know that allowing $_{\text{BIG}}$ million viewers to watch the World Cup is better than taking $_{\text{BIG}}$ gambles on $_{\text{LOW}}$ of saving an employee. From Anti-stubbornness, we know that taking $_{\text{BIG}}$ gambles on $_{\text{LOW}}$ of saving an employee is better than saving one for certain. By transitivity, it is better to allow $_{\text{BIG}}$ million viewers to watch the World Cup than to certainly save one employee.

> **Anti-superiority**: There is some *n* such that allowing *n* million people to watch the World Cup is better than saving one employee.

So, there is some number of people watching the World Cup ($_{\text{BIG}}$ million) such that we should not cut the transmission for the sake of saving an employee from extremely painful electric shocks.

4. Denying CAT

Many axiologies claim to uphold both Superiority and Separability. These approaches generally treat value as a vector quantity—a quantity expressed as a list of components that cannot be reduced to a single scale.[53] Call these components *superior value* and

[52] Perhaps some patient-affecting deontological views would deny that saving two employees is better than saving one in the event that the identity of the two are unknown, but the identity of the one is known. Frick (2015: 215), for instance, endorses an *ex ante* contractualism that implies that it is better to certainly save one identified person than certainly save two unidentified people. Some simple alterations to the case will deal with views like this. Suppose that cutting each transmission raises
the probability of saving Jones *and* Smith but does not affect the probability of saving any other employee. Frick's view should accept that saving Jones and Smith is better than merely saving Jones, since both Jones and Smith are identified.

[53] Within the domain of population ethics, this kind of approach is taken by Parfit (1986, 2016); Griffin (1986); Rachels (2001); Crisp (1992); Glover (1977); Edwards (1979); Lemos (1993); Portmore

inferior value. The standard approach is to order these components lexicographically, such that some object *x* is better than another object *y* iff:

1. *x* has greater superior value than *y*, or
2. *x* and *y* have equal superior value, and *x* has greater inferior value.

On this view, no number of mediocre lives is more valuable than any number of great lives because inferior value will only be consulted as a tiebreaker for superior value.[54]

We now know that these axiologies (assuming they are internally coherent and do, in fact, uphold Separability) deny at least one CAT principle. In most cases, the details of the theory will not have been worked out sufficiently to determine which principle they reject. But a common approach is to deny Weak Agglomeration by discounting small probabilities to zero. This strategy—call it *discounting*—has become increasingly popular in recent years.[55] Discounting is useful for a variety of reasons, including allowing Superiority theories to uphold Continuity: We can use a lexicographical ordering to ensure that some quantity of *x* is better than any quantity of *y* but use discounting to ensure that a sufficiently small *probability* of *x* is worse than some quantity of *y* (since we discount this probability to zero).[56] If we adopt discounting, we cannot prove Anti-recklessness: We ignore the (sufficiently small) probability that an individual Longshot planet contains great lives, so it is better for each planet to be Safe, irrespective of the other planets. However, for a sufficiently large number of Longshot planets, we will *not* ignore the probability that *one of* them contains great lives, so it is better for them all to be Longshot than Safe. Since

---

(1999); Riley (1999, 2009, 1993, 2008); Nebel (2022) For an extension of this approach to absolutist deontological theories, see Lee-Stronach (2018).

[54] Lexicographical orderings have been criticized for being implausibly extreme: An arbitrarily small difference to superior value outweighs an arbitrarily large difference in inferior value. However, the lexicographical ordering, while the most orthodox approach, is not the only decision rule these theories can use. For instance, Nebel (2022) offers a theory with weaker *better than* conditions, which allows inferior value to come into play when there is a sufficiently small difference in superior value—rather than only when there is a tie.

[55] The most common context in which it is endorsed is as a solution to the St. Petersburg Paradox (e.g., Bernoulli 1738; d'Alembert 1761; Buffon 1777; Condorcet 1785; Borel 1962; Jordan 1994: 217-18; Buchak 2013: 73-74; Smith 2014, 2016; Robert 2018; Monton 2019).

[56] Discounting is endorsed for this reason by Kagan (1989: 89-92); Aboodi, Borer, and Enoch (2008); Hawley (2008); Haque (2012); Bjorndahl, London, and Zollman (2017); Lazar (2017); Lee-Stronach (2018); Tarsney (2018); Kosonen (2021).

it is better for each *individual* planet to be Safe irrespective of the others, and yet it is better that *every* planet is Longshot, Weak Agglomeration is false. Given the intuitiveness of Weak Agglomeration, its denial seems to be a substantial cost to those who endorse discounting. In the next section, I show that—whatever the merits and demerits of discounting—an approach that endorses discounting *and* Superiority faces uniquely challenging problems.

5. Problems

There is an important similarity between Superiority and discounting: A commitment to either involves attending to some outcomes and ignoring others. Superiority theories instruct agents to ignore outcomes that do not contain the superior good (except in the event of a tie); discounting makes agents ignore outcomes attached to sufficiently improbable states. A theory with both features is especially troublesome because we can construct a decision problem where all but one outcome is ignored for one reason or the other. Unsurprisingly, decisions made on the basis of a single (relatively improbable) outcome are liable to be extremely unreasonable.

Due to this similarity, we will use the same strategy to draw out problems with both discounting and Superiority. First, we identify the boundary separating the class of objects the theory tells us to care about and the class we are instructed to disregard. Discounters care about non-negligible probabilities and disregard negligible ones, while proponents of Superiority care about superior objects and disregard inferior ones. Then, we will posit pairs of objects at an arbitrarily small distance from each side of this boundary. The problem is that our *Superiority Discounter* will refuse to take out insurance against their gamble on the superior good for any arbitrarily small price to the *value* of this good or their *probability* of acquiring it. The results will be demonstrated for each of our three Superiority theories.

We will start with the paradigmatic case in which the Superiority Discounter prefers a non-negligible probability of the superior good to certainty of the inferior good. In Table 27, the green square in the decision matrix is the only outcome influencing the Superiority Discounter's decision. The threshold below which we should discount probabilities to zero might vary contextually, but for illustration, we will suppose it is fixed at one in 1 million.

(Let *Superior* refer to some quantity of the superior good that is better than any quantity of the inferior good and *Inferior* refer to some quantity of the inferior good.)

TABLE 27

	State 1 (1 – 1/1m)	State 2 (1/1m)
Gamble		Superior
Safe	Inferior	Inferior

5.1 *Discounting*

The first change we will make to the decision problem involves shifting some of State 1's probability into a third state, which will be just under the threshold. The outcomes in this state will be extreme, but the Superiority Discounter will ignore them due to the state's improbability. In Table 28, this change is reflected for Population Superiority, where *Great* corresponds to a great life, *Mediocre* corresponds to a mediocre life, *Heaven* corresponds to the best life imaginable, and *Hell* corresponds to the worst life imaginable. We might suppose that if the planet is Longshot, it has a small probability of containing 10 billion great lives, an approximately equal probability of containing 10 billion hellish lives, and otherwise none. But if it is Safe, it is certain to contain 10 billion lives that are likely mediocre but possibly heavenly.

TABLE 28

	State 1 (1 – 1/1m)	State 2 (1/1m)	State 3 (1/1m – ε)
Longshot		10b Great	**10b Hell**
Safe	10b Mediocre	10b Mediocre	**10b Heaven**

Both prospects now have an approximately equal probability of producing great lives, except Safe's great lives would be considerably better than Longshot's. Longshot is further dampened by taking an approximately equal risk of hellish lives, while Safe is further sweetened by guaranteeing *at least* mediocre lives. But since the Superiority Discounter ignores State 3, they still prefer Longshot to Safe.

Table 29 makes an analogous change to our Well-being Superiority theory. Suppose you must choose whether to place Sleeping Beauty inside the experience machine today. If the prince arrives, there is an approximately equal probability that he will enter the tower on the North or South side. If he enters the North side, he will find Sleeping Beauty's bed before he finds the experience machine; if she is in the bed, he will wake her; if she is not, he will leave. If he enters the South side, he will find the experience machine first; if she is inside it, he will wake her, and they will sell the machine and use the money to live for 100 years at an extremely high welfare level (call this outcome *Century of Ecstasy*); if she is not inside it, he will fiddle with the machine, causing it to explode and killing them both painfully. Table 29 formalizes your decision whether to place her in the experience machine.

TABLE 29

	No Prince (1 – 1/1m – (1/1m – ε))	North (1/1m)	South (1/1m – ε)
Bed		50 Years	**Painful Death**
Machine	Artificial Life	Artificial Life	**Century of Ecstasy**

Both prospects offer an approximately equal probability of living in the real world, with Machine's real-world life being longer, at a higher welfare level, coming without the risk of a painful death, and with the insurance of at least an artificial life. Nevertheless, the Superiority Discounter ignores State 3 and so prefers Bed.

Finally, consider the transmitter room case. Suppose that Jones could be saved whether or not the transmission is cut; cutting it would merely raise the probability of saving him by an arbitrarily small amount. However, cutting the transmission would also introduce an approximately equal probability of killing Jones.

TABLE 30

	State 1 (1 – 1/1m – (1/1m – ε))	State 2 (1/1m)	State 3 (1/1m – ε)
Cut		Save	**Kill**
Continue	Broadcast	Broadcast	**Broadcast + Save**

Cutting the transmission is unlikely to change Jones's situation, but if it does, it is (approximately) equally likely to save or kill him. If you continue the transmission, you will have about the same chance of saving him without any risk of killing him, alongside the bonus of allowing many people to watch the broadcast. Still, the Superiority Discounter believes it is impermissible to leave the broadcast running.

Adopting discounting allows Superiority advocates to avoid chasing arbitrarily small probabilities at any cost—thus upholding Continuity—without accepting the Repugnant Result. But introducing a discounting threshold makes them chase arbitrarily small *increases* in probability at any cost (when the probabilities are on either side of this threshold). It is not clear that this result is any less undesirable: Since Inferior is better than a sufficiently small probability of Superior, it is strange that *no quantity* of Inferior can justify an arbitrarily small *decrease* in the probability of Superior.

5.2 *Population Superiority*

Having seen how to derive problems for discounting, a similar strategy will produce problems for Superiority. Again, we will begin with the paradigm case in which the Superiority Discounter prefers a non-negligible probability of Superior to any quantity of Inferior, as in Table 31. Then, we will identify the border separating the outcomes we must attend to from those we should ignore. However, instead of adding additional states to the decision problem, we will alter the existing outcomes in ways that the Superiority advocate will not be sensitive to.

TABLE 31

	State 1 (1 – 1/1m)	State 2 (1/1m)
Gamble		10b Great
Safe	10b Mediocre	10b Mediocre

As we know, our Population Superiority theory maximizes expected great lives, looking to mediocre lives to break ties. Naturally, there is some threshold separating great lives (the superior good) from mere mediocre lives (the inferior good). Wherever this threshold is, call a life that is an arbitrarily small distance below the threshold a *great*- life.

TABLE 32

	State 1 (1 – 1/1m)	State 2 (1/1m)
Longshot		10b Great
Safe	**10b Great-**	**10b Great-**

Now Longshot offers a one-in-1-million gamble on great lives, while Safe *guarantees* a life at an approximately equal welfare level. However, the Superiority Discounter will refuse to take the insurance at an arbitrarily small price to the value of the life.

We can run a similar strategy to dampen Longshot. We have described Population Superiority as a commitment to a preference among valuable objects rather than disvaluable objects. Still, the theory must count disvaluable lives towards either the *superior* value or the *inferior* value of the prospect. The most charitable interpretation assigns superiority to avoiding sufficiently bad (call them *horrible*) lives.[57] But just as the Superiority Discounter is insensitive to positive-but-not-great lives, they will be insensitive to negative-but-not-horrible lives. Call a life that is an arbitrarily small distance above the horrible threshold a horrible+ life.

TABLE 33

	State 1 (1 – 1/1m)	State 2 (1/1m)
Longshot	**10b Horrible+**	10b Great
Safe	10b Great-	10b Great-

[57] If they assigned superiority to avoiding *all* disvaluable lives, then they would be unwilling to take a small (but non-negligible) risk of a low negative life for the sake of near-certainty of a great- life. If they did not assign superiority to avoiding *any* disvaluable lives, they would be willing to risk near-certain hell for the sake of a small probability of a great life. Since both of these results are implausible, the most charitable interpretation of the position assigns superiority to avoiding *sufficiently bad* lives.

Even though Longshot is near-certain to produce horrible+ lives, the Superiority Discounter will chase its small probability of producing great lives, even though these great lives are an arbitrarily small amount better than the lives *guaranteed* by Safe.

5.3 *Well-being Superiority*

As we saw, Population Superiority posits superiority between extrema on a spectrum: Great lives are on the upper end of a welfare spectrum, and mediocre lives are on the lower end. As a result, finding borderline cases simply requires finding lives barely below the greatness threshold. However, things are not so straightforward for Well-being Superiority: The difference between real and artificial lives seems to be one of kind, not degree. So, we will have to take a different approach to find borderline cases.

Our Well-being Superiority theory claims that *some quantity* of time in the real world is better than any quantity inside the experience machine. This kind of view is called *Weak Superiority*; Strong Superiority claims that *any* quantity of the superior good is better than any quantity of the inferior good. Using the weak version seems correct in this case: We would not want to say that one millisecond in the real world is better than any amount of time in the experience machine. However, a second strategy for teasing out problems with Superiority involves introducing outcomes that do not quite have a sufficient quantity of the superior good. In Table 34, we have increased Machine's insurance payout to 49 years in the real world.

TABLE 34

	State 1 (1 – 1/1m)	State 2 (1/1m)
Bed		50 Years
Machine	**49 Years**	**49 Years**

Since Well-being Superiority judges that *at least* 50 years in the real world is better than any amount of time in the experience machine, our discounting Superiority theory will judge outcomes in which she lives 49 years as providing only inferior (tie-breaking) value.

Again, we have not filled in the details of our Superiority theory's attitude toward disvaluable outcomes. However, a charitable interpretation would not make Well-being Superiority insensitive to arbitrarily bad outcomes. Just as some quantity of time in the real world is better than any quantity of time in the experience machine, *avoiding* some quantity of time in the *torture machine* is presumably better than any quantity of time in the experience machine: So, our Superiority Discounter will endure one second in the torture machine for the sake of some (sufficiently valuable) artificial life, but no artificial life would be worth enduring a millennium in the torture machine. At some point between one second and one millennium, a threshold will separate the superior good from the inferior good. Call some quantity of time inside the torture machine slightly less than the threshold *torture+*.

TABLE 35

	State 1 (1 – 1/1m)	State 2 (1/1m)
Bed	**Torture+**	50 Years
Machine	49 Years	49 Years

Now, the Superiority Discounter prefers near-certain torture+ for the sake of 50 years of life instead of near-certainty of 49 years in the real world.

5.4 *Ethical Superiority*

The strategies used to extract problems with Population Superiority and Well-being Superiority will not work for Ethical Superiority: Duty violations (like allowing Jones to suffer) and consequentialist goods (like World Cup enjoyment) are not extrema on a spectrum (like great lives and mediocre lives); nor is this theory a Weak Superiority theory (whereby only a *sufficient quantity* of the superior good is superior to the inferior good, like Well-being Superiority), since small duty violations of any size are impermissible.

The approach to finding borderline cases for Ethical Superiority involves noticing vagueness. We will pick out a feature of both the superior and inferior objects, where the degree to which that feature is present determines which class it falls into. For instance, the deontologist believes we have a duty to save Jones from fifteen minutes of suffering. If cutting the transmission would stop his suffering after one second, then doing so would save him—even though it would not prevent the full fifteen minutes of suffering. But if cutting the transmission would only truncate his suffering by a fraction of a second, then you would be unable to save Jones and so not duty-bound to cut the transmission. (Presumably, the deontologist would not want to say that we have a duty to prevent an arbitrarily small amount of suffering at any cost.) Now we can treat Ethical Superiority on a spectrum: For some amount of time *t*, shortening Jones's suffering by *t* is to fail to save him, but shortening it by *t* + ε is to save him. Let *save-* refer to shortening Jones's suffering by *t*.

TABLE 36

	State 1 (1 – 1/1m)	State 2 (1/1m)
Cut		Save
Continue	Broadcast + **Save-**	Broadcast + **Save-**

Again, we can dampen Cut by filling in some details about Ethical Superiority's attitude towards disvalue. A charitable interpretation would allow the deontologist to posit a duty to avoid killing: After all, we would not want our deontologist to kill to save Jones. Suppose saving Jones requires lifting some heavy equipment off him. If we drop it on him and he succumbs to his injuries a minute later, we would have killed him; if we drop it on him and his injuries eventually hasten his death by a few milliseconds, we would not have killed him. Let a *killing-* refer to shortening his life by an arbitrarily small amount less than the amount that would constitute a killing.

TABLE 37

	State 1 (1 – 1/1m)	State 2 (1/1m)
Cut	**Kill-**	Save
Continue	Broadcast + Save-	Broadcast + Save-

According to Ethical Superiority, we have a duty to near-certainly kill- Jones for a small chance of saving him, rather than certainly at least saving- him.

5.5 *Preposterous Propositions*

As seen in 5.1, wherever we set the discounting threshold, it will be possible to place pairs of states arbitrarily small distances from either side. The discounter will prefer a small chance of Superior and the same chance of a horrible outcome rather than an approximately equal chance of Superior+ with the insurance of Inferior. Similarly, as we have seen in 5.2–5.4, whatever the goods assigned a superiority relation, there will be objects at arbitrarily small distances from either side of the boundary. Even though these objects are almost identical, the Superiority theorist will prefer a small probability of one to certainty of the other. Both results are problematic, but when they arise in a single decision problem, the result is *preposterous*.

In Table 38, both problems for Population Superiority are combined into a single decision problem.

TABLE 38

	State 1 (1 – 1/1m – (1/1m – ε))	State 2 (1/1m)	State 3 (1/1m – ε)
Longshot	10b Horrible+	10b Great	10b Hell
Safe	10b Great-	10b Great-	10b Heaven

Of course, since the perfectionist discounter is only sensitive to A's outcome in State 2, we can increase the number of people in each other outcome without affecting the perfectionist discounter's preference.

TABLE 39

	State 1 (1 – 1/1m – (1/1m – ε))	State 2 (1/1m)	State 3 (1/1m – ε)
Longshot	*n* Horrible+	10b Great	*n* Hell
Safe	*n* Great-	*n* Great-	*n* Heaven

Longshot will almost certainly result in an arbitrarily great number of people living borderline horrible lives, and otherwise an approximately equal probability of either 10 billion people living a great life or an arbitrarily great number of people living hellish lives. Safe guarantees an arbitrarily great number of people at least borderline great lives and otherwise heavenly lives. The fact that, according to Population Superiority, Longshot is better than Safe, is the Preposterous Conclusion.

Both problems for Well-being Superiority are combined in Table 40.

TABLE 40

	No Prince (1 – 1/1m)	North (1/1m)	South (1/1m – ε)
Bed	Torture+	50 Years	Painful Death
Machine	49 Years	49 Years	Century of Ecstasy

Bed would almost certainly result in torture+ and otherwise has an approximately equal probability of giving Sleeping Beauty 50 years of life and causing her painful death. Machine would guarantee at least 49 years of life, with a small chance of the Century of Ecstasy. Well-being Superiority endorses a Preposterous Proposition—that Bed is better than Machine.

Finally, the problems are combined for Ethical Superiority in Table 41.

TABLE 41

	State 1 (1 – 1/1m – (1/1m – ε))	State 2 (1/1m)	State 3 (1/1m – ε)
Cut	Kill-	Save	Kill
Continue	Broadcast + Save-	Broadcast + Save-	Save

Continuing the broadcast, in addition to satisfying many viewers, is certain to at least save-Jones and otherwise save him. Instead, the deontological discounter prefers to cut the transmission, near-certainly killing- Jones for the sake of a one-in-a-million chance of extending his life by one second and otherwise killing him. And again, we can magnify the problem by increasing the number of people in the equation.

TABLE 42

	State 1 (1 – 1/1m – (1/1m – ε))	State 2 (1/1m)	State 3 (1/1m – ε)
Cut	Kill- × *n*	Save	Kill × *n*
Continue	(Broadcast + Save-) × *n*	(Broadcast + Save-) × *n*	Save × *n*

Continuing the transmission, in addition to satisfying *n* million viewers, is certain to at least save- *n* lives and otherwise save *n* lives. Instead, the deontological discounter prefers to cut the transmission, near-certainly killing- *n* people, for the sake of a one-in-a-million shot at extending one life by a second longer than Continue certainly would, and an approximately equal probability of killing *n* people. Ethical Superiority endorses the Preposterous Proposition that Cut is better than Continue.

6. Conclusion

This paper shows that Superiority and Separability are incompatible, given Continuity, Weak Agglomeration, and Transitivity. Given the plausibility of Continuity (considering how implausibly extreme Recklessness and Stubbornness are) and transitivity, our best bet of upholding both Superiority and Separability comes from denying Weak Agglomeration. We can deny Weak Agglomeration in a way that prevents the argument from going through by discounting small probabilities to zero. However, the resultant view avoids Repugnant Results only at the cost of Preposterous Propositions, which are much worse.

References

Aboodi, R., Borer, A., and Enoch, D. 2008. 'Deontology, Individualism, and Uncertainty: A Reply to Jackson and Smith.' *The Journal of Philosophy,* 105: 259-72.

Alexander, L. 2018. 'The Need to Attend to Probabilities—for Purposes of Self-Defense and Other Preemptive Actions.' *San Diego Law Review,* 55: 223-29.

Arrhenius, G. and Rabinowicz, W. 2015. Value Superiority. In *The Oxford Handbook of Value Theory*, ed. I. Hirose and J. Olson, 225-48: Oxford University Press USA.

Arrow, K. J. 1971. *Essays in the Theory of Risk-Bearing*. Amsterdam: North-Holland Pub. Co.

Ashford, E. 2003. 'The Demandingness of Scanlon's Contractualism.' *Ethics,* 113: 273-302.

Bassett, G. W., Jr. 1987. 'The St. Petersburg Paradox and Bounded Utility.' *History of Political Economy,* 19: 517-23.

Bernoulli, D. 1738. 'Exposition of a New Theory on the Measurement of Risk.' *Econometrica,* 22: 23-36.

Bjorndahl, A., London, A. J., and Zollman, K. J. S. 2017. 'Kantian Decision Making under Uncertainty: Dignity, Price, and Consistency.' *Philosophers' Imprint,* 17.

Borel, E. 1962. *Probabilities and Life*. New York: Dover.

Bostrom, N. 2009. 'Pascal's Mugging.' *Analysis,* 69: 443-45.

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

Buchanan, A. 1979. 'Revolutionary Motivation and Rationality.' *Philosophy & Public Affairs,* 9: 59-82.

Buffon, G.-L. L. d. 1777. Essai D'arithmétique Morale. In *Supplement a l'histoire Naturelle*. Paris: L'Imprimerie.

Chalmers, A. 2017. An Offer You Can't (Rationally) Refuse: Systematically Exploiting Utility-Maximisers with Malicious Gambles, University of Sydney.

Clarke, R. 2013. 'Belief Is Credence One (in Context).' *Philosophers' Imprint,* 13: 1-18.

Colyvan, M., Cox, D., and Steele, K. S. 2010. 'Modelling the Moral Dimension of Decisions.' *Noûs,* 44: 503-29.

Condorcet, J.-A.-N. d. C. 1785. *Essai Sur L'application De L'analyse à La Probabilité Des DéCisions Rendues à La Pluralité Des Voix*. Paris: Imprimerie royale.

Crisp, R. 1992. 'Utilitarianism and the Life of Virtue.' *The Philosophical Quarterly,* 42: 139-60.

d'Alembert, J. L. R. 1761. *Opuscules MathéMatiques, Ou MéMoires Sur DifféRens Sujets De GéOméTrie, De MéChanique, D'optique, D'astronomie &C*. Paris: David.

Edwards, R. B. 1979. *Pleasures and Pains: A Theory of Qualitative Hedonism*. Ithica, NY: Cornell University Press.

Frick, J. 2015. 'Contractualism and Social Risk.' *Philosophy & Public Affairs,* 43: 175-223.

Fried, B. H. 2012. 'What Does Matter? The Case for Killing the Trolley Problem (or Letting It Die).' *The Philosophical Quarterly,* 62: 505-29.

Glover, J. 1977. *Causing Death and Saving Lives*. Harmondsworth; New York: Penguin.

Greco, D. 2015. 'How I Learned to Stop Worrying and Love Probability 1.' *Philosophical Perspectives,* 29: 179-201.

Griffin, J. 1986. *Well-Being: Its Meaning, Measurement, and Moral Importance*. Vol. 100: Clarendon Press.

Hájek, A. 2014. 'Unexpected Expectations.' *Mind,* 123: 533-67.

Hansson, S. O. 2013. *The Ethics of Risk: Ethical Analysis in an Uncertain World*: Palgrave Macmillan.

Haque, A. A. 2012. 'Killing in the Fog of War.' *Southern California Law Review,* 86: 63-116.

Hare, C. 2016. 'Should We Wish Well to All?' *Philosophical Review,* 125: 451-72.

Hawley, P. 2008. 'Moral Absolutism Defended.' *Journal of Philosophy,* 105: 273-75.

Hayenhjelm, M. and Wolff, J. 2012. 'The Moral Problem of Risk Impositions: A Survey of the Literature.' *European Journal of Philosophy,* 20: e26-e51.

Holm, S. 2016. 'A Right against Risk-Imposition and the Problem of Paralysis.' *Ethical Theory and Moral Practice,* 19: 917-30.

Huemer, M. 2010. 'Lexical Priority and the Problem of Risk.' *Pacific Philosophical Quarterly,* 91: 332-51.

Isaacs, Y. 2014. 'Duty and Knowledge.' *Philosophical Perspectives,* 28: 95-110.

Jackson, F. 1985. 'On the Semantics and Logic of Obligation.' *Mind,* 94: 177-95.

Jackson, F. and Pargetter, R. 1986. 'Oughts, Options, and Actualism.' *The Philosophical Review,* 95: 233-55.

Jackson, F. and Smith, M. 2006. 'Absolutist Moral Theories and Uncertainty.' *Journal of Philosophy,* 103: 267-83.

Jackson, F. and Smith, M. 2016. The Implementation Problem for Deontology. In *Weighing Reasons*. New York: Oxford University Press.

Jeffrey, R. C. 1965. *The Logic of Decision*. Vol. 77: University of Chicago Press.

Jordan, J. 1994. 'The St. Petersburg Paradox and Pascal's Wager.' *Philosophia,* 23: 207-22.

Kagan, S. 1989. *The Limits of Morality*: Oxford University Press.

Kosonen, P. 2021. 'Discounting Small Probabilities Solves the Intrapersonal Addition Paradox.' *Ethics,* 132: 204-17.

Lazar, S. 2017. 'Deontological Decision Theory and Agent-Centered Options.' *Ethics,* 127: 579-609.

Lazar, S. 2017. 'Risky Killing.' *Journal of Moral Philosophy,* 16: 1-26.

Lazar, S. and Lee-Stronach, C. 2019. 'Axiological Absolutism and Risk.' *Noûs,* 53: 97-113.

Lee-Stronach, C. 2018. 'Moral Priorities under Risk.' *Canadian Journal of Philosophy,* 48: 793-811.

Lee-Stronach, C. 2021. 'Morality, Uncertainty.' *Philosophical Quarterly,* 71: 334-58.

Lemos, N. M. 1993. 'Higher Goods and the Myth of Tithonus.' *Journal of Philosophy,* 90: 482-96.

McClennen, E. F. 1994. Pascal's Wager and Finite Decision Theory. In *Gambling on God: Essays on Pascal's Wager*, ed. J. Jordan, 115-38. Boston: Rowman & Littlefield.

McKerlie, D. 1986. 'Rights and Risk.' *Canadian Journal of Philosophy,* 16: 239-51.

McTaggart, J. M. E. 1927. *The Nature of Existence Volume Ii*: Cambridge University Press.

Mill, J. S. 1998. *Utilitarianism*. Oxford: Oxford University Press. Original edition, 1863.

Miller, F. and Sartorius, R. 1979. 'Population Policy and Public Goods.' *Philosophy and Public Affairs,* 8: 148-74.

Monton, B. 2019. 'How to Avoid Maximizing Expected Utility.' *Philosophers' Imprint,* 19: 1-25.

Nebel, J. M. 2019. 'An Intrapersonal Addition Paradox.' *Ethics,* 129: 309-43.

Nebel, J. M. 2022. Totalism without Repugnance. In *Ethics and Existence: The Legacy of Derek Parfit*, ed. J. McMahan, T. Campbell, J. Goodrich and K. Ramakrishnan, 200-31: Oxford University Press.

Nozick, R. 1974. *Anarchy, State, and Utopia*. Vol. 7: Basic Books.

Parfit, D. 1986. Overpopulation and the Quality of Life. In *Applied Ethics*, ed. P. Singer, 145-64: Oxford University Press.

Parfit, D. 2016. 'Can We Avoid the Repugnant Conclusion?' *Theoria,* 82: 110-27.

Persson, I. 2004. The Root of the Repugnant Conclusion and Its Rebuttal. In *The Repugnant Conclusion: Essays on Population Ethics*, ed. T. Tännsjö and J. Ryberg, 187-99. Dordrecht: Springer Netherlands.

Peterson, M. 2002. 'What Is a De Minimis Risk?' *Risk Management,* 4: 47-55.

Peterson, M. 2017. *An Introduction to Decision Theory*. 2 ed, *Cambridge Introductions to Philosophy*. Cambridge: Cambridge University Press.

Portmore, D. W. 1999. 'Does the Total Principle Have Any Repugnant Implications?' *Ratio,* 12: 80-98.

Rachels, S. 2001. 'A Set of Solutions to Parfit's Problems.' *Noûs,* 35: 214-38.

Rachels, S. 2004. Repugnance or Intransitivity: A Repugnant but Forced Choice. In *The Repugnant Conclusion: Essays on Population Ethics*, ed. T. Tännsjö and J. Ryberg, 163-86. Dordrecht: Springer Netherlands.

Rawls, J. 1971. *A Theory of Justice*. Cambridge, MA: Harvard University Press.

Riley, J. 1993. 'On Quantities and Qualities of Pleasure.' *Utilitas: A Journal of Utilitarian Studies,* 5: 291-300.

Riley, J. 1999. 'Is Qualitative Hedonism Incoherent?' *Utilitas: A Journal of Utilitarian Studies,* 11: 347-58.

Riley, J. 2008. 'Millian Qualitative Superiorities and Utilitarianism, Part I.' *Utilitas: A Journal of Utilitarian Studies,* 20: 257-78.

Riley, J. 2009. 'Millian Qualitative Superiorities and Utilitarianism, Part Ii.' *Utilitas: A Journal of Utilitarian Studies,* 21: 127-43.

Robert, D. 2018. 'Expected Comparative Utility Theory: A New Theory of Rational Choice.' *Philosophical Forum,* 49: 19-37.

Ross, W. D. 1930. *The Right and the Good. Some Problems in Ethics*. Vol. 41: Clarendon Press.

Samuelson, P. A. 1977. 'St. Petersburg Paradoxes: Defanged, Dissected, and Historically Described.' *Journal of Economic Literature,* 15: 24-55.

Savage, L. J. 1954. *The Foundations of Statistics*. Vol. 11: Wiley Publications in Statistics.

Scanlon, T. 1998. *What We Owe to Each Other*. Vol. 66: Belknap Press of Harvard University Press.

Schwitzgebel, E. 2017. '1% Skepticism.' *Noûs,* 51: 271-90.

Smith, N. J. J. 2014. 'Is Evaluative Compositionality a Requirement of Rationality?' *Mind,* 123: 457-502.

Smith, N. J. J. 2016. 'Infinite Decisions and Rationally Negligible Probabilities.' *Mind,* 125: 1199-212.

Sobel, D. 2012. 'Backing Away from Libertarian Self-Ownership.' *Ethics,* 123: 32-60.

Tarsney, C. 2018. 'Moral Uncertainty for Deontologists.' *Ethical Theory and Moral Practice,* 21: 505-20.

Temkin, L. 2011. *Rethinking the Good: Moral Ideals and the Nature of Practical Reasoning*: Oxford University Press USA.

Temkin, L. S. 1987. 'Intransitivity and the Mere Addition Paradox.' *Philosophy & Public Affairs,* 16: 138-87.

Tenenbaum, S. 2017. 'Action, Deontology, and Risk: Against the Multiplicative Model.' *Ethics,* 127: 674-707.

Thomas, T. 2022. Separability and Population Ethics. In *The Oxford Handbook of Population Ethics*, ed. G. Arrhenius, K. Bykvist, T. Campbell and E. Finneron-Burns: Oxford University Press.

Von Neumann, J. and Morgenstern, O. 1944. *Theory of Games and Economic Behavior*: Princeton University Press.

Williamson, T. 2000. *Knowledge and Its Limits*. Oxford: Oxford University Press.