# Probability Discounting and Money Pumps

Petra Kosonen

# Tiny probabilities of huge payoffs

- On standard decision theory, a rational agent always maximizes expected utility.
- However, this seems to lead to counterintuitive choices in cases that involve very small probabilities of huge payoffs. Consider, for example, the following case:<sup>1</sup>

#### Pascal's Hell

Satan offers Pascal a deal: He will create a million Graham's number of happy Earth-like planets if a coin lands on heads. But if the coin lands on tails, then everyone on Earth will suffer excruciating pain until life on Earth is no longer possible. The probability of heads is one-in-Graham's-number.

 $<sup>^1 {\</sup>rm Kosonen}$  (2022, pp. 2-4). This case is based on Bostrom's (2009) Pascal's Mugging.

### Fanaticism

Should Pascal accept the offer?

- The probability of the positive payoff is tiny, so accepting the offer will almost certainly result in a negative outcome. However, as the possible payoff is enormous, Pascal is forced to conclude that the expected value of accepting the offer is positive.
- More generally, maximizing expected utility (with unbounded utilities) leads to

#### **Probability Fanaticism**

For any probability p > 0 and any (finitely) good outcome o, there is some great enough outcome O such that probability p of O (and otherwise nothing) is better than certainty of o.<sup>2</sup>

<sup>&</sup>lt;sup>2</sup>Wilkinson (2022, p.449) and Beckstead and Thomas (2020, p. 2).

# Probability Discounting

- In response to cases like this, some have argued that we ought to discount very small probabilities down to zero—let's call this Probability Discounting.
- For example, Monton (2019) argues that one ought to discount very small probabilities down to zero, while Smith (2014) argues that it is rationally permissible, but not required, to do so.
- There are many ways of making Probability Discounting precise.
- Let  $X \succeq Y$  mean that X is at least as preferred as Y.
- Also, let EU(X)<sub>pd</sub> denote the expected utility of prospect X when small probabilities have been discounted down to zero (read as 'the probability-discounted expected utility of X').
- Also, let a negligible probability be a probability below the discounting threshold, that is, a probability that should be discounted down to zero.

Then, one of the simplest versions of Probability Discounting—let's call it Naive Discounting—states:

#### Naive Discounting

For all prospects X and Y,  $X \succeq Y$  if and only if  $EU(X)_{pd} \ge EU(Y)_{pd}$ , where  $EU(X)_{pd}$  and  $EU(Y)_{pd}$  are obtained by conditionalizing on the supposition that some outcome of non-negligible probability occurs.

## EU axioms

- Given that Probability Discounting differs from Expected Utility Theory, it has to violate at least one of the following axioms that together entail Expected Utility Theory: Completeness, Transitivity, Independence and Continuity.<sup>3</sup>
- Violating these axioms renders probability discounters vulnerable to exploitation as there are money-pump arguments for each of these axioms.<sup>4</sup>
- In this talk, I'll show that Probability Discounting violates Independence and Continuity.
- It is therefore vulnerable to exploitation in the money pumps for Independence and Continuity.

<sup>4</sup>Gustafsson (2022).

 $<sup>^3</sup>$ Von Neumann and Morgenstern (1947), Jensen (1967, pp. 172–182) and Hammond (1998, pp. 152–164).

- Next, I'll discuss two ways in which Probability Discounting violates Continuity.
- First, I'll show that views that discount probabilities below some discounting threshold violate Continuity.
- Then, I'll show that views that discount probabilities up to some threshold violate another version of Continuity.

# Continuity

- Let X ≻ Y mean that X is strictly preferred (or simply 'preferred') to Y.
- ► Also, let XpY be a risky prospect with a p chance of prospect X obtaining and a 1 − p chance of prospect Y obtaining.
- Continuity then states the following:

#### Continuity

If  $X \succ Y \succ Z$ , then there are probabilities p and  $q \in (0, 1)$  such that  $XpZ \succ Y \succ XqZ$ .

- ▶ For example, suppose a coin is flipped, and an agent gets X with heads and Z with tails. Suppose further that the bias of the coin can be changed.
- Continuity requires that, with some bias, the agent prefers the coin flip to certainly getting Y, but with some other bias, the agent prefers certainly getting Y to flipping the coin.

- Views that discount probabilities below some threshold violate Continuity.
- To see how the Continuity violation happens, consider the following prospects:

### Continuity Violation

*Prospect* A Gives probability t of some very good outcome (and otherwise nothing).

*Prospect B* Certainly gives a good outcome.

*Prospect C* Certainly gives nothing.

### Continuity Violation

*Prospect* A Gives probability t of some very good outcome (and otherwise nothing).

*Prospect B* Certainly gives a good outcome.

*Prospect C* Certainly gives nothing.

- Let *t* be the discounting threshold.
- Then, all probabilities less than t will be discounted down to zero, but probabilities at least as great as t will not be discounted.
- Also, suppose that A is better than B, which is better than C; a non-negligible probability of a very good outcome (and otherwise nothing) is better than a certain good outcome, which is better than certainly getting nothing.

Next, consider the following mixed lottery:

### Continuity Violation

*Prospect ApC* Gives probability p of A and probability 1 - p of C (i.e., probability  $t \cdot p$  of a very good outcome and otherwise nothing).

CONTINUITY VIOLATION			
	$p \cdot t$	$1-p\cdot t$	
ApC	Very good	Nothing	
B	Good	Good	

### Commence Manager

- Given that t is the discounting threshold, t multiplied by any probability p < 1 must be below the discounting threshold.</p>
- Consequently, t · p is discounted down to zero, and ApC only gives a negligible probability of a positive outcome.
- And, given that *B* certainly gives a good outcome, *B* must be better than ApC for all probabilities  $p \in (0, 1)$ .
- So, now we have that A is better than B, which is better than C, but B is better than ApC for all probabilities  $p \in (0, 1)$ —which is a violation of Continuity.

Continuity Violation		
	$p \cdot t$	$1-p\cdot t$
ApC	Very good	Nothing
B	Good	Good

- There is also a money-pump argument for Continuity.
- A money-pump argument intends to show that agents who violate some alleged requirement of rationality would make a combination of choices that lead to a sure loss.
- In so far as vulnerability to this kind of exploitation is a sign of irrationality, Probability Discounting is untenable as a theory of instrumental rationality.
- The money-pump argument for Continuity goes as follows:<sup>5</sup>

<sup>5</sup>Gustafsson (2022, p. 66). Gustafsson calls this the Lexi-Pessimist Money Pump.

# Continuity Money Pump



 $A \succ A^- \succ AqC$  for all probabilities  $q \in (0, 1)$ .

- In this decision tree, the square represents a choice node and the circle represents a chance node. Going up at a choice node means accepting a trade and going down means refusing a trade.<sup>6</sup>
- The agent starts with AqC. AqC is arbitrarily similar to A; it results in the same outcome as A with a probability arbitrarily close to one.
- However, no matter how close q is to one, AqC will only give a negligible probability of a positive outcome.

<sup>6</sup>Rabinowicz (2008, p. 152).

# Continuity Money Pump



 $A \succ A^- \succ AqC$  for all probabilities  $q \in (0, 1)$ .

- Next, the agent is offered A<sup>-</sup> in exchange for AqC. A<sup>-</sup> is like A except that the agent has some amount ε less money.
- A<sup>-</sup> gives the threshold probability of a positive outcome, while AqC only gives a negligible probability of a positive outcome.
- Thus, the agent prefers  $A^-$  over AqC and accepts the trade.
- However, this means that the exploiter gets a fixed payment with only an arbitrarily small chance of having to give up anything. The situation is therefore arbitrarily close to pure exploitation.

## Discounting threshold

- The previous Continuity violation happens because the discounting threshold multiplied by any probability below one results in a probability below the discounting threshold.
- This happens because the discounting threshold is the lowest probability not discounted down to zero. Hence, the set of non-discounted values is closed (i.e., it is an interval of the form [t, 1]).
- One may think that Probability Discounting obviously violates Continuity, but that is because one is thinking of the threshold as the lowest probability not discounted.
- However, instead of the threshold being the lowest probability not discounted down to zero, it might be the highest probability that is discounted.
- If so, it is not so obvious that Probability Discounting violates Continuity.

### Discounting threshold

- In that case, there is no lowest non-negligible probability, and the set of non-discounted values is open on one side (i.e., it is an interval of the form (t, 1]).
- Consequently, A will only have positive probability-discounted expected utility if it gives at least a t + ε probability of a positive outcome, where ε is positive but arbitrarily close to zero.
- ▶ But in that case, one can always find some probability p (that may be very close to one), such that p(t + ε) > t.
- In other words, for all probabilities above the discounting threshold, there is some probability p such that their product is still above the discounting threshold.
- Consequently, Probability Discounting can avoid the previous violation of Continuity by letting the discounting threshold be the highest probability discounted down to zero.

# Mixture Continuity

However, this view violates another version of Continuity:

#### Mixture Continuity

For all prospects *X*, *Y* and *Z*, the set of probabilities {  $p \in [0,1]$  } with property  $XpZ \succeq Y$  and the set of probabilities {  $q \in [0,1]$  } with property  $Y \succeq XqZ$  are closed.<sup>7</sup>

- In effect, this principle states that if prospect XpZ is at least as good as prospect Y with some probability p, then there must be some highest and some lowest probability with which XpZ is at least as good as Y.
- (Similarly, if prospect Y is at least as good as prospect XqZ, then there must be some highest and some lowest probability with which Y is at least as good as XqZ).

<sup>&</sup>lt;sup>7</sup>This is axiom 2 in Herstein and Milnor (1953, p. 293).

# Mixture Continuity Violation

To see how this view violates Mixture Continuity, consider the following prospects:

#### Mixture Continuity Violation

- *Prospect A* Certainly gives a very good outcome.
- *Prospect B* Certainly gives a good outcome.
- *Prospect C* Certainly gives nothing.
  - Again, A is better than B, which is better than C.
  - Moreover, suppose that the very good outcome is sufficiently great so that ApC is at least as good as B for all p > t.
  - Given that t is discounted down to zero, it is not the case that AtC is at least as good as B. So, there is no lowest probability p with which ApC is at least as good as B.
  - For all p > t, ApC is at least as great as B; when p = t, ApC is worse than B.
  - This is a violation of Mixture Continuity.

## Mixture Continuity Violation $\rightarrow$ Money Pumped

- Furthermore, even though this view avoids the first Continuity violation, it is still vulnerable to the Continuity Money Pump.
- Let A<sub>t+ε</sub> be a prospect that gives probability t + ε of a very good outcome (and otherwise it gives nothing).
- A<sub>t+ε</sub> has positive probability-discounted expected utility for all ε > 0, no matter how close ε is to zero.
- Also, let A<sub>t+ε</sub>pC be a prospect that gives probability p(t + ε) of a very good outcome (and otherwise it gives nothing).

### Mixture Continuity Violation $\rightarrow$ Money Pumped

- If ε is very close to zero, A<sub>t+ε</sub>pC will only have positive probability-discounted expected utility if p is very close to one—otherwise the probability of a positive outcome would be at most t, and thus, discounted down to zero.
- As ε can be arbitrarily close to zero, A<sub>t+ε</sub>pC does not have positive probability-discounted expected utility with probabilities arbitrarily close to one; as long as p(t + ε) is at most t, A<sub>t+ε</sub>pC is at most marginally better than nothing.
- ► Consequently, even when p is very close to one, probability discounters would be willing to pay some fixed amount in order to trade A<sub>t+ε</sub>pC for A<sub>t+ε</sub> in the Continuity Money Pump.

## Mixture Continuity Violation $\rightarrow$ Money Pumped

- So, if we fix p, no matter how close to one, we can find a version of the Continuity Money Pump where the exploiter wins with probability p as long as we choose ε sufficiently close to zero.
- ► Therefore, an exploiter can get a fixed payment (up to the value of A<sub>t+ε</sub>) from the agent with only an arbitrarily small chance (1 − p) of having to give up anything.
- To summarize, views on which probabilities up to some discounting threshold are ignored violate Mixture Continuity. They are also vulnerable to exploitation in the Continuity Money Pump.

# Vulnerability to the Continuity Money Pump

- Probability discounters are vulnerable to exploitation in the Continuity Money Pump because arbitrarily small increases in probability, from just below the discounting threshold to just above it, can make a large difference to the value of a prospect.
- So, the Continuity Money Pump illustrates how probability discounters, who wish to ignore very small probabilities, do care a great deal about very small *changes* in probabilities.
- Nevertheless, unlike in the Independence Money Pump (discussed later), at least probability discounters would be paying for something, namely, for a small increase in the probability of a positive outcome (from just below the discounting threshold to just above it).
- Therefore, this money pump is not as worrisome as the Independence Money Pump.

# Summary

- To summarize, I've discussed two ways in which Probability Discounting might violate Continuity.
- First, I showed that views that discount probabilities below some threshold violate Continuity.
- Next, I showed that views that discount probabilities up to some threshold violate Mixture Continuity.
- Preferences that violate Continuity in these ways are vulnerable to exploitation by a money pump.

- ▶ Next, I'll show that Probability Discounting violates Independence.
- Then, I'll show how violating Independence renders probability discounters vulnerable to exploitation in a money pump for Independence.
- Finally, I'll discuss possible ways of avoiding exploitation in this case.

To see how Probability Discounting violates Independence, consider the following prospects:

#### Independence Violation

*Prospect* A Gives probability q of some very good outcome (and otherwise nothing).

*Prospect B* Certainly gives a good outcome.

*Prospect C* Certainly gives nothing.

- Let q be a probability that is above the discounting threshold but less than one.
- Suppose that the very good outcome is sufficiently great so that A is better than B.

Next, consider the following mixed lotteries:

#### Independence Violation

*Prospect ApC* Gives a probability p of A and a probability 1 - p of C (i.e., probability  $p \cdot q$  of a very good outcome and otherwise nothing).

*Prospect BpC* Gives a probability p of B and a probability 1 - p of C (i.e., probability p of a good outcome and otherwise nothing).

	p		1-p
	$p\cdot q$	p(1-q)	1-p
ApC	Very good	Nothing	Nothing
BpC	Good	Good	Nothing

### A VIOLATION OF INDEPENDENCE

- Given that B certainly gives a positive outcome, while A gives only a probability q of a positive outcome, we can mix A and B with C so that A mixed with C (i.e., ApC) gives only a negligible probability of a positive outcome but B mixed with C (i.e., BpC) gives a non-negligible probability of a positive outcome.
- ► This is so because there must be some probability p ∈ (0, 1) such that the result of q multiplied by p is below the discounting threshold, but p itself is above that threshold.

	p		1-p
	$p\cdot q$	p(1-q)	1-p
ApC	Very good	Nothing	Nothing
BpC	Good	Good	Nothing

### A VIOLATION OF INDEPENDENCE

- Suppose that the outcomes in question are monetary and that the utility of money equals the monetary amount.
- Then, there must be some p such that the probability-discounted expected utility of ApC is zero, but BpC has positive probability-discounted expected utility.
- In that case, Probability Discounting judges ApC to be worse than BpC.

	p		1-p
	$p\cdot q$	p(1-q)	1-p
ApC	Very good	Nothing	Nothing
BpC	Good	Good	Nothing

### A VIOLATION OF INDEPENDENCE

- Now, we have that A is better than B, but ApC is worse than BpC for some p ∈ (0, 1].
- ▶ This is a violation of the following axiom of Expected Utility Theory:

#### Independence

#### If $X \succ Y$ , then $XpZ \succ YpZ$ for all probabilities $p \in (0, 1]$ .<sup>8</sup>

- Informally, Independence is the idea that a lottery's contribution to the value of a mixed lottery does not depend on the other lotteries.
- The previous violation of Independence happens because, by mixing gambles together, one can reduce the probabilities of states or outcomes until their probabilities end up below the discounting threshold.
- As A gives a lower probability of a positive outcome than B does, with some values of p, ApC only gives a negligible probability of a positive outcome, while BpC still gives a non-negligible probability.

<sup>&</sup>lt;sup>8</sup>Jensen (1967, p. 173).

Violating Independence renders probability discounters vulnerable to exploitation in the Independence Money Pump. The case is as follows:<sup>9</sup>

The Independence Money Pump



 $A \succ B$ , and  $BpC \succ B^-pC^- \succ ApC$ .

The agent starts with prospect BpC: probability p of a good outcome and otherwise nothing.

<sup>9</sup>This money pump is from Gustafsson (2021, p. 31n21; 2022, p. 57). Also see Hammond (1988a, pp. 292–293; 1988b, pp. 43–45).

- At node 1, the agent is offered a trade from BpC to B<sup>-</sup>pC<sup>-</sup>, where B<sup>-</sup>pC<sup>-</sup> is just like BpC except that the agent has less money.
- If the agent turns down this trade and BpC results in the agent going up at chance node e, then at node 2, the agent will be offered a trade from B (certain good outcome) to A (probability q of a very good outcome and otherwise nothing).
- Both chance nodes depend on the same chance event *e*.

The Independence Money Pump



- ▶ The agent can use *backward induction* to reason about this case.
- This means that the agent considers what they would choose at later choice nodes and then takes those predictions into account when making choices at earlier choice nodes.<sup>10</sup>
- ▶ As the agent prefers A to B, they would accept the trade at node 2.

The Independence Money Pump



<sup>10</sup>Selten (1975) and Rosenthal (1981, p. 95).

- By using backward induction at node 1, the agent can reason that the prospect of turning down the trade at node 1 is effectively ApC, and the prospect of accepting the trade is B<sup>-</sup>pC<sup>-</sup>.
- ► Given that the agent prefers BpC to ApC, it seems plausible that there is some price ext{e} that they would be willing to pay to get the former instead of the latter.
- So, the agent pays that price and ends up with  $B^-pC^-$ .
- But they have ended up with B<sup>-</sup>pC<sup>-</sup> even though they could have kept BpC for free had they gone down at both choice nodes.
- Therefore, they have given up money for the exploiter.

The Independence Money Pump



# Avoiding exploitation in the Independence Money Pump

- Next, I'll discuss how probability discounters (and others who violate Independence) can avoid exploitation in the Independence Money Pump.
- I'll argue that none of the standard views, such as Resolute Choice and Self-Regulation, work.
- I'll also argue that even if vulnerability to exploitation is not a sign of irrationality, Probability Discounting has untenable implications in a version of the Independence Money Pump that might result in a loss.

- One decision policy that has been proposed as a solution to money pumps is Self-Regulation.<sup>11</sup>
- Self-Regulation forbids (if possible) choosing options that may lead via a rationally permissible route to a final outcome that is unchoiceworthy by the agent's own lights.
- The idea is that one ought not choose options that may (following one's preferences) lead to an outcome that one would not choose in a direct choice of all final outcomes.

<sup>&</sup>lt;sup>11</sup>Self-Regulation helps avoid exploitation in money pumps against cyclic preferences. See Ahmed (2017).

# Self-Regulation

- Unlike Resolute Choice (discussed later), Self-Regulation is forward-looking.
- When an agent's present choices determine the options available to them in the future, they should now choose so that their future choices lead to what they now consider acceptable in light of what is now available.
- If the agent now wants to avoid some final outcome O, and they know what they are going to do at later choice nodes, then they should (if possible) now choose in such a way that, given those later choices, they will not end up with O.

## Self-Regulation for Plans

- Self-Regulation in its original formulation does not help in the Independence Money Pump, as it was intended for money pumps that do not involve chance.<sup>12</sup>
- The Independence Money Pump involves chance nodes, so the agent does not know what the final holding will be.
- One way to adapt Self-Regulation to cases that involve chance is to apply it to plans.
- A plan specifies a sequence of choices to be taken by an agent at each choice node that can be reached from that node while following this specification.

 $<sup>^{12}\</sup>mbox{Rabinowicz}$  (2021, n. 13) writes: "[H]e [Ahmed, 2017] only shows how self-regulation allows the agents with cyclic preferences to avoid dynamic inconsistency. It is unclear whether and how this approach can be extended to agents who violate Independence."

## Self-Regulation for Plans

Self-Regulation with respect to plans then states the following:

#### Self-Regulation for Plans (i.e., Avoid Unchoiceworthy Plans)

If possible, one ought not choose options that may (following one's preferences) lead one to follow a plan that one would not choose in a direct choice of all plans (assuming one was able to commit to following some available plan).

Self-Regulation for Plans is a partial characterization of what it means to follow one's preferences: It involves, if possible, not choosing options that may, following one's preferences, lead one to follow an unchoiceworthy plan.

- The available plans at node 1 of the Independence Money Pump correspond to prospects ApC, BpC and B<sup>-</sup>pC<sup>-</sup>.
- One would not choose ApC or  $B^-pC^-$  in a direct choice between these plans.
- Therefore, one should not (if possible) choose any option that may lead via a rationally permissible route to one following ApC or B<sup>-</sup>pC<sup>-</sup>.

## Self-Regulation for Plans

- However, both accepting and rejecting the trade at node 1 of the Independence Money Pump lead the agent to follow one of these plans via rationally permissible routes.
- Rejecting the offer leads one to follow ApC; accepting the offer leads one to follow B<sup>-</sup>pC<sup>-</sup>.
- So, Self-Regulation for Plans is silent in this case because it is not possible to make choices that do not lead to unchoiceworthy plans via rationally permissible routes.
- Thus, Self-Regulation for Plans does not help avoid exploitation in the Independence Money Pump.
- To get out of trouble, probability discounters need to find some other decision policy.

### Avoid Exploitable Plans

Instead of accepting Self-Regulation for Plans, one might restrict the set of forbidden plans and accept the following decision rule:

#### Avoid Exploitable Plans

If possible, one ought not choose options that may (following one's preferences) lead one to pay for a plan that one could keep for free.

- Avoid Exploitable Plans forbids accepting the trade at node 1 of the Independence Money Pump because accepting it would be paying for something that one could keep for free.
- However, Avoid Exploitable Plans does not forbid choosing A over B at node 2 because doing so would not be paying for a plan that one could keep for free.
- Thus, at node 2, an agent using Avoid Exploitable Plans would choose A over B, given that they prefer the former.
- So, if one uses Avoid Exploitable Plans, one can avoid getting money pumped in the Independence Money Pump.

### The Three-Way Independence Money Pump

- However, in another decision problem, someone using Avoid Exploitable Plans would pay a higher price for something they could have obtained cheaper.
- I won't go into this case, but it is printed below:

The Three-Way Independence Money Pump



### Avoid Dominated Plans

- ▶ The focus on avoiding monetary exploitation may be misplaced.
- Instead, one might prefer adopting a decision rule that forbids all dominated plans whether or not they involve monetary exploitation:

#### Avoid Dominated Plans

If possible, one ought not choose options that may (following one's preferences) lead one to pay more for a plan that one could obtain for less money.

Avoid Dominated Plans forbids accepting the offer at node 1 of the Independence Money Pump because B<sup>-</sup>pC<sup>-</sup> is dominated by BpC.





### Avoid Dominated Plans

- However, Avoid Dominated Plans seems a too narrow decision policy.
- Self-Regulation for Plans forbids choices that lead to plans that are unchoiceworthy by the agent's own lights.
- In contrast, Avoid Dominated Plans only forbids choices that lead to dominated plans but allows choices that lead to unchoiceworthy plans (such as ApC).
- It seems difficult to motivate such a decision policy.
- Why would it be irrational to choose an option that leads to a dominated plan but not irrational to choose an option that leads to an unchoiceworthy plan? Allowing the latter but forbidding the former seems arbitrary.
- Moreover, it leads one to something that is worse than the dominated plan, namely, ApC.

### Avoid Dominated Plans

- Furthermore, if we change the probabilities in the Independence Money Pump slightly, then Avoid Dominated Plans no longer avoids exploitation, at least entirely.
- Now, instead of B<sup>−</sup>pC<sup>−</sup>, the agent faces B<sup>−</sup>qC<sup>−</sup>, where q is arbitrarily close to p (and q < p).</p>
- Then, given that B<sup>-</sup>qC<sup>-</sup> and BpC do not give the exact same probabilities of the relevant outcomes, Avoid Dominated Plans no longer forbids accepting the trade at node 1; it is not the case that B<sup>-</sup>qC<sup>-</sup> is like BpC except that the agent has less money, so Avoid Dominated Plans is silent.
- Consequently, a probability discounter who uses Avoid Dominated Plans will choose B<sup>-</sup>qC<sup>-</sup> even though they could have kept BpC for free, and q is arbitrarily close to p.
- So, Avoid Dominated Plans is vulnerable to a scheme that is arbitrarily close to exploitation.

## **Resolute Choice**

- Self-Regulation (and related principles) do not help probability discounters avoid monetary exploitation.
- But perhaps Resolute Choice will?
- A resolute agent chooses in accordance with any plan they have adopted earlier as long as nothing unexpected has happened since the adoption of the plan.<sup>13</sup>
- If one accepts Resolute Choice, one can make a plan that one will not trade B for A in node 2 of the Independence Money Pump.
- Even though one would usually prefer A over B, one is now committed to keeping B regardless.
- Consequently, one can safely refuse the trade at node 1, as one is then choosing BpC over B<sup>-</sup>pC<sup>-</sup>; one will not get money pumped nor choose the inferior prospect ApC.

<sup>&</sup>lt;sup>13</sup>Strotz (1955-1956) and McClennen (1990, pp. 12–13).

# Independence Violation (Negative)

- However, combining Probability Discounting with Resolute Choice gives untenable results in another case.
- Consider the following prospects:

### Independence Violation (Negative)

Prospect A Certainly gives nothing.

*Prospect B* Gives probability r of some very bad outcome and probability 1 - r of a barely positive outcome.

*Prospect C* Certainly gives a barely positive outcome.

- Let r be a probability above the discounting threshold but less than 1 r (i.e., less than 0.5).
- Suppose the very bad outcome in B is sufficiently bad so that A is better than B; certainly getting nothing is better than a non-negligible chance of a very bad outcome and otherwise a barely positive outcome.

# Independence Violation (Negative)

- Now, we get the same Independence violation as before (for similar reasons as before): A is better than B, but BpC is better than ApC for some p ∈ (0, 1].
- The probability of a very bad outcome is above the discounting threshold in B but below the discounting threshold in the mixed lottery BpC.

		p	1-p
	$p \cdot r$	p(1-r)	1-p
ApC	Nothing	Nothing	Barely positive
BpC	Very bad	Barely positive	Barely positive

#### **INDEPENDENCE VIOLATION (NEGATIVE)**

## Resolute Choice and the Independence Money Pump

- Recall that a probability discounter who uses Resolute Choice would commit to keeping B in node 2 of the Independence Money Pump.
- In this case, the agent would then choose a prospect that gives a non-negligible probability r of some very bad outcome and otherwise a barely positive outcome over the certainty of getting nothing.
- Earlier, we assumed that r is above the discounting threshold but less than 1 r. So, it could be, for example, 0.49.
- Then, the agent would choose a prospect that gives a 0.49 probability of a very bad outcome and otherwise a barely positive outcome over certainly getting nothing.
- Furthermore, the very bad outcome can be arbitrarily bad, while the barely positive outcome can be arbitrarily close to getting nothing.
- No reasonable theory recommends making this choice.

## Resolute Choice and Probability Discounting

- Appeals to Resolute Choice seem to provide a general means of answering dynamic choice arguments against various patterns of preferences.
- However, Probability Discounting combined with Resolute Choice leads to disastrous results.
- So, although Resolute Choice may help others who violate Independence avoid exploitation in the Independence Money Pump, it does not help probability discounters.

### How worrisome are the Independence Money Pumps?

- Probability discounters might argue that these money pumps are not worrisome because, for example, the agent only really faces prospects ApC and B<sup>-</sup>pC<sup>-</sup> at node 1 of the Independence Money Pump, given that they would choose A at node 2.<sup>14</sup>
- Thus, given the agent's preferences, in a way BpC is not even available to the agent. So, by choosing B<sup>-</sup>pC<sup>-</sup>, the agent does not end up paying for something they could have kept for free.





<sup>14</sup>See Levi (1997, p. 82n10) and Levi (2002, p. S241) for this point.

### How worrisome are the Independence Money Pumps?

- However, a money-pump argument is supposed to show that a given set of preferences is irrational because they lead to the agent paying for something they could have kept for free (if they had some other preferences).
- Therefore, it is not an adequate defense of those preferences that, given those preferences, the agent did not have any other option but to pay for something they could have kept for free.
- ▶ The target of the money pump is the structure of preferences.<sup>15</sup>
- If one's preferences lead one to pay for something one could have kept for free (if one had some other preferences), then the money pump has succeeded in showing that those preferences are irrational.

<sup>&</sup>lt;sup>15</sup>Steele (2010, p. 474) and Gustafsson (2022, p. 8n. 29, 14).

### How worrisome are the Independence Money Pumps?

- Furthermore, even if being exploited is not a sign of irrationality as this argument claims, the violation of Independence in the case that includes negative payoffs is worrisome independently of the exploitation it leads to.
- The reason for this is that the agent would choose to lock in a choice of keeping B (at node 2) if that was somehow possible at node 1.
- This means they would lock in a choice of a prospect that gives a 0.49 probability of a very bad outcome and otherwise a barely positive outcome over certainly getting nothing—which seems irrational.
- So, even if probability discounters do not accept Resolute Choice, they would still make the same choice of *B* over *A* if offered the chance to lock in the choice at node 1.
- This makes Probability Discounting less plausible as a theory of instrumental rationality.

# Summary

- To conclude, I've discussed possible ways of avoiding exploitation in the Independence Money Pump.
- First, I showed that Self-Regulation for Plans does not avoid exploitation in the Independence Money Pump.
- An agent who uses Avoid Exploitable Plans would pay too much for a plan in the Three-Way Independence Money Pump.
- Avoid Dominated Plans solves the Three-Way Independence Money Pump, but it is vulnerable to a scheme that is arbitrarily close to pure exploitation.
- Finally, Resolute Choice leads to untenable results in the negative version of the Independence Money Pump.

# Summary

- I also argued that locking in the choice of B over A at node 2 of the negative version of the Independence Money Pump is irrational—and that this is something probability discounters would do regardless of whether they accept Resolute Choice or not.
- So, even if vulnerability to exploitation is not a sign of irrationality, Probability Discounting has untenable implications in the negative version of the Independence Money Pump.
- All in all, what we learn from these money pumps is that the various possible ways of avoiding exploitation do not ultimately work.
- In addition, we learn that Probability Discounting gives untenable implications even if exploitation is not a sign of irrationality.

# Conclusion part 1/2

- Probability Discounting is one way to avoid fanatical choices in cases that involve tiny probabilities of huge payoffs.
- However, it faces some serious problems.
- First, I discussed two ways in which Probability Discounting might violate Continuity.
- I showed that views that discount probabilities below some discounting threshold violate Continuity.
- Also, I showed that views that discount probabilities up to some discounting threshold violate Mixture Continuity.
- As a result of these Continuity violations, Probability Discounting is vulnerable to exploitation in the Continuity Money Pump.

# Conclusion part 2/2

- In addition to violating Continuity, Probability Discounting also violates Independence. This renders probability discounters vulnerable to exploitation in the Independence Money Pump.
- I discussed some possible ways of avoiding exploitation (Self-Regulation for Plans, Avoid Exploitable Plans, Avoid Dominated Plans, Resolute Choice).
- However, these either failed to avoid exploitation in some version of the Independence Money Pump or they had otherwise untenable implications.
- To conclude, I've shown that Probability Discounting is vulnerable to exploitation in the money pumps for Independence and Continuity.
- The former is more worrisome than the latter, and it is difficult to see how Probability Discounting can respond to this challenge.

### References I

- Ahmed, A. (2017), 'Exploiting cyclic preference', *Mind* **126**(504), 975–1022.
- Beckstead, N. and Thomas, T. (2020), 'A paradox for tiny probabilities and enormous values'. Global Priorities Institute Working Paper No.10. URL: https://globalprioritiesinstitute.org/nick-beckstead-and-terujithomas-a-paradox-for-tiny-probabilities-and-enormous-values/
- Bostrom, N. (2009), 'Pascal's Mugging', Analysis 69(3), 443-445.
- Gustafsson, J. E. (2021), 'The sequential dominance argument for the Independence axiom of Expected Utility Theory', *Philosophy & Phenomenological Research* **103**(1), 21–39.
- Gustafsson, J. E. (2022), *Money-Pump Arguments*, Cambridge University Press, Cambridge.
- Hammond, P. J. (1988a), 'Orderly decision theory: A comment on professor Seidenfeld', *Economics and Philosophy* **4**(2), 292–297.
- Hammond, P. J. (1988b), 'Consequentialist foundations for expected utility', *Theory and Decision* **25**(1), 25–78.

## References II

- Hammond, P. J. (1998), Objective expected utility: A consequentialist perspective, *in* S. Barberà, P. J. Hammond and C. Seidl, eds, 'Handbook of Utility Theory Volume 1: Principles', Kluwer, Dordrecht, pp. 143–211.
- Herstein, I. N. and Milnor, J. (1953), 'An axiomatic approach to measurable utility', *Econometrica* **21**(2), 291–297.
- Jensen, N. E. (1967), 'An introduction to Bernoullian utility theory: I. Utility functions', *The Swedish Journal of Economics* **69**(3), 163–183.
- Kosonen, P. (2022), Tiny Probabilities of Vast Value, PhD thesis, University of Oxford.
- Levi, I. (1997), The Covenant of Reason: Rationality and the Commitments of Thought, Cambridge University Press, Cambridge.
- Levi, I. (2002), 'Money pumps and diachronic books', *Philosophy of Science* **69**(S3), S235–S247.
- McClennen, E. F. (1990), *Rationality and Dynamic Choice: Foundational Explorations*, Cambridge University Press, Cambridge.

# References III

- Monton, B. (2019), 'How to avoid maximizing expected utility', *Philosophers' Imprint* **19**(18), 1–24.
- Rabinowicz, W. (2008), Pragmatic arguments for rationality constraints, *in* M. C. Galavotti, R. Scazzieri and P. Suppes, eds, 'Reasoning, Rationality, and Probability', CSLI, Stanford, CA, pp. 139–163.
- Rabinowicz, W. (2021), Between sophistication and resolution—wise choice, *in* R. Chang and K. Sylvan, eds, 'The Routledge Handbook of Practical Reason', Routledge, Abingdon, Oxon; New York, NY, pp. 526–540.
- Rosenthal, R. W. (1981), 'Games of perfect information, predatory pricing and the chain-store paradox', *Journal of Economic Theory* **25**(1), 92–100.
- Selten, R. (1975), 'Reexamination of the perfectness concept for equilibrium points in extensive games', *International Journal of Game Theory* **4**(1), 25–55.

## References IV

- Smith, N. J. J. (2014), 'Is evaluative compositionality a requirement of rationality?', *Mind* **123**(490), 457–502.
- Steele, K. S. (2010), 'What are the minimal requirements of rational choice? Arguments from the sequential-decision setting', *Theory and Decision* **68**(4), 463–487.
- Strotz, R. H. (1955-1956), 'Myopia and inconsistency in dynamic utility maximization', *The Review of Economic Studies* **23**(3), 165–180.
- von Neumann, J. and Morgenstern, O. (1947), *Theory of Games and Economic Behavior*, 2 edn, Princeton University Press, Princeton.

Wilkinson, H. (2022), 'In defence of fanaticism', Ethics 132(2), 445-477.