Money Pumps and the Instrumentalist Foundations of Decision Theory

Johanna Thoma, Universität Bayreuth

Chapter forthcoming in *Problems of Choice: Normativity, Rationality, Value, and Morality*, edited by *M. S. Sagdahl* and A. Tanyi for *Routledge Studies in Ethics and Moral Theory*

Abstract: Many decision theorists take standard normative decision theory to be a theory of instrumental rationality, and its core requirements to be requirements of instrumental rationality. To justify these requirements instrumentally, decision theorists need to be clear on what they take to be the standard of instrumental rationality: What are the attitudes that are considered to pick out an agent's ultimate ends and how do they find decision-theoretic representation? This chapter argues that instrumentalist arguments in favour of one key decision-theoretic requirement, namely money pump arguments in favour of acyclicity, are guilty of a fatal equivocation about the standard of instrumental rationality. If the standard of instrumental rationality is preferences understood as all-things-considered attitudes to the fully described outcomes of our actions, we cannot show that being money-pumped is instrumentally irrational. If, however, the standard of instrumental rationality is partial attitudes to the various features of the outcomes of our actions – what we might think of as the multifarious instrumental reasons our all-things-considered attitudes are typically based on – we cannot show that instrumentally rational agents must adopt acyclical preferences to avoid being money-pumped. The chapter ends on a more optimistic note, that there is at least a promising conditional instrumentalist argument to be made that applies to many agents given some common human ends.

1. Introduction

Instrumental rationality requires agents to take the best means to their ends, but is silent on what ends agents ought to have. Many decision theorists assume that standard normative decision theory is concerned with instrumental rationality alone. For the most part, this instrumentalist, or 'Humean' interpretation of decision theory is so entrenched, it is rarely stated or explicitly argued

for.[1] To defend standard decision theory as a theory of instrumental rationality, we need to show that its requirements are requirements of instrumental rationality. Some requirements of standard decision theory — such as the requirement that agents should maximise with regard to their preferences — have appeared to many to be obvious requirements of instrumental rationality. Others, such as the requirement to have acyclical preferences, have been defended as requirements of instrumental rationality by appealing to various instrumentalist arguments.

Instrumentalist arguments point out that agents who violate the requirements of decision theory are prone to making a sure loss in some choice scenarios. According to the Money Pump Argument, which will be the main focus of this paper, an agent with preferences that form a strict cycle (e.g., she strictly prefers A to B, B to C, and C to A) can be offered series of choices where she will end up paying for something she could have had for free. Since being money-pumped in this way is taken to be instrumentally irrational, it is then argued that instrumental rationality requires agents to have acyclical preferences.

This paper aims to establish that at the heart of this instrumentalist defence of acyclicity lies an equivocation about the *standard of instrumental rationality*. Before we can even address the question of instrumentalist justification of core requirements of decision theory, we must answer a question of interpretation. Most importantly, which of the agent's attitudes are taken to pick out her ends, and how do these relate to the elements of standard decision theory? This question is important, because those attitudes will serve as the standard against which the agent's choices and other attitudes will be evaluated.

It is a pervasive assumption in philosophical decision theory that the agent's preferences over the objects of choice – understood as all-things-considered attitudes – form this standard of instrumental rationality against which the agent's actions and other attitudes are evaluated. In the case of choice under certainty, we can consider those objects of choice to be outcomes. The requirement that our choices ought to be guided by our preferences over outcomes, as captured,

[1] For an explicit statement of this instrumentalist interpretation of decision theory, see, e.g., the opening of Joyce's (1999) *The Foundations of Causal Decision Theory*, p.9. Lewis (1988) defends Humeanism about decision theory by arguing that one major form of non-Humeanism is not compatible with decision theory.

for instance, by the standard requirement to maximise, is plausible according to this preference-based notion of instrumental rationality. However, I will argue that the Money Pump Argument fails according to this standard, because we cannot establish that being money-pumped is instrumentally irrational. The intuition that it is instrumentally irrational to be money-pumped in fact relies on a different understanding of the standard of instrumental rationality, one that relates to underlying partial attitudes to features of outcomes – what we can think of as the multifarious instrumentalist reasons that all-things-considered preferences are based on. But according to this partial attitudes standard, the requirement to be guided by one's preferences in action, and for those preferences to be stable, is no longer justifiable in the context of the Money Pump Argument. And the Money Pump Argument only gets off the ground if agents act in accordance with stable preferences.

I thus show that the Money Pump Argument fails according to both ways of understanding the standard of instrumental rationality. If preferences understood as all-things-considered attitudes are the standard of instrumental rationality, being money-pumped can't be shown to be instrumentally irrational. If partial attitudes to features of outcomes are the standard of instrumental rationality, agents need not adopt acyclical preferences to avoid being money-pumped. They can simply act counter-preferentially, or temporarily adjust their preferences at the right points in time. Either way, the instrumentalist case in favour of acyclicity fails. Still, I will argue that the second way of understanding the standard of instrumental rationality is in fact more plausible. Moreover, I will conclude by showing that the partial attitudes-based standard of instrumental rationality also makes it possible, at least, to provide an instrumental justification for acyclicity that is conditional on the agent having certain desires. This, I argue, is the best instrumentalist case we can make for acyclicity.

2. Preference Cycles

Preference cycles over outcomes are both wide-spread and often appear sensible. Two types of cases are commonly used to motivate the idea that cyclical preferences may sometimes not be irrational. In one kind of case, outcomes differ along various dimensions the agent cares about.

And in the other kind of case, outcomes are in some respects seemingly indistinguishable to her. To start with the first kind of case, suppose I am looking for an apartment. Three apartments are available for the same rent, which I can afford. They differ only in terms of their size, their views, and the length of the daily commute to work. All three of these are factors I care about.

> **Apartment A:** 40 m^2 large; view onto a garden; 5-minute commute.
> **Apartment B:** 70 m^2 large; view onto the skyline, lake and woods; 60-minute commute.
> **Apartment C:** 100 m^2 large; view onto a brick wall; 30-minute commute.

When it comes to choosing where to live, my pair-wise preferences over the outcomes of living in each of these apartments (denoted by 'Apartment A', 'Apartment B', 'Apartment C') may well be cyclical. Suppose I have the following preferences over outcomes, where $\prec$ represents strict preference:

$$\text{Apartment A} \prec \text{Apartment B} \prec \text{Apartment C} \prec \text{Apartment A}$$

What may make these preferences seem defensible is that I can explain them in the following way: I prefer Apartment B over Apartment A because Apartment B is larger and has such a lovely view, and this outweighs the fact that it has a longer commute. I prefer Apartment C over Apartment B because Apartment C is even larger, and has a shorter commute, and this outweighs the fact that it does not have a good view. And I prefer Apartment A over Apartment C, because it has an even shorter commute, and a better view, and this outweighs the fact that it is smaller.

The second kind of case is best illustrated with the Puzzle of the Self-Torturer, first introduced by Quinn (1990). Suppose an evil scientist straps a device to your arm that causes you pain with electric shocks. The device has 1,000 different settings. At the first setting, it causes you no pain. At the highest, the pain is excruciating. However, adjacent settings differ so little in their electric current that you cannot distinguish them by the pain you feel when experienced subsequently. Now suppose you are offered \$10,000 in exchange for each setting you are willing to go up. And so each setting of the device is associated with an amount of money. Let S_1, S_2, S_3, ...S_{1000} be the outcomes of ending up with the level of pain and amount of money associated with the 1,000

different settings. It seems reasonable to have the following, cyclical preferences over these outcomes:

$$S_1 \prec S_2 \prec S_3 \prec \ldots \prec S_{1000} \prec S_1$$

Out of two adjacent settings, you always prefer the higher one. After all, you cannot detect a difference in pain between them when experienced subsequently, and $10,000 is a substantial amount of money. However, when you consider the highest setting, you find the amount of pain so unbearable that you would gladly forego the fortune associated with it in order to be pain-free at the lowest setting.

According to all well-known decision theories, the preferences we described in both cases are irrational, since they violate the requirement of acyclicity. Let X be the set of outcomes the agent's actions may bring about. Let $\succcurlyeq$ represent weak preference between outcomes: $x \succcurlyeq y$ if and only if the agent either strictly prefers x to y, or is indifferent between x and y. *Acyclicity* requires the following:

Acyclicity: For all $x_1, x_2, \ldots x_n \in X$, $x_1 \succ x_2, x_2 \succ x_3, \ldots x_{n-1} \succ x_n$ implies that $x_1 \succcurlyeq x_n$.

What I want to investigate in the following is whether we can provide an instrumental justification for acyclicity, as it seems we would need to, were we to provide instrumentalist foundations for decision theory.

3. The Money Pump Argument

Instrumentalist defences of requirements regarding the *structure* of an agent's preferences, such as acyclicity, typically point out that an agent with the preferences in question is prone to making a sure loss. In the case of acyclicity, this takes the form of the Money Pump Argument, first

formulated by Davidson et al. (1955), but going back to ideas in Ramsey (1950/1928).[2] We can illustrate it with our examples. To start with the first, suppose a rental agency gives me the chance to choose between Apartment A and Apartment B. Choosing according to my preference between these two apartments, I go with Apartment B. But then the agency offers me the opportunity to switch to Apartment C instead. Choosing again in accordance with my binary preference, I choose C. Now suppose I get offered the chance to switch to Apartment A, in exchange for a small fee, say $25. I prefer more money to less, other things being equal. But seeing that I have a strict preference for A over C, I probably still prefer A, even when I must pay $25 for it. If not, there will be a small enough positive amount of money ϵ that I will be willing to pay. My preferences are thus:

$$\text{Apartment A} \prec \text{Apartment B} \prec \text{Apartment C} \prec \text{Apartment A - } \epsilon$$

If I choose in accordance with my binary preferences at every point in time, I will end up with Apartment A having lost $25. But I could have had Apartment A without losing that money, if I had only refused to trade. Ending up with Apartment A having paid $25 seems instrumentally criticisable. Moreover, the rental agency could potentially repeat offering me this series of swaps, effectively turning me into a 'money pump'.

The same fate could meet you in the second example. Suppose you are offered the chance to go up by one setting every week, in exchange for the $10,000. Going with your binary preference between two adjacent settings, you should always go up by one setting, all the way to the highest setting. This way, you would turn yourself into the eponymous 'self-torturer'. If somebody now offers you the chance to go back to the lowest setting, in exchange for giving up your entire fortune, plus an additional $25 (or a small enough amount of money ϵ), you would gladly accept. But again, you could have been pain-free for less, making you apparently instrumentally criticisable. Moreover, the cycle could be repeated, turning the self-torturer into a money pump.[3]

[2] See Gustafsson (2022) for a comprehensive recent discussion of Money Pump Arguments for all key requirements of orthodox decision theory.

[3] Note that, in the examples as we have described them, foresight could help our agents. Schick (1986) originally proposed that if an agent can foresee that she is going to be offered the series of trades we just described, she will stop trading early on. Debates in dynamic choice theory have shown, however, that there are variations of these choice

Susceptibility to being money-pumped is widely held to provide a good justification for the requirement to have acyclical preferences. One rough way of cashing out the Money Pump Argument is as follows:

> **P1.** Agents with cyclical preferences can be placed in situations where they can't rationally avoid being money-pumped while retaining their cyclical preferences.
>
> **P2.** Being money-pumped is an instance of instrumental irrationality: Agents who end up money-pumped are not serving their ends well.
>
> **C.** Therefore, agents with cyclical preferences can't be instrumentally rational. Acyclicity is a requirement of instrumental rationality.

In addition, it is often taken to follow that, insofar as their preferences are under their control, the Money Pump Argument provides agents with cyclical preferences with a reason, grounded in their ends, to adopt acyclical preferences instead.

As stated, it is unclear why the conclusion should follow, given that some agents may never face the kinds of situations where there is a threat of being money-pumped. And presumably, to some extent, agents can actively avoid facing such situations. The following will not depend on how one fills in the details here. Instead, I am concerned with P1 and P2, which I take it will feature in some form in any plausible reconstruction of the Money Pump Argument. What I wish to argue in this paper is that there is no conception of the standard of instrumental rationality that allows us to establish both.

scenarios where foresight alone doesn't help. All we need, as Rabinowicz (2000) has shown, is persistence on the side of the money pumper, such that agents are offered further trades even after they have refused one.

4. Preference-Based Instrumental Rationality and the Money Pump Argument

At this point, let us return to the fundamental question of interpretation that the instrumentalist about decision theory faces, which I posed in the introduction: Which of the agent's attitudes are taken to pick out her ends, and how do these relate to the elements of standard decision theories? That is, which attitudes will serve as the standard of instrumental rationality? The most common response appears to be that, in the case of choice under certainty at least, this role is played by the preferences over outcomes that decision theory ascribes to agents. This is a natural position to take if we think of preferences over outcomes as all-things-considered attitudes to the objects of choice – after all, such attitudes aim to take into account everything the agent cares about. In the following, I will refer to this conception of the standard of instrumental rationality as 'preference-based instrumental rationality'.

In one respect, preference-based instrumental rationality seems to help us make the Money Pump Argument. It provides the best explanation of why instrumental rationality should require agents to act in accordance with their preferences in binary choice contexts. As we have seen, the Money Pump Argument relies on agents choosing in such a way. In fact, preference-based instrumental rationality seems to justify a more general requirement of *preference-guidance*, a requirement to fulfil, and to avoid frustrating one's preferences as much as possible in each of our choices. If instrumental rationality is about doing well by our preferences over outcomes, the best way of doing so would seem to be to act in accordance with those preferences.

Ultimately, however, the Money Pump Argument fails under preference-based instrumental rationality, because its proponents cannot establish the truth of P2. The most common way of defending P2 is as follows: Agents who are money-pumped fail to maximise with regard to their preferences. But maximisation is a requirement of instrumental rationality, which agents who are money-pumped thus violate. Maximisation is indeed a standard requirement of orthodox decision theory, and one way of spelling out the more general requirement of preference-guidance:

> **Maximisation:** Agents ought to choose an outcome such that no other available outcome is strictly preferred to it.

Note, however, that the agents in our money pump scenarios do not actually violate a requirement to maximise on any individual choice they face. As we said, they always choose in accordance with their binary strict preference between the two options they are facing. Money-pumped agents do fail to maximise *over time*, however, and thus only violate a diachronic version of the alleged requirement to maximise. Still, as proponents of the Money Pump Argument also sometimes highlight, agents with cyclical preferences will also fail to maximise in simple synchronic choices over the whole set of outcomes over which they have cyclical preferences (Gustafsson, 2013; Levi, 2002; Davidson et al., 1955).

Does the failure to maximise explain the irrationality of being money-pumped? As Andreou (2016) compellingly argued, on this way of justifying P2, the Money Pump Argument is question-begging. The problem is that those who don't already believe that cyclical preferences are irrational may simply respond that maximisation is not a plausible choice rule for agents with cyclical preferences, given it is impossible for them to follow the rule without changing their preferences. What Gustafsson, Levi and Davidson et al. seem to ignore is that the requirement to maximise is itself a principle that needs to be instrumentally justified, if decision theory is supposed to be a theory of instrumental rationality. If preferences are the standard of instrumental rationality, then choice rules should be justified in terms of their ability to satisfy the agent's preferences, rather than the other way around. It must be shown that agents need to abide by a choice rule to serve their preferences well. When the agent's preferences are cyclical, the question is thus what it would take for the agent to serve her cyclical preferences well. Unlike in the case of agents with acyclical preferences, the requirement to maximise does not qualify as a good general principle of choice for such agents, because it may be impossible for those agents to maximise without becoming a different kind of agent. But that need not mean that instrumentally rational choice is impossible for them.

In fact, choice rules that extend to cyclical preferences have been proposed that seem to capture preference-guidance for agents with cyclical preferences. For instance, according to a rule proposed in Schwartz (1972), an agent should choose a member of a subset of the available outcomes such that (i) no outcome outside of the subset is strictly preferred to any member of the

subset, and (ii) no proper subset of this subset fulfils condition (i). In our examples, if the agent is given a single choice between the outcomes over which she has cyclical preferences, according to this rule she is permitted to choose any of the options. An agent with cyclical preferences may thus have to frustrate some of her preferences. But the rule identifies a set of outcomes that seems to ensure no preferences are frustrated unnecessarily. Given the availability of plausible alternative choice rules for cyclical preferences, the violation of maximisation can't be what makes being money-pumped irrational.

There is another way in which one could argue in favour of P2. Rather than argue that agents with cyclical preferences violate some other requirement of instrumental rationality, namely maximisation, one could try to establish directly that, when they are money-pumped, agents with cyclical preferences are not serving their ends well. How might this alternative way of justifying P2 work within preference-based instrumental rationality? Essentially, what the argument would need to establish is that having different preferences would serve your preferences better. This appears to be only so if having different preferences comes with autonomous benefits, in terms of the preferences you currently hold – something you really want is only attainable if you adopt new preferences. For instance, if an evil demon were to severely punish you for continuing to have the preferences you currently have, and it is implied by your current preferences that you want to avoid such punishment at all cost (including the cost of abandoning some preferences you are strongly attached to), then it serves your preferences as they are now to have different preferences.

Is susceptibility to being money-pumped like this? To say so, we would have to show that being money-pumped is bad in terms of the agent's cyclical preferences over outcomes. That is, we must show that having cyclical preferences leads to an outcome that is bad in terms of those cyclical preferences, and that adopting different, acyclical preferences would lead to a better outcome according to those cyclical preferences. But preference-based instrumental rationality does not allow us to say so. First, many critics of cyclical preferences claim that cyclical preferences mean that there is no outcome that it would be rational for the agent to choose precisely because cyclical preferences do not pick out any outcome as 'best'. In fact, this is exactly what the first interpretation of the argument, which we just considered, relies on. But if we believe that, it would

also seem like those preferences cannot act as a standard of what alternative preference relation may serve the agent better.

Second, note that the Money Pump Argument exploits the following fact about the preferences of agents with cyclical preferences who prefer more money to less. If an agent displays a preference cycle over some outcomes, then she also has cyclical preferences over a set of outcomes that includes one of the original outcomes with some small amount of money deducted. In the apartment example, if I have these cyclical preferences:

$$\text{Apartment A} \prec \text{Apartment B} \prec \text{Apartment C} \prec \text{Apartment A}$$

I also have these cyclical preferences:

$$\text{Apartment A} \prec \text{Apartment B} \prec \text{Apartment C} \prec \text{Apartment A} - \epsilon \prec \text{Apartment A}$$

If preferences over outcomes are all we can go by, then if there is any outcome that it would be rational in terms of my cyclical preferences to end up with, at least in the first case it seems like each of these choices should be permitted. For all we have said, the agent's preferences treat each of the outcomes symmetrically. At the same time, to make the instrumentalist argument we are considering here, it needs to be instrumentally irrational, in terms of the agent's preferences, to end up with A - ϵ. There thus needs to be a difference between the cyclical preferences over A, B, and C, and the set of cyclical preferences over A, B, C and A - ϵ.

Intuitively, we in fact see the preference of A over A - ϵ as one that may not rationally be frustrated, in contrast to the other preferences in the cycle. It is this intuition that the Money Pump Argument relies on. But preference-based instrumental rationality cannot accommodate this intuition in a satisfactory way. The best explanation of why this preference is special has to do not with the structure of the agent's preferences over outcomes, but with the nature of the outcomes involved, and with further, more fundamental attitudes that the agent has to features of these outcomes.[4] This

[4] As discussed in more detail in Thoma (2025) and Andreou's (2025) response, there are some criteria appealing, at least explicitly, only to preferences over outcomes that one could use to rule out A - ϵ but not A, B, and C in many

calls for adopting the alternative conception of the standard of instrumental rationality introduced in the next section. As we will see, there are also independent good reasons for adopting this alternative standard.

5. Partial Attitudes-Based Instrumental Rationality and the Cost of Being Money-Pumped

We have seen that preference-based instrumental rationality cannot establish P2, despite the intuitive irrationality of being money-pumped. Preference-based instrumental rationality is in fact implausible on independent grounds.[5] I here want to briefly motivate an alternative,[6] before showing that while it can establish P2, it fails to support P1.

Preference-based instrumental rationality takes, in the case of choice under certainty, preferences over outcomes to form the standard against which actions are judged. Outcomes, in turn, are descriptions of all the circumstances an action may lead to that the agent cares about. This means that decision-theoretic preferences, if we understand them as attitudes, are all-things-considered

money pump scenarios, once we consider plausible additional preferences. For instance, plausibly, A - ϵ is "covered" by A in the sense first introduced by Miller (1980) in the context of tournament theory: A is preferred to it, and it ranks no higher than any other available option. This would be so if B is preferred to A - ϵ. However, ultimately it is not clear whether this criterion can be given a purely preference-based rationale. And moreover, it cannot always rule out agents being money-pumped, while intuitively, being money-pumped is always irrational. To start with worries that the criterion cannot be given a preference-based rationale, imagine we replace A - ϵ with some other outcome D that stands in no special relation to A — it is simply a fourth apartment. Yet, the preference relations are the same as just described, and D is covered by A. On a preference-based view, why should D be ruled out? After all, it is preferred to another option, C, which is not ruled out. Moreover, ending up with D involves frustrating precisely two preferences among the available options, which is no more than the number of preferences frustrated by ending up with C, which is not covered and hence allowed by Miller's rule. And so, the preference-based rationale for Miller's rule is at least not clear. More importantly, there are potential cases of agents being money-pumped that Miller's criterion does not rule out, and preference-based instrumental rationality cannot rule out for any other reasons, but that nevertheless strike us as irrational. In our case, it is intuitively irrational to end up with A - ϵ even if the agent's preferences are such that A - ϵ is not covered. For instance, this would be so if A - ϵ was preferred to B. Now one might say that given we have assumed ϵ is something the agent values, and she prefers B to A, it would be irrational for her to prefer A - ϵ to B. But this would be appealing to what, below, I will call the partial attitudes-based account of instrumental rationality, as it is ultimately information beyond outcome-preferences that makes it so plausible that the agent would prefer B to A - ϵ, given she prefers B to A. The fundamental problem is that the sense in which ϵ is an unambiguous loss for the agent relates to the features of the options involved and can only be made directly rationally relevant with partial attitudes-based standards, as they are introduced below.

[5] As also argued in more detail in Andreou (2022), who ultimately defends a different alternative standard.

[6] See also my Thoma (2021a) for a more detailed defence.

kinds of attitudes. In our first example, at a minimum, the outcome Apartment A would consist in a description of the size of the apartment, the length of my daily commute, and the views from the apartment. In the self-torturer case, the outcomes would include descriptions of both the level of pain I will feel, and the amount of money I will have. In line with preference-based instrumental rationality, decision theorists have implemented this idea in a way that appeals to only preferences over outcomes. According to Joyce (1999, p.52), whenever there is some circumstance such that an agent would strictly prefer an outcome in the presence of that circumstance to the same outcome in the absence of that circumstance, the outcome has been underspecified. Clearly, this rule for specifying outcomes will lead to outcomes being very detailed descriptions of states of affairs that may come about as the result of my actions. In fact, moving to more detailed descriptions of outcomes is a common move made in order to accommodate apparent violations of the standard axioms of decision theory, including acyclicity. And so those who want to defend acyclicity need to embrace very fine-grained outcomes as the object of preference. But it is the fine-grained nature of these outcomes that makes them implausible candidates for the object of the conative attitude that should form the standard of instrumental rationality.

Let me first highlight that we do not ordinarily think of such detailed descriptions of states of affairs as the objects of our desires, as it seems the proponent of preference-based instrumental rationality would need to hold. What we usually claim to desire in ordinary discourse are simpler states of affairs, which are features of fully described outcomes. For instance, I might say, 'I desire to drink a glass of fizzy water right now', 'I desire to have more time to practise viola', or 'I desire to sail the Inside Passage'. In the last case, the object of my desire is not a complete description of one course that my life could take in which I sail the Inside Passage - complete with the description of what flavour ice-cream I will have after dinner tonight. The object of my desire is simply my sailing the Inside Passage. We also often speak of preferences regarding such simpler states of affairs. I may say that I prefer sailing the Inside Passage to sailing the Northwest Passage, for instance. The objects of this preference are not fully specified outcomes. In fact, there are many outcomes involving me sailing the Northwest Passage that I would prefer to many outcomes involving me sailing the Inside Passage. I will prefer an outcome involving me sailing the Northwest Passage if that outcome also involves you giving me a million dollars by the end of it. But this does not make it any less true that I prefer sailing the Inside Passage to sailing the

Northwest Passage in the ordinary sense just described. My desire to also have a million dollars is irrelevant to this preference.

The alternative account of the standard of instrumental rationality I want to propose here evaluates the agent's actions in terms of her attitudes, be they desires or preferences, over such simple states of affairs. Outcomes, as full descriptions of everything the agent may care about, comprise many simpler states of affairs. These simple states of affairs are features of full outcomes. Attitudes to such features of outcomes are in that sense *partial* attitudes. I will hence refer to the alternative conception of the standard of instrumental rationality as "partial attitudes-based". This alternative notion of instrumental rationality is not only more plausible than preference-based instrumental rationality. As we will see, it can also express what is intuitively instrumentally irrational about being money-pumped.

One reason to err towards such a partial attitudes-based notion of instrumental rationality is that it intuitively seems like desires for (or preferences over) simpler states of affairs are more basic and explain preferences we may have over outcomes. If I prefer Apartment A to Apartment C, it is because I desire to have a beautiful view and a short commute, and this outweighs my desire to have a large living space. Pettit (1991) goes so far as to refer to this idea as a 'platitude of desiderative structure'.[7] Reflecting on the way we make decisions in the context of conflicting desires provides further support for this 'platitude'. In these contexts, we do not generally come readily equipped with all-things-considered preferences over full outcomes, but rather explicitly consult the 'pros and cons' of the various options.

Partial attitudes-based instrumental rationality goes further than this explanatory claim and says that our actions are ultimately rationally answerable to our attitudes to features of outcomes, not to our all-things-considered preferences over outcomes. This again seems intuitively plausible. The kind of reasoning we are engaged in when forming preferences over outcomes on the basis of our underlying partial attitudes appears to be instrumental reasoning. If this reasoning goes wrong,

[7] See my Thoma (2021b) for arguments that preferences over outcomes as they feature in decision theories in fact do not offer satisfactory folk psychological explanations on their own, without appealing to attitudes to features of outcomes.

and I act on my outcome-preferences, this seems to be an instrumental failure. In fact, it is a kind of rational failure that is familiar, especially when forming preferences over outcomes involves weighing up many kinds of consideration. Yet, preference-based instrumental rationality cannot cast this as an instrumental failure.

Assuming that this alternative account of the standard of instrumental rationality is correct, much more needs to be said on exactly how an agent's preferences over outcomes and her actions should be based on her partial attitudes. In fact, there is a growing literature in decision theory on how all-things-considered preferences are formed from more basic values (e.g., Dietrich and List, 2013). Luckily, to account for the irrationality of being money-pumped, we only need to appeal to one minimal requirement of partial attitudes-based instrumental rationality: one should not frustrate one of one's partial attitudes unnecessarily, that is, without in return better satisfying another partial attitude to a greater extent. This is the principle the money-pumped agent violates: She is deprived of money without receiving anything else she values in return, and her desire for money is unnecessarily frustrated.

More formally, I propose the following partial attitudes-based principle to capture the irrationality of being money-pumped:

Q: It is irrational to make a series of choices when

a. it leaves one with an outcome y, when there would have been an alternative series of choices which leaves one with outcome x, which is identical to y, except that it has one additional feature one desires under the circumstances, or except that it differs to y with regard to only one feature such that x's feature is preferred to y's under the circumstances,[8] and

[8] I mean features of outcomes to be individuated such that they themselves are not desired or preferred by the agent in one respect, while she is averse to them in another respect. For instance, if there is some respect in which I am averse to having more money, I do not violate principle Q when I am money-pumped. I will assume for the sake of argument that money is not like that for our agent, and that she unambiguously desires to have more money under her present circumstances. If this is too far-fetched, we can reformulate the Money Pump Argument with appeal to a more basic good.

b. there is an alternative series of choices of which (a) is not true, and within which each individual choice is permissible given one's partial attitudes concerning the outcomes still available at the time of action.

This principle claims that it would be irrational for an agent to be money-pumped if she could have avoided being money-pumped by taking a series of actions each of which is itself unproblematic in terms of partial attitudes-based instrumental rationality. It is a diachronic principle, as it must be in order to apply to the diachronic money pump scenarios.[9] Q provides us with a plausible partial attitudes-based account of what is irrational about being money-pumped, and thus with support for P2.

6. Partial Attitudes-Based Instrumental Rationality and Alternative Ways to Avoid Being Money-Pumped

We said above that the success of the Money Pump Argument relies on us being able to establish both P2, that being money-pumped is indeed instrumentally irrational, and P1, that agents with cyclical preferences can find themselves in situations where they can't rationally avoid being money-pumped while retaining their cyclical preferences. We have argued that preference-based instrumental rationality cannot establish P2 for agents with cyclical preferences. The partial attitudes-based alternative we sketched in the last section, however, arguably can. Moreover, we argued that this is the more plausible account of the standard of instrumental rationality for independent reasons. The important question now is whether we can establish P1 according to this alternative account of the standard of instrumental rationality. I will argue that we cannot.

Some authors have defended dynamic choice rules that may help agents with cyclical preferences avoid being money-pumped without giving up their cyclical preferences. One such choice strategy is what McClennen (1990) calls 'resolute choice'. Resolute agents pick a sequence of actions in

[9] This principle is similar to the diachronic part of Andreou's (2016, p.1454) principle P, adapted to the partial attitudes-based account of instrumental rationality developed here, and weakened to have less problematic implications for cases where an agent's ends (here conceived of in terms of partial attitudes) change over time, as in what are sometimes called 'temptation cases' (e.g. Gauthier, 1996).

the beginning of a series of choices, in accordance with their preferences at that point, and then simply go through with that sequence of actions, even if it may conflict with the preferences the agent has over the continuation of the decision problem at future choice points. In our money pump scenarios, a resolute agent would choose as she would in a synchronic choice over all the outcomes she may reach in the series of trades she will be offered. Partial attitudes-based instrumental rationality would require her not to choose to be money-pumped in such a synchronic choice. Resolute choice, if it is rationally defensible, would thus keep an agent with cyclical preferences from being money-pumped. In fact, being resolute is not the only way of avoiding being money-pumped while keeping one's cyclical preferences. Any way of stopping the trading early will do so.[10]

Standard worries about these dynamic choice rules are usually expressed in terms of preference-based instrumental rationality, and thus, I think, miss the point in the context of our discussion. These worries tend to focus on the supposed instrumental irrationality of the counter-preferential choices presumably licensed by these alternative choice rules. Counter-preferential choice may be obviously instrumentally irrational if preferences themselves form the standard of instrumental rationality. But, as I have argued above, preference-based instrumental rationality cannot ground the Money Pump Argument anyway. What defenders of the Money Pump Argument need to show is that alternative ways of avoiding being money-pumped are incompatible with instrumental rationality according to the partial attitudes-based alternative we described in the last section. But the irrationality of counter-preferential choice is not so obvious on this picture. I will argue in the following that indeed, partial attitudes-based instrumental rationality allows for alternative ways of avoiding being money-pumped, both by choosing counter-preferentially, and by temporarily altering one's preferences.

According to partial attitudes-based instrumental rationality, attitudes to features of outcomes are the fundamental conative attitudes that pick out an agent's ends, and that instrumental rationality requires her to serve effectively. This raises the question of how the preferences over outcomes that feature in formal decision theories relate to those attitudes. As long as we continue to think of

[10] See, for instance, Rabinowicz (2014), and Ahmed (2017) for alternative dynamic choice strategies that also allow agents to avoid being money-pumped.

preferences over outcomes as conative attitudes in their own right, as is standard, the most natural way to understand our preferences over outcomes is as all-things-considered comparative evaluations of outcomes on the basis of our attitudes over all the states of affairs the outcomes comprise.[11]

Does partial attitudes-based instrumental rationality rule out counter-preferential choice? We can immediately note one caveat regarding the more general requirement of preference-guidance under partial attitudes-based instrumental rationality: Unless we take outcome-preferences to be infallible as all-things-considered attitudes that express the agent's partial attitudes, it can happen that the agent's preferences misrepresent her underlying concerns. We already noted the intuitive possibility of the instrumental failure involved in such misrepresentation. And so, maximisation, or any other norm of preference-guidance, can be required at best conditionally: Agents ought to maximise if their outcome-preferences represent their partial attitudes correctly.

In its conditional form, preference-guidance seems, at first sight, well-supported by partial attitudes-based instrumental rationality. As a matter of fact, the consequences of our choices are full outcomes: I choose to live in an apartment, along with all that that implies. I do not only choose a beautiful view. If my outcome-preferences correctly capture all the partial attitudes to the states of affairs the outcome comprises, then acting on the preference means that my action is still ultimately based on those partial attitudes.

But preference-guidance cannot be justified in full generality even this conditional form. The core problem is that letting one's preferences guide one's actions is rationally required by partial attitudes-based instrumental rationality only if there is only one preference relation over outcomes that is admissible given the agent's underlying partial attitudes. Let me call this condition *uniqueness*. If it holds, then for each pair of outcomes, the agent's partial attitudes uniquely determine whether the agent should prefer one over the other or be indifferent. In that case, it seems like partial attitudes-based instrumental rationality indeed requires preference-guidance. But suppose uniqueness fails: Several different preference relations over outcomes would equally well express the agent's partial attitudes. And suppose the agent only adopts one of these

[11] See Hausman (2012) for a comprehensive defence of such an interpretation of preference.

permissible preference relations.[12] Partial attitudes-based instrumental rationality now no longer requires that the agent is guided by the preferences she adopted, which is what we would need for the Money Pump Argument to go through. If she chooses in accordance with preferences she does not have, but that would have been admissible given her partial attitudes then she is not instrumentally criticizable according to partial attitudes-based instrumental rationality. She would be serving her partial attitudes no worse than had she acted on her actual preferences.

There are good independent reasons to think that uniqueness frequently fails, in particular when there is some form of fuzziness (vagueness, incompleteness, imprecision) in an agent's attitudes, as argued in my Thoma (forthcoming). How common non-uniqueness is precisely depends on how restrictive our full theory of partial attitudes-based instrumental rationality is beyond the minimal principle Q. But I here merely need to point out that proponents of the Money Pump Argument are not entitled to assume uniqueness, in which case they cannot establish preference-guidance and rule out alternative ways of avoiding being money-pumped. On the partial attitudes-based picture, the aim of the instrumentalist justification of acyclicity would be to show that agents are required to form acyclical preferences in order to serve their partial attitudes well. For there to be any hope for this endeavour, for any agent, there must be at least one acyclical preference relation that correctly captures her underlying partial attitudes. I will grant this here. But now the question of uniqueness arises: Is there one such acyclical preference relation that uniquely expresses the agent's partial attitudes correctly?

This can't in fact be so if we want to deliver a Money Pump Argument that does any justificatory work. Suppose that for some agent with cyclical preferences, uniqueness holds regarding some acyclical preference relation. In that case, her actual cyclical preferences must be a mistaken expression of her underlying partial attitudes. Now, either the Money Pump Argument establishes this, or we know this independently. If we know this independently, then the Money Pump Argument does no work. But proponents of the Money Pump Argument presumably think the

[12] Another possibility would be that in these kinds of cases, an agent should adopt a family of preference relations, expressing an 'imprecise preference'. However, in that case, we would already have given up on preference-guidance in the form of maximisation or Schwartz's rule described above, since we would have to formulate new choice rules for families of preference relations. How we could then give an instrumental justification for each preference relation in the family being acyclical is unclear.

argument actually helps to justify acyclicity. Moreover, the intuitive plausibility of cyclical preferences in our examples in fact speaks against us always having independent reason to think the preferences are mistaken.

It must thus be the Money Pump Argument that establishes that the agent's cyclical preferences are mistaken representations of her underlying partial attitudes. The reasoning could be that if the agent's preferences over outcomes capture her partial attitudes fully, then actions guided by the preferences should not frustrate those very attitudes. But this argument takes for granted that agents should always be guided by their preferences in action, which begs the question.

Therefore, we cannot assume uniqueness if we want to make a successful Money Pump Argument. We should accept non-uniqueness at least in the sense that both the original cyclical preferences, as well as at least one acyclical preference relation are permissible representations of the underlying partial attitudes. The latter is needed if the Money Pump Argument is to be successful, and the former takes the intuitive plausibility of the agent's original preferences seriously. The problem now is that any kind of non-uniqueness implies that there is no requirement, in terms of partial attitudes-based instrumental rationality, that the agent's actions should be guided by the preferences she actually adopts, for the reasons given above.

What does this mean for the Money Pump Argument? The Money Pump Argument only gets off the ground if the agent is guided by her preferences in action at every point in time, and moreover if those preferences are stable. If she is to be guided by stable preferences, then she needs to adopt acyclical preferences to guarantee she can't be put in a situation where she is money-pumped. We have just found that, under partial attitudes-based instrumental rationality, we can no longer give an independent justification for preference-guidance. But one might think that the Money Pump Argument provides a *joint* instrumental justification for acyclicity and preference-guidance. After all, given stable preferences over time, acyclicity and preference-guidance together guarantee that the agent is not money-pumped in the ways we described.

But the Money Pump Argument does not provide a joint justification for acyclicity and preference-guidance, because there are at least two kinds of ways of avoiding being money-pumped that

involve violating one or both of these principles. First, one could violate both and avoid being money-pumped by keeping one's cyclical preferences and refraining from always being guided by them in action. This alternative response to money pump scenarios may make good sense in cases where cyclical preferences are especially tenacious, such as in the Self-Torturer Problem. In light of the pairwise indiscernibility in pain of adjacent settings, you might find your strict preference for the higher of any two adjacent settings hard to shake. But you might still acknowledge that serving your respective desires for money and for being pain-free well requires you to stop at a sensible setting – against your preference at that point in time. In fact, this coheres well with at least one prominent solution to the puzzle of the self-torturer (Tenenbaum and Raffman, 2012), and more generally coheres with the standard way of understanding resolute choice and other alternative dynamic choice rules.

Second, and less obviously, one could avoid being money-pumped while holding on to acyclical preferences even without violating preference-guidance, namely if one displays unstable preferences. In the presence of non-uniqueness, one could achieve this by switching between different rationally permissible preference relations over time. What we need for an agent to not be money-pumped is for her to stop trading at some point before she has lost money. Even for an agent that abides by preference-guidance, it is enough for the agent to have a preference to stop at that one point in time when somebody attempts to money pump her. But this is compatible with her having cyclical preferences at every point in time.

Consider the Self-Torturer Problem, for instance. For the agent to avoid being money-pumped, it just needs to be true that at some point, she will prefer to stick with the lower of two settings she is offered. Although that temporarily breaks the preference cycle that we originally characterised her with, her preference relation as a whole need not be acyclical, even at that point in time. Suppose, for instance, that at the point in time when she is offered the choice between S_{300} and S_{301}, she prefers to choose S_{300}. She could at the same time retain all her other preferences, including, for instance, the following cycle:

$$S_1 \prec S_2 \prec \ldots S_{300} \prec S_{302} \prec \ldots S_{1000} \prec S_1$$

Let us assume that S_{300} is also the setting that would be picked out as the last permissible stopping point by the acyclical preference relation the agent would adopt, were she to pick one. If both the original cyclical preference relation as well as an acyclical preference relation with the last permissible stopping point of S_{300} are appropriately responsive to the agent's underlying partial attitudes, it is eminently plausible that the preferences just described also express the agent's partial attitudes correctly. After all, they are only minimally different from the cyclical preferences that seemed most natural to the agent. They just include the minimal change necessary to keep the agent from being money-pumped. If anything, they are thus more accurately based on the agent's underlying attitudes than fully acyclical preferences would be.

Of course, given the preference cycle still contained in the agent's preferences, the agent is still in danger of being money-pumped when she is offered a different series of trades — namely one that leaves out setting S_{301}. But actually being money-pumped would require that the agent's preferences remain stably what they are over that series of choices. Again, the agent could avoid being money-pumped by adopting a crucial preference to not trade any further at some point in time in that new series of trades. If we give up on the idea that the agent needs to have stable preferences over time, and across different choice situations, then agents can avoid being money-pumped without having fully acyclical preferences, and even while abiding by preference-guidance. In fact, we could even interpret the resolute strategy and other alternative dynamic choice rules discussed earlier along these lines: Instead of involving counter-preferential choice, we could understand these choice rules as inducing a preference change within a specific dynamic choice problem only.

Importantly, this second alternative way of avoiding being money-pumped also applies to an alternative understanding of what preferences are than the one we have been working with. Instead of as all-things-considered conative attitudes to outcomes – as we must think of them when working within preference-based instrumental rationality – preferences are sometimes interpreted behaviourally, as dispositions to choose, or simply as stable choice patterns, in particular in economics. Elsewhere, I have in fact argued in favour of such an interpretation of preference (Thoma 2021c). Much of what we have said about preferences would still apply to such a notion of preference: It would have to have as its object fine-grained outcomes, and, within partial

attitudes-based instrumental rationality, should be ideally response to the sum of an agent's underlying partial attitudes. Non-uniqueness should for the same reasons be assumed in the context of the Money Pump Argument. But, one might think, counter-preferential choice is conceptually impossible on such a notion of preference. If preference just is choice, how could one make choices out of line with one's preferences? This could be taken to help make the Money Pump Argument, because it would rule out the alternative way of avoiding being money-pumped involving counter-preferential choice. However, it follows from what we have said that even under this notion of preference, the conclusion of the Money Pump Argument can be avoided. There is a second alternative way of avoiding being money-pumped: Agents can retain acyclical preferences by displaying unstable preferences, that is, in the present context, unstable patterns of choice.

The Money Pump Argument goes through only if adopting stable, acyclical preferences and acting in accordance with them is the only rationally permissible way of avoiding being money-pumped. In the face of the two kinds of alternatives presented here, is there nevertheless something that counts in favour of this route? One potential consideration is that we are often not in a position to know what further choices we will be offered, and we sometimes forget what choices we were offered in the past. If that is so, we might fail to act counter-preferentially, or fail to adopt preferences that are unstable in just the right way to avoid being money-pumped. A strategy that requires me to adopt, in one way or another, a disposition to choose specifically to avoid being money-pumped requires me to keep track of potential money pumps to try and avoid them. But often, this will be difficult. Adopting stable, acyclical preferences circumvents this difficulty. With such preferences, agents cannot just stumble into being money-pumped.

However, there is a related disadvantage to the strategy of adopting stable, acyclical preferences. And that is that it requires an agent to have specific acyclical preferences even in situations where there is no danger of her being money-pumped. And that may be unnecessarily restrictive in cases where cyclical preferences seem at least initially most natural. This comes out especially in the Self-Torturer Problem. Suppose that the agent avoids being money-pumped in the series of choices we described above by adopting acyclical preferences that have her stop at S_{300}. If those are her stable preferences, then the agent is presumably also required to choose S_{300} over S_{301} at other

points in time, in situations where she is only offered that one choice. But that seems unnecessarily restrictive. Much speaks in favour of choosing S_{301} in that situation.

To offer a Money Pump Argument in favour of acyclicity as an unconditional requirement, we need agents to decisively side with the strategy of adopting stable, acyclical preferences in the face of potential money pumps. And there is no such decisive reason, independently of the agent's other desires. At best, there are competing considerations that count in favour of adopting one or the other strategy of avoiding being money-pumped. Partial attitudes-based instrumental rationality thus does not require agents to adopt acyclical preferences independently of the content of the agent's desires. The project of providing a general instrumental justification for acyclicity has thus reached a dead end.

7. Conclusions and a Potential Way Forward

If we want to understand decision theory as a theory of instrumental rationality, we need to provide instrumental justifications for its central requirements. One such central requirement is acyclicity. The standard instrumentalist defence of this requirement is the Money Pump Argument. The argument aims to show that agents who violate acyclicity can be placed in situations where they end up money-pumped. Being money-pumped, in turn, is deemed to be instrumentally irrational. I have argued that this argument fails to provide a general instrumental justification of acyclicity. In fact, the common acceptance of the argument seems to rely on a fatal equivocation about the standard of instrumental rationality.

The Money Pump Argument, in its standard form, presupposes that agents choose in accordance with stable preferences over outcomes. A requirement of preference-guidance in fact seems to be well justified if we take preferences over outcomes to be the fundamental conative attitude which forms the standard of instrumental rationality. But in order to show that being money-pumped is indeed instrumentally irrational, we need to reject preferences over outcomes as the standard of instrumental rationality. Instead, I argued, we should adopt a partial attitudes-based notion of instrumental rationality. According to this notion, instrumental rationality requires an agent to do

well by her partial attitudes to features of outcomes. But on this notion, preference-guidance no longer holds as a general requirement in money pump scenarios, and neither do agents generally have to stick to a stable preference relation over time. And then agents can rationally avoid being money-pumped without adopting acyclical preferences.

Confidence in the Money Pump Argument thus seems to be based on equivocation about the standard of instrumental rationality. On one way of understanding the standard of instrumental rationality, preference-guidance is plausible, so that agents with cyclical preferences who abide by it indeed end up money-pumped. But we cannot show that this is instrumentally irrational. On the other way of understanding the standard of instrumental rationality, we can explain why being money-pumped is instrumentally irrational. But agents with cyclical preferences can rationally avoid being money-pumped. Ultimately, the Money Pump Argument fails to provide an unconditional instrumental defence of acyclicity on either way of thinking about the standard of instrumental rationality.

What we haven't ruled out, however, is that the Money Pump Argument could successfully provide a conditional instrumental defence of acyclicity. I here want to point to a promising way of doing so under the assumption of partial attitudes-based instrumental rationality, which we argued provides the more plausible picture of instrumental rationality in any case. If Q is a requirement of partial attitudes-based instrumental rationality, then being money-pumped is arguably instrumentally irrational. The problem was that there are ways of avoiding being money-pumped that do not involve adopting acyclical preferences, and that are not generally ruled out by instrumental rationality. However, some agents may have desires that favour the response of adopting acyclical preferences, which would rationally rule out the other possible responses. For those agents, the Money Pump Argument would justify a requirement of acyclicity. The Money Pump Argument can thus provide us with an instrumental defence of acyclicity that is conditional on the desires in question.

When we are operating with the behavioural conception of preference sketched in the last section, one plausible candidate for such a desire is the desire to have stable preferences, that is, stable choice dispositions, over time and across different choice contexts. Such a desire helps because it

effectively resolves the problem of non-uniqueness: If an agent has a desire for stability of preference, then once she has settled on a preference relation, this preference relation becomes the only relation that expresses her partial attitudes correctly from then on. In the money pump scenario, the strategy of avoiding being money-pumped by adopting cyclical preferences that are unstable in just the right way then performs worse according to the agent's partial attitudes than the strategy of adopting stable acyclical preferences. The agent will be frustrating her desire for stable preferences when another available course of action, which would have also been permissible given the agent's other partial attitudes, would have not done so. In fact, avoiding being money-pumped without adopting stable acyclical preferences would thus also violate principle Q.

We can hence make a successful Money Pump Argument in favour of acyclicity that is conditional on a desire for stable choice dispositions. And in fact, there may be something attractive about being settled in one's choice dispositions in this way. In addition to requiring less alertness to potential sure loss-inducing choice situations, as mentioned in the last section, once we look at decision problems involving interpersonal interaction, stability and thus reliability may also be an advantage in cooperation. Still, and importantly, instrumental rationality cannot *require* us to have the desire to have stable preferences over time. This is because instrumental rationality is supposed to be silent about what desires an agent may have. As long as we remain Humeans about decision theory, the best we can say about acyclicity is that agents who have the right kinds of desires, such as a desire for stable preferences, must abide by it. For the rest of us, instrumental rationality is more permissive.

References

Ahmed, A. (2017). Exploiting cyclic preference. *Mind*, 126(504), 975–1022.

Andreou, C. (2016). Cashing out the money-pump argument. *Philosophical Studies*, 173(6), 1451–1455.

Andreou, C. (2022). *Choosing Well: The Good, the Bad, and the Trivial*. Oxford University Pres.

Andreou, C. (2025). Reflections on choosing well: Reply to commentators. *Philosophia*, 52, 1279–1288.

Davidson, D., McKinsey, J. C. C. , & Suppes, P. (1955). Outlines of a formal theory of value, I. *Philosophy of Science*, 22, 140–160.

Dietrich, F., & List, C. (2013). A reason-based theory of rational choice. *Nous*, 47(1), 104–134.

Gauthier, D. (1996). Commitment and choice. In F. Farina, S. Vannucci, and F. Hahn (Eds.), *Ethics, Rationality, and Economic Behaviour*, (pp. 217–243). Oxford University Press.

Gustafsson, J.E. (2013). The irrelevance of the diachronic money-pump argument for acyclicity. *The Journal of Philosophy*, 110(8), 460–464.

Gustafsson, J.E. (2022). *Money-Pump Arguments*. Cambridge University Press.

Hampton, J. (1995). Does Hume have an instrumental conception of reason? *Hume Studies*, 21(1), 57–74.

Hausman, D. (2012). *Preference, Value, Choice, and Welfare*. Cambridge University Press.

Hume, D. (2007/1739). *A Treatise of Human Nature*. Clarendon Press.

Joyce, J. (1999). *The Foundations of Causal Decision Theory*. Cambridge University Press.

Levi, I. (2002). Money pumps and diachronic books. *Proceedings of the Philosophy of Science Association*, 3, S235–S247.

Lewis, D. (1988). Desire as belief. *Mind*, 97, 323–332.

McClennen, E. (1990). *Rationality and Dynamic Choice: Foundational Explorations*. Cambridge University Press.

Miller, N. (1980). A new solution set for tournaments and majority voting: Further approaches to the theory of voting. *American Journal of Political Science*, 24(1), 68–96.

Pettit, P. (1991). Decision theory and folk psychology. In M. Bacharach & S. Hurley (Eds.), *Foundations of Decision Theory: Issues and Advances* (pp. 147–175). Blackwell.

Quinn, W. (1990). The puzzle of the self-torturer. *Philosophical Studies*, 59(1), 79–90.

Rabinowicz, W. (2000). Money pump with foresight. In M. Almeida (Ed.), *Imperceptible Harms and Benefits* (pp. 123–154). Kluwer.

Rabinowicz, W. (2014). Safeguards of a disunified mind. *Inquiry*, 57(3), 356–383.

Ramsey, F.P. (1950/1928). Truth and probability. In R.B. Braithwaite (Ed.), *The Foundations of Mathematics and other Logical Essays* (pp. 52–94). Routledge.

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

Schwartz, T. (1972). Rationality and the myth of the maximum. *Nous*, 6, 97-117.

Tenenbaum, S., & Raffman, D. (2012). Vague projects and the puzzle of the self-torturer. *Ethics*, 123(1), 86–112.

Thoma, J. (2021a). Judgementalism about Normative Decision Theory. *Synthese*, 198:6767–6787.

Thoma, J. (2021b). Folk Psychology and the Interpretation of Decision Theory. *Ergo*, 7.

Thoma, J. (2021c). In Defence of Revealed Preference Theory. *Economics and Philosophy* 37(2), 163-187.

Thoma, J. (2025). Preferences: What we can and can't do with them. *Philosophia*, 52, 1269–1278.

Thoma, J. (forthcoming). The Precision Fallacy: On the Futility of Preference Purification. *Studies in History and Philosophy of Science*.

Williams, B. (1979). Internal and external reasons. In R. Harrison (Ed.), *Rational Action* (pp. 101–113). Cambridge University Press.