



Irrationality and Indecision

Jan-Paul Sandmann¹

Received: 8 March 2022 / Accepted: 17 February 2023
© The Author(s), under exclusive licence to Springer Nature B.V. 2023

Abstract

On the standard interpretation, if a person holds cyclical preferences, the person is prone to acting irrationally. I provide a different interpretation, tying cyclical preferences not to irrationality, but to indecision. According to this alternative understanding – coined the indecision interpretation – top cycles in a person’s preferences can be associated with a difficulty in justifying one’s choice. If an agent’s justificatory impasse persists despite attempts to resolve the cycle, the agent can be deemed undecided. The indecision interpretation is compatible with the standard interpretation of such cycles as instances of irrationality. Yet indecision corresponds to a first-person, non-instrumental perspective on the problem, whereas irrationality usually corresponds to a third-person, instrumental perspective. Due to these differences, interpreting cycles in terms of indecision offers a novel conceptual perspective, pertinent both for explanatory purposes and for the aim of normative theorizing.

Keywords Indecision · Instrumental irrationality · Preference cycles · Money pump argument

1 Introduction

Much of philosophical decision theory has concerned itself with explicating a standard of instrumental rationality. On the picture familiar from microeconomics, instrumental rationality requires agents to rank alternatives in a coherent manner by avoiding cycles at the top of their preference ranking. An individual thus is deemed

✉ Jan-Paul Sandmann
jsandmann@fas.harvard.edu

¹ Department of Government, Harvard University, 1737 Cambridge St, Cambridge, MA 02138, USA

rational when she does not cycle back and forth between her top-ranked alternatives, and irrational if she does cycle back and forth between these alternatives.¹

By focusing on the relationship between cycles and rationality, a different interpretation of preference cycles has however received comparatively little attention (yet see Chang, 1998). My aim in this paper is to explore such a hitherto uncommon interpretation of cyclical preferences. According to this alternative understanding, cyclical preferences can indicate that the agent entertains conflicting views about the value of the options she is presented with. In light of such conflicting views, the agent may face a justificatory impasse, making her undecided about her choice. When the agent, moreover, cannot resolve her difficulty in justifying the choice to herself, she faces a genuine problem of indecision.

Put another way, if one cycles back and forth between some options at the top of one's preference ranking, then one may not be able to select an option and justify this choice to oneself. Given one's problem to justify the choice on the basis of one's preferences, one may instead be left undecided. Cyclical preferences, besides capturing someone's potential for instrumental irrationality, may indicate that one is undecided.

In the second section, I spell out more clearly how a person with cyclical preferences may be led into a justificatory impasse. I further illustrate when such an initial impasse results in indecision.

In the third section, I draw several distinctions to related sources of indecision, seeking to clarify the supposed link between cycles and indecision. In particular, I distinguish between someone's being undecided because of a preference cycle and indecision which is connected to the options being *indifferent* from the agent's viewpoint, them being *incomparable*, and *uncertainty* about the right option to choose. Importantly, my account of how preference cycles can be a source of indecision abstracts from facts about the right choice for an agent. Rather, I am concerned with how the agent can justify a choice to herself, without reference to facts about the goodness of options.

This leads me to the fourth section, in which I explore how the interpretation of cycles in terms of indecision is related to the familiar idea of tying cycles to irrationality. I argue that the two interpretations of cyclical preferences are principally compatible, but that they come apart in their underlying explanatory ambition. We tend to think about someone's being undecided primarily from a first-personal and non-instrumental perspective, but are pushed toward evaluating someone's being irrational from a third-personal, instrumental perspective. This shift in explanatory perspective, I argue, marks an important difference between connecting cycles to indecision and tying them to rationality.

In the fifth and final section, I sketch some of the implications the indecision interpretation has for purposes of explaining and evaluating agents' decisions. In particular, the interpretation lends itself to spelling out when an individual exercises reasoned control over herself. Exercising reasoned control can be valuable independently of the outcomes an agent obtains by doing so. The final section thus points to

¹ One may question whether instrumental rationality can reasonably be cashed out in terms of acyclical preference orderings (Thoma, manuscript). But such worries do not cast doubt on the fact that individuals are commonly viewed as rational because they evade contradictions in their preferences.

the kinds of explanatory and normative contributions thinking about cyclical preferences in non-instrumental terms can make.

2 Cycles and indecision

To illustrate my claims, it helps to consider cases in which a decision needs to be made. To take a simple example that I will return to repeatedly, think about a student who is tasked with choosing which subject to major in. Our student, call her Sophie, prefers to study physics over philosophy, philosophy over political science, but political science again over physics. Sophie thus ranks the options physics, philosophy, and political science cyclically: physics \succ philosophy \succ political science \succ physics (...).

A well-known piece of reasoning, coined the *money pump argument*, shows that individuals with cyclical preferences are prone to act against their own interests (Hampton, 1994; Andreou, 2016). Let me illustrate how the argument works in general before returning specifically to Sophie's case. Suppose an individual has a cyclical preference ordering of the form $a \succ b \succ c \succ a$ (...). Starting with c , the individual may be offered to trade c for b by paying a small amount of money. In virtue of her strict preference ordering, the individual agrees to this trade. Subsequently, our individual could be offered b for a . In light of the same argument, the individual again agrees to pay some amount of money while giving up b . She then faces a choice between c and a . By the same logic, she would accept to trade the two options.

As a result of the sequence of trades, the individual ends up with the option she started with but lost a fraction of her wealth. If this sequence continues indefinitely, the individual would end up significantly worse off. The money pump argument hence shows that individuals with cyclical preferences can be drawn to act against their own interest. Agents who choose in this way would exhibit irrational behavior from an instrumental viewpoint.

In the case of Sophie's major choice, however, arguments centered on instrumental irrationality seem beside the point. There will only be one choice Sophie needs to make. As such, Sophie cannot be made to act against her own interest. Instead, Sophie may ignore an underlying preference cycle and pick any of the available majors. She could, for instance, reduce her attention to a binary choice between political science and physics (leaving philosophy out of the picture), and choose to major in political science, given her pairwise preference for this.

Yet even if there is only a single choice Sophie needs to make, choosing despite the underlying cycle in her preferences appears troublesome, independent of the reasons the money pump argument captures. Why exactly is that the case? Why would there be something troublesome with Sophie focusing on any particular pairwise comparison and choosing accordingly when her preferences over all options, taken together, are contradictory? The point is that, by reducing the choice to any particular pairwise comparison, some information Sophie may deem relevant to the decision gets lost.

Let us assume that Sophie, knowing that her major choice will greatly affect the course of her life, wishes to justify this choice to herself. Even though Sophie is able

to rank the majors relative to one another in pairwise comparisons, when she considers the three options *jointly*, she cycles back and forth between them. My claim is that agents like Sophie who seek to justify their choice, but rank the available options cyclically, may be led into a justificatory impasse.

Justificatory impasse: A cyclical preference ordering among the top-ranked options in a given set may lead to an impasse for the person who needs to choose a single option from this set.

A key distinction between an agent's preferences and her choice function (familiar in the rational choice literature)² helps to characterize why such a situation may arise. Let me illustrate this by continuing to draw on Sophie's example. Sophie's set of options includes physics, philosophy, and political science. Let this set be denoted by X , so that $X = \{physics, philosophy, politicalscience\}$. As explained, Sophie's preferences over the options in this set happen to be cyclical as she ranks physics \succ philosophy, philosophy \succ political science, political science \succ physics. Given Sophie's preference ordering, we are interested in the kind of choice function, $C(X)$ – a mapping from every non-empty subset A of X to a non-empty subset of A – which brings out some options as choice-worthy. For an option to be choice-worthy, it must be one that the agent can justify to herself. Seen in this way, the choice function captures the justificatory standard which an agent applies to a given decision.

There are different, more or less demanding, such standards that agents can impose on themselves. One possibility would be for an agent to look for an *optimal* choice, requiring one to select an option which is at least as good as any other option (so that, formally, $\alpha \in C(A) \iff \forall \beta \in A, \alpha \succeq \beta$). Another less demanding possibility could be for the agent to select only those options which are *maximal*, in the sense of being no worse than any other options (so that, formally, $\alpha \in C(A) \iff \nexists \beta \in A, \beta > \alpha$).

The point is that even in the less demanding case – where Sophie strives only for a maximal choice – her cycle made up of strict preferences yields no option that can justifiably be chosen (which formally is represented by an empty choice set). Sophie's choice function, $C(X)$, be it a one defined by optimality or maximality, returns the empty set if she takes into account all options.

Of course, the choice set will not be empty for any of the subsets, including the pairwise comparisons $\{physics, philosophy\}$, $\{physics, politicalscience\}$, $\{philosophy, politicalscience\}$.³ If Sophie's decision were reducible to any of the pairwise comparisons between the majors, then the choice function yields a permissible option to choose. The point, yet, is that the justification Sophie requires for her choice does not allow her to simplify the decision situation. Such an inability on her part leads to an impasse at the level of the entire set X over which the choice function only returns the empty set.

² For instance, see Sen (1971, 1997).

³ Of course, Sophie's choice function is also defined over all other subsets of X : the singleton sets, including each option separately, as well as the empty set.

In the way that I have characterized Sophie, she understands the justificatory demand which applies to her choice before ranking the available options in an order of preference. Her cyclical ordering then gives rise to the impasse conditional on such standard. But there may also be agents whose preference ranking logically precedes the demand by which they justify a choice.

Yet even when an agent takes their preferences to be logically prior to a given justificatory demand, cycles can likewise be connected to a justification problem. A justification problem, connected to cyclical preferences, could also arise when an agent was not initially clear about such a standard. One's preferences may, in fact, help one realize the fitting justificatory demand in the first place. The general point is that an agent's cyclical ranking among her top-preferred alternatives can be associated with a problem in justifying her choice – both when a justificatory demand comes prior to an agent's preferences and when the preferences themselves clarify what the justificatory demand should be.

I have thus illustrated how cyclical preferences can lead agents into an impasse as they struggle to justify the choice to themselves. The concerns such a situation raises are conceptually distinct from the issues connected with the money pump argument. The money pump argument brought forth the possibility that someone with cyclical preferences could be drawn to acts against their own interest. Given a single choice, however, one simply cannot be drawn to act against one's own interest. But even if one only needs to take a single decision, choosing on the basis of cyclical preferences is not free from problems, as one may fail to justify a choice to oneself. There thus is a separate reason why taking a decision in light of underlying cyclical preferences may be problematic.

Thus far I have supposed that a person is confronted with a set of options and demanded that the person choose a single option from this set. I then focused on the case where the person fails to identify an ordering of the alternatives on the basis of which she could justifiably pick a single option. The person either ranks the top-ranked options cyclically or avoids the cycle among these options by disregarding something she deems necessary to justify her choice. In that case, I claimed, the person is (at least provisionally) led into an impasse. Such a problem arises either in light of an existing justificatory standard or because cyclical preferences make such a justificatory standard apparent in the first place.

The point I wish to convey in addition is that the *persistence* of the person's impasse indicates that she is undecided. Depending on whether an agent is able to resolve the cycle, the agent can either merely provisionally, or more solidly, be undecided. An agent is *provisionally undecided* when she can successfully resolve her cycle or change the justification she requires for her choice. There are multiple ways the agent could go about this. One strategy would be to adjust the applicable justificatory standard or reduce the set of options taken into account. Another strategy would be to hold on to the justificatory standard while considering different options or deliberating further about the benefits of the available options. But there will also be cases when none of these strategies resolve the problem to the agent's satisfaction. When the agent deems none of these means sufficient to resolve either her cycle or remains unwilling to amend the justification she needs to choose, then the agent can be viewed as being *solidly undecided*.

To illustrate how provisional undecidedness differs from solid indecision, let me go through the different ways in which Sophie could circumvent her impasse. After thinking for some time about the virtues of the major options, Sophie may wish to get her major choice out of the way and move on. To justify her decision, she may thus no longer strive to select an option which beats all others in pairwise comparison. If Sophie can do this to her own satisfaction, she faces no genuine problem of indecision. And yet, given that she initially struggled to justify a choice on the basis of her cyclical preferences, she revealed herself to have been provisionally undecided.

Of course, we could also think about individuals who have not considered all options, and whose reasoning is incomplete as a result. Perhaps Sophie has failed to ponder majoring in mathematics, which could trump all of physics, philosophy, and political science. If that is the case, then majoring in mathematics appears to be good enough for justification purposes. Again, Sophie here is able to resolve the impasse; this time by directly resolving her cycle, rather than indirectly adjusting her justificatory standard. Yet the conclusion in this case parallels the earlier one: Even though Sophie turns out not to be solidly undecided, she initially was in a position of indecision.

Alternatively assume that after having considered all options, Sophie still ends up with a cycle at the top of her ranking. Even though she has no further options to consider, she may come to conceive of the available options differently. In other words, while the options themselves remain the same, Sophie's framing of these options may change. How Sophie views the alternatives affects whether she continues to rank these cyclically.

Presume, for instance, Sophie realizes that majoring in philosophy allows her to take classes not only in political philosophy, but also in the philosophy of physics. Since Sophie is primarily interested in the fundamental debates underpinning quantum theory, she comes to change her preferences. She now prefers majoring in philosophy given that she can take philosophy of physics classes on quantum theory to majoring in physics. And she still prefers majoring in philosophy to majoring in political science (after all, she can continue to take political philosophy classes). Now, majoring in philosophy, given that Sophie can take the relevant classes, trumps the other options from her point of view.

Sophie thereby resolves her cyclical ordering. She can now pick philosophy and justify choosing this major because of its comparative virtues vis-à-vis the other available majors. Sophie's mental reconfiguration of the options has led her to avoid the cycle. And this has been the case without the options themselves having changed. Even though Sophie was initially undecided, she turned out to find a suitable resolution of her situation. Once again, Sophie did not face genuine indecision.

In contrast to these cases, presume Sophie cannot do any of this. She is unwilling to lower the standard required to justify her choice, remains unable to think of a subject that would match her excitement for physics, philosophy, or political science, and fails to reconceive of these options so that she could avoid her cyclical ordering. Despite deliberating more and more, Sophie cannot find a persuasive way to adjust the way in which she could justify the choice to herself. She also cannot think of a plausible resolution of the cycle among her top-ranked options. Sophie's cyclical ordering persists in the face of continued reflection. In contrast to the foregoing

examples, Sophie strikes me as solidly undecided in light of her preferences in this case. Sophie faces a genuine problem of indecision.

Let me take stock. I have argued that if an agent seeks to justify her choice according to the virtues of all the options she is presented with, then a cyclical ordering among these options may lead the agent into an impasse. The persistence of such an impasse, I reasoned, suggests that the person is undecided. There are different ways to think about how a justification problem, and a person's resulting indecision, could come to be resolved (and so cease to persist). The chooser's evaluative conflict could, for instance, be evaded by reducing the set; by expanding the set with the result of singling out an option preferred to the options caught in the cycle; or by reconceiving of the options such that at least one preference ranking reverses.

There are cases in which an agent can successfully resolve her cycle and corresponding impasse by any of these ways of re-describing the problem to herself. In these cases, the agent was merely provisionally undecided. But there also cases in which, upon further reflection, the agent simply fails to resolve her problem by any of these means. In these cases, the agent is solidly undecided. The agent faces a genuine problem of indecision.

Overall, I have given an account of how to link a person's cyclical ordering of a set of alternatives with her state of being undecided. The precise connection between cycles and indecision can take subtly different shapes, as the distinction between an agent being provisionally and solidly undecided illustrates. The crucial point, however, is that there indeed appears to be a connection between ranking some options cyclically, being led into an impasse because of one's underlying demand to justify the choice, and being undecided about one's choice.

3 Sources of indecision: cycles, indifference, incomparability, and uncertainty

I would like to scrutinize my claim that cyclical preferences can be associated with indecision. Specifically, in this section, I will distinguish between indecision which arises from cyclical preferences and several key notions connected to value incomparability; namely, indifference, incomparability, and uncertainty. Drawing these distinctions clarifies the connection between someone's cyclical preferences and her state of indecision.

First, consider the difference between indecision which is associated with preference cycles and indecision connected to indifference. Just as Sophie who could rank the major options in pairwise comparisons, a person who takes herself to be indifferent between two or more options has a well-defined preference ranking. The difference is simply that the person who is indifferent between some options does not have any reasons to prefer one option to another one in a pairwise comparison.

Whereas a strict preference cycle results in an empty choice set both on accounts of optimality and maximality, an indifference ranking entails a choice set which includes all options, on the same accounts of optimality and maximality. Every option among which the agent is indifferent is optimal since it is weakly better than any alternative option. And any option is maximal since there is no alternative option which is

strictly preferable to it. Thus interpretationally, whereas a cycle – conditional on a given justification standard (in this example, either optimality or maximality) – leads to a situation in which *no* option can justifiably be chosen, an indifference ranking – relative to the same justification standard – entails a situation in which *any* option can justifiably be chosen.⁴

Moreover, indifference can give rise to indecision even if the set of options includes only two alternatives. A cycle, however, can only arise with three or more options. Thus, for cyclical preferences to ground an agent's indecision, the agent's choice set needs to include a larger number of options as compared to when the agent's indecision stems from pairwise indifference.

Second, let me turn to the contrast with *incomparability*. It should be apparent from my discussion that if someone is undecided because of an underlying cycle in her preferences it does not follow that the options are incomparable from the person's perspective. Two options are incomparable from the agent's viewpoint when she cannot rank them. She neither prefers one option to another. Nor can she say that she is indifferent between the options. By contrast, someone who is undecided due to a preference cycle nonetheless is able to rank the options in pairwise comparisons (recall Sophie who could make precisely these comparisons between her major options).

An agent who cannot rank some alternatives could not even begin to think about what an optimal or a maximal choice would require of her. There is nothing which tells her that *no* option can be chosen, unless she has a special further reason to think that. And so, similar to a person with an indifference ranking, an agent who deems some options to be incomparable could pick any option. Yet, unlike the person indifferent between some options, she would need to justify this choice to herself independent of considerations about maximality or optimality.

Ranking some options cyclically, taking oneself to be indifferent, or viewing the options to be incomparable, thus is importantly different. As a result, the connection between these orderings and indecision differs too. And yet, the varying preference orderings are united as they can all be associated with a problem for the agent's justification of her choice; which, if persistent, grounds an agent's indecision. Thus, diverse ways of evaluating some options – indifference, incomparability, or cycles – can all be *sources* of indecision.

To clarify what it means to be a source of indecision, consider how the agent may resolve her indecision. On pain of selecting a single option, the agent could ponder different resolutions of her indecision. When she cycles back and forth between some options (say n of them), it appears best for her not to prioritize any particular option and instead treat all options equally. To do so, she could randomize and pick any particular alternative with uniform probability ($1/n$ in the case of n options).

Interestingly, as Ullmann-Margalit and Morgenbesser (1977) have noted, in certain contexts it may be good enough for the agent to *pick* an option (say, in an order

⁴ Both situations can be grounds for indecision, since a full menu of undifferentiable alternatives may also pose a problem to the agent (to fully explain this point, see below for a distinction between an agent's choice problem and her justification problem). Yet, importantly, the way in which these situations are associated with indecision differs as the agent's relationship between her preferences and the applicable justificatory demand varies.

of convenience) rather than choose (on the basis of a proper justification). In situations in which it would be worse for an agent to not choose at all, and in which the options appear indistinguishable from the agent's viewpoint (even though a more fine-grained evaluation would differ),⁵ the agent seems justified in picking any one option without giving a reason for doing so. More precisely, Ullmann-Margalit and Morgenbesser claim that the agent need not give any *comparative* reason, singling out an option as better relative to another one, to justify her choice.

That there exists a resolution of the agent's indecision, however, does not question that the agent is initially put in a position prone to indecision because of a particular way of ranking the options. Randomization or picking resolutions are principally applicable where the agent is undecided in light of a cyclical ordering, deems some options to be incomparable, or views herself as being indifferent between them.

Crucially, in all of these cases, there are two questions we may ask:

Question I: In light of what is the agent undecided?

Question II: And what, if anything, can the agent do to no longer be undecided?

I have given an answer to **Question I**. That the agent is undecided can be explained by the cycle at the top of her preference ordering *in connection* with the demand for an optimal or maximal choice. But, a cyclical ordering, of course, is not the only explanation for indecision. The agent could face other, however clearly disanalogous, difficulties in choosing when she deems the options to be incomparable or sees herself as being indifferent. My claim simply is that a cycle among the agent's top preferences could be one ground (among multiple possible grounds) for indecision.

With respect to **Question II**, there are two possibilities: either picking or randomizing resolves the agent's initial justificatory impasse or picking or randomizing does not resolve such an impasse. These two options recapture my previous distinction between an agent who is merely provisionally undecided and one who stands as solidly undecided. For the former type of agent, some randomization or picking solution may conceivably be good enough to resolve the justification problem she initially faces. For the latter type of agent, by contrast, no such resolution resolves her indecision.

Many agents in various situations fall within the first category. A given cycle initially makes them undecided, but grounds no genuine indecision over time. But there may also be situations in which agents fall within the second category. As such, an agent may simply not want to pick or randomize in the face of her cyclical ordering. From her viewpoint, neither randomizing nor picking is good enough to resolve her impasse. She may realize that randomizing or picking puts her in a better position than being stuck with no option to choose would. And still, she may be discontent with any such resolution.

The reason she could give for her discontentment is that randomizing or picking means that she must give up on her initial justificatory demand. For some crucial decisions, randomizing or picking may indeed provide no adequate resolution of the

⁵ On this particular point, see Ullmann-Margalit and Morgenbesser (1977: 775).

justification problem in light of which the agent struck us to be undecided. The initial example of Sophie aptly illustrates this.

To further clarify this point, it is crucial to separate the agent's *choice problem* from her *justification problem*. The choice problem refers to an agent's ability to decide, whereas the justification problem refers to an agent's capacity to justify a given choice to herself.

Agents who are solidly undecided face both problems. Such agents cannot decide (i.e., have a choice problem) because they cannot separate their ability to choose from their need to justify the choice themselves and, moreover, are unable to find the right kind of justification (a persistent justification problem). In the vein of Buridan's ass, a solidly undecided agent will proclaim to herself: "Since I cannot justify selecting a given option, I do not choose. Not even randomizing will do the trick for me!"

By contrast, for merely provisionally undecided agents, the choice and the justification problems can come apart. Such agents will at some point agree on a path by which they resolve their choice problem. There are two paths by which their indecision can be overcome.

First, the agent's choice problem could be resolved because she has given in to relaxing her justificatory demand. For instance, Sophie may give up on either maximization or optimization. In this case, while the choice problem remains connected to the justification problem, the choice problem is resolved as the justificatory demand has been relaxed.

Second, the agent's impasse could be resolved because she has altogether decoupled her choice problem from her justification problem. While this case is harder to imagine, we could picture a Sophie, who, in principle, clings on to her initial justificatory demand, say maximization. Yet, after having chosen, Sophie continues to be unhappy with her choice and accepts that there was no way she could have justified this choice to herself. Still, she realizes that a Buridan's ass-type situation is worse an evil than decoupling her choice problem from her justification problem.

The point of detailing these distinctions is to emphasize that, just as an agent could be put in a position of indecision because of indifference or incomparability among the options, she can be drawn to an impasse as a result of a preference cycle. The potential *persistence* of the agent's *choice problem* in light of her cyclical preferences indeed shows this. That is the crucial point I have been aiming to defend all along. Importantly, this point is not questioned by the fact that, in some cases, a resolution of the agent's initial impasse exists.

Third, being undecided is relevantly different from being *uncertain* as to which of the options most closely tracks some "goodness" facts about value. Crucially, I have characterized the problem the agent faces as a purely internal one. The problem arises solely from the agent's viewpoint. The agent is tasked with selecting a single option from a set of alternatives but must justify the choice to herself. This is sharply different from assuming that there are certain facts that determine what the *right* option to choose is. If that were so, then the agent's deliberation would be geared at picking the right option.

In that case, the agent could similarly be undecided and fail to justify her choice to herself. But the reason for which she cannot justify her choice is not wholly internal. By being uncertain about the facts, she faces a justification problem of a different

kind. On the picture I have drawn, the agent's indecision is not tied up with uncertainty in epistemic terms. The agent is not undecided because she either lacks, or has suspended, belief as to what the right option is. Instead, the agent is undecided because she cannot evaluate the alternatives in a way that would resolve the impasse to her satisfaction. Her evaluation may be altogether free from considerations as to what the best option would be according to some external standard.⁶

Let us again pause and take stock. That a person is undecided because of an underlying cyclical preference ordering does not mean that the alternatives are incomparable or that the person is indifferent when she makes pairwise comparisons. These ways of ranking some options can all be associated with justification problems in light of which agents are undecided: they are all *sources* of indecision. And still, the exact character of the impasse an agent faces differs depending on whether she deems herself to be indifferent, takes the options to be incomparable, or ranks them cyclically.

Moreover, an agent who faces a difficulty in justifying her choice in light of a preference cycle is *not necessarily* in a position of genuine indecision. One's difficulty to justify a choice due to a preference cycle is *not* synonymous with, or constitutive of, indecision. My claim, rather, is that if the agent's choice problem persists because of the cycle, then the agent is solidly undecided.

As a result of preference cycles, choosers may find themselves *prone* to indecision. The cycle can, to reiterate important terminology, be a *source* of indecision. But such temptations towards indecisiveness can be overcome, emphasizing the distinction between the choice problem and the justification problem an agent faces.

It is, finally, important to clarify that, based on how I have fleshed out the connection between cycles and indecision, a person's indecision may stem from a purely internal difficulty in justifying the choice. The person would face a different kind of problem when she is unsure as to which option is most valuable according to an external standard of goodness.

Table 1 Sources of indecision

source of indecision	minimum number of options, \geq	type of justification problem
strict preference cycle	3	no option can justifiably be chosen (according to optimality and maximality)
indifference	2	all options can justifiably be chosen (according to optimality and maximality)
incomparability	2	all options can justifiably be chosen (independent of optimality and maximality)
uncertainty	1	unsureness about tracking goodness facts

⁶ To continue a previous comparison (noted in Table 1 above): whereas cycles can be a source of indecision only among sets of three or more options, uncertainty could apply to sets of just a single option (and so is unlike both incomparability or indifference which apply only to sets of two or more options).

4 Indecision and irrationality

This naturally leads one to ask: What is the relationship between the familiar idea that captures irrationality in terms of cyclical preferences – call this the *irrationality interpretation* – and the less familiar connection – call it the *indecision interpretation* – which, as just explained, ties cyclical preferences to a justification problem and ultimately to indecision? My view is that interpreting cyclical preferences in terms of irrationality or indecision is not necessarily contradictory; on the contrary, the two interpretations are principally compatible. It is even possible that the two interpretations complement each other.

To clarify how the irrationality and indecision interpretations can be compatible, consider the two possible cases: either the agent acts on the basis of her cyclical preferences or she does not. If the agent acts upon her cyclical preferences (and does so stably over time), then she is irrational, according to the standard brought out by the money pump argument. Clearly, in that case, the agent is not solidly undecided. It is possible, although not necessary, that the agent was initially undecided in light of her cycle. She could have realized that her preferences are cyclical and thought of the cycle as posing a justificatory challenge to herself, but after some deliberation chose to act on the basis of these preferences.

If the agent, by contrast, does not act based on her cyclical preferences, but continues to hold on to these, then she may be formally irrational. From a purely instrumental viewpoint, however, her irrationality is inconsequential. The agent will not be made to act against her own interest. My aim has been to call attention to the idea that, even in that case, the agent's cyclical preferences are not wholly inconsequential. If such preferences can be associated with a persistent choice problem for the agent, then the agent can be deemed undecided.

Thus, in both cases (where the agent either acts or does not act upon her preferences), interpreting the agent as being irrational and interpreting her to be undecided can principally be compatible. Yet besides such compatibility, the key point I seek to convey is that the two differing interpretations of cyclical preferences have a divergent *explanatory focus*. A difference in the interpretations' explanatory focus entails dissimilar implications for how we evaluate decisions. Thus, whether we interpret contradictory preferences in terms of irrationality or indecision ultimately has noteworthy consequences (as will become apparent in the final section).

The earlier discussion of the money pump argument highlights the different explanatory claims underpinning the two interpretations. Whereas the money pump argument (connected to the irrationality interpretation) focused on sequences of choices, the case of Sophie (linked to the indecision interpretation) takes into account only a single choice. In the former case, we are interested in the consequences that deciding over time has for the person (holding stable her cyclical preferences). In the latter case, we seek an account of how the person can justify the choice to herself, even if she faces only a single choice and so is not prone to be made worse off in the first place.

As pointed out, we may say that Sophie's initial preference ordering makes her prone to an irrational sequence of actions. As such, the absence of cycles continues to be a formal requirement for rational action. From the perspective of *instrumental*

rationality, however, Sophie's proclivity to irrational actions would remain inconsequential if she does not act upon her preferences. But from a *non-instrumental viewpoint*, the cyclical ordering is not inconsequential since such preferences may put Sophie in a position of indecision. Thus, seen from a justificatory angle, disconnected from concerns about instrumental rationality, it is nonetheless significant to pick up on Sophie's cyclical ordering. The indecision interpretation, therefore, helps us to grasp a part of what makes choosing in light of preference cycles troublesome. The justificatory problem Sophie faces could not be adequately captured by the rationality interpretation alone. A divergence in the two interpretations' explanatory focuses can be turned into an explanatory contribution.

But there is a further way to sharpen how the rationality and indecision interpretations differ in their explanatory ambition. To bring out this difference, let me turn to a problem that has frequently been discussed within the familiar setting of connecting cycles with irrationality (see, e.g., Broome, 1993; Bradley, 2016): the problem of the *reframing of options*. The problem runs as follows when applied to the rationality interpretation. If one puts no boundaries on how a person can conceive of the options, then the person turns out to be instrumentally rational virtually by definition. Put another way, suppose we think preference orderings free from contradictions capture a central standard of instrumental rationality, but agents can come to adopt such orderings by reframing the options as they please. If that is the case, then a central standard of instrumental rationality stands as vacuous.

The worry, when applied to the indecision interpretation, runs similarly. Anyone who strikes us as provisionally undecided could, by viewing the alternatives differently, resolve the contradiction and therewith cease to be undecided. Analogous to how agents can turn out to be vacuously rational as a result of picturing the alternatives differently, they can resolve their state of indecision in a seemingly frivolous fashion. It is unclear, then, whether cycling back and forth between some options tracks that someone is undecided.

And yet, whereas the reframing worry may lead us to question a robust connection between preference cycles and irrationality, it does not *as strongly* cast doubt on the connection between cycles and indecision. The reason for this claim is that there is an underlying difference in the explanatory aims of the irrationality and indecision interpretations. Namely, being undecided can be made sense of largely in *first-personal* terms, whereas standards of rationality appear to have a bite only if they are given a clear *third-personal* meaning.

The link between cycles and indecision I have described points to a state that the person is in: if *I* cannot come up with a way of resolving cycling back and forth between majoring in physics, philosophy, or political science (or resolving my corresponding justification problem), then *I* am undecided. In an analogous manner, one could explicate rationality from a first-person view: if *I* cannot come up with a way of resolving cycling back and forth between majoring in physics, philosophy, or political science, then *I* am in a state of irrationality.

But whereas the former first-personal claim about indecision may take us quite far in our explanations, the latter first-personal claim about rationality does not seem to do as much work. Rationality is commonly theorized from a third-personal perspective. We are pushed toward having a third-personal account of rationality because

we need to distinguish between rational and irrational states people are in. We may hence aim for an *evaluative* standard of rationality according to which certain preference orderings count as rational and others as irrational. Similarly, we may seek the standard to provide *action guidance*, prescribing certain (sequences of) actions as reasonable and others as unreasonable.

By showing how agents with cyclical preferences are prone to act against their own interests, the money pump argument captures precisely how such a standard can be applied and reveal someone's irrationality to an outside observer. As such, the standard of rationality is formulated from a point of view that extends beyond the agent. It can ultimately be employed to evaluate the agent's actions and guide her in taking future decisions. *Prima facie*, we are not as strongly pushed toward evaluating someone as being decided or undecided. We commonly also do not as unambiguously aspire to guide someone whether to be undecided or not. Therefore, I think, there is an important difference in the explanatory orientation of a largely first-personal account of indecision and a third-personal standard of irrationality.

This difference accounts for why reframing seems more problematic in the case of connecting cycles with rationality than in the case of connecting cycles with indecision. In the former case, that persons could resolve cycles by reframing the options as they please questions whether the rationality standard can be coherently applied. In the latter case, it is not as clear that we need a similarly unambiguous third-personal standard of someone's being undecided. That agents can resolve their indecision by any reframing of the options that comes to mind hence does not seem as problematic.

The example of Sophie's major choice is a case in point. The fact that Sophie realizes that majoring in philosophy allows her to take not only classes in political philosophy, but also on the philosophy of quantum theory, indeed resolves her indecision. Sophie comes to circumvent her cycle, sorts out her justification problem, and thereby resolves her initial indecision. And that strikes me as plausible after all.

So far in this paper, I have concentrated on a connection between cycles and indecision solely from the agent's perspective. That the agent can internally reframe how she views the options and thereby resolve her indecision does not seem worrisome from a *purely conceptual* angle. As the example of Sophie's major choice illustrates, simply zooming in on the connection between cycles and indecision from the agent's viewpoint underscores that cyclical preferences can be a source of indecision.

Yet, to be clear, there are decisions importantly dissimilar from Sophie's major choice. Certain choices, for instance, have weighty moral implications in ways that Sophie's choice does not. In these cases, we may be interested in how a person *ought* to resolve her indecision. And thus, we are drawn to put some constraints on how they may justify their choices to themselves. A person's justification problem would thereby be no longer purely internal. Instead, it would need to be seen in light of certain external standards as to what "good reasoning" amounts to. In that case, that agents can resolve their indecision by any conceivable reframing of the set of options would be problematic.

In the subsequent final section, I will tie the indecision interpretation to a picture of a chooser capable of reasoned control and thereby capture an evaluative angle from which to view a person's indecision. By drawing a link to an account of reasoned control, it further becomes apparent how interpreting preference cycles in terms

of indecision comes apart from viewing such cycles through the lens of, primarily instrumental, irrationality.

5 Reasoned control

This section thus outlines how the indecision interpretation, which I have so far developed purely as a conceptual tool, can feature in a normative account of reasoned control. To embark on my inquiry into reasoned control, let me begin with a presumption.

Reasoned control presumption: there is a value in reasoned decisions because such decisions allow one to exercise a worthy form of control over oneself; call this type of control *reasoned control*.

What arguments do we have to endorse the presumption? Why exactly would there be a value in exercising reasoned control over oneself? And how is the indecision interpretation related to reasoned control? To start answering these questions, it is instructive to consider some deviations from the exercise of either control, reason, or both.

Harry Frankfurt's (1971) example of a wanton neatly illustrates an agent who lacks the ability to control herself appropriately through reason. A wanton takes decisions simply in line with the most piercing desire that presents itself. Despite their deficiencies, wantons are capable of having clearly defined pairwise preferences. Presented with any two options a and b , a wanton simply selects a over b if and only if a is more in line with the most piercing desire than b . Although a wanton is able to make such pairwise comparisons, she could not strive for a more elaborate justificatory standard.

A wanton's deliberations happen at the level of first-order desires and never reach a higher level from which one could effectively control these desires itself. In principle, since the wanton's first-order desires can conflict with one another, the wanton may be unsure about the course of action she should take. Importantly, however, a wanton's conflict will never present itself as a justification problem *about* her preferences. A wanton will not be in a position to cognize about the source of her indecision in the way Sophie did.⁷

⁷ In my depiction of a wanton, I am going a step further than Frankfurt did. Frankfurt (1971: 11) held that a wanton is capable of developing second-order desires: "The fact that a wanton has no second-order volitions does not mean that each of his first-order desires is translated heedlessly and at once into action. He may have no opportunity to act in accordance with some of his desires. Moreover, the translation of his desires into action may be delayed or precluded either by conflicting desires of the first order or by the intervention of deliberation. For a wanton may possess and employ rational faculties of a high order. Nothing in the concept of a wanton implies that he cannot reason or that he cannot deliberate concerning how to do what he wants to do." Hence, on Frankfurt's picture, the point is not that a wanton is incapable of deliberation, but that he cannot effectively control these desires through a high order decision which is wilful. Frankfurt (1971: 11) continues: "What distinguishes the rational wanton from other rational agents is that he is not concerned with the desirability of his desires themselves. He ignores the question of what his will is to be." Importantly, a wanton thus is separated from a wilful agent (in his example, an unwilling

In the wanton's case, indecision could stem from first-order conflicts, but not from a deeper reflective impasse connected to a higher-order justification problem. A wanton ultimately lacks control precisely because she fails to act on the basis of reason. Reason demands that agents' deliberations go beyond the most minimal consideration of pairwise desire comparisons, that these deliberations enter the second-order level when these desires conflict, and that such conflicts can be arbitrated through deliberating as well.

Another illustrative example of a person unable to exercise reasoned control is a chooser – call her the randomizer – who fully commits herself to go along with whatever a random device, a coin toss or an oracle, decides on her behalf. The randomizer, by being fully reliant on the outcome of the random device, avoids going through any comparisons between the options herself. She does not even perform the plain desire comparisons a wanton is engaged in. Randomizers, in this respect, could strike one as more extreme than wantons.

Unlike a wanton, who may face action-inhibiting conflicts among their desires, a randomizer is fully decisive. At the same time, a randomizer may exercise a kind of control over herself that a wanton would be incapable of. Randomizers expressly decide to give themselves into the oracle. They commit themselves to go along with whatever option the oracle selects on their behalf. A randomizer thus is capable of devising and acting upon second-order desires: the desire to go along with the result the random device elicits. By committing herself so uncompromisingly to a result over which she has no immediate influence, the randomizer is capable of a form of control which evades the wanton.

Still, there clearly seems to be something insufficient about the kind of control the randomizer exercises. Even though the randomizer has a second-order desire (for the random device to choose on her behalf) and is able to commit herself to resulting decisions in this way, her control does not stem from the reason she *herself* exercises. The randomizer circumvents the task of making up her own mind about the comparative value of some options. Similar to a wanton, the randomizer therefore lacks the ability to control herself *through reason*.

One may think that there is something deeply deficient about agents who approach weighty decisions in the way that a wanton or a random chooser does. Such a judgment may stem not only from an instrumental viewpoint. The judgment appears to have an important non-instrumental component as well. To illustrate why, suppose that one is a particularly bad reasoner. As a result, either replicating the wanton's thought process or taking decisions arbitrarily puts oneself in a better position than personally deliberating about the comparative value of some options would. Would it be reasonable to act in such a way? It does not seem so. Even in the case where wantonness or arbitrariness would be to one's advantage, there continues to be something wanting about one's choices, perhaps even about oneself as a person. If this presump-

addict) in the following way: "The unwilling addict identifies himself, however, through the formation of a second-order volition, with one rather than with the other of his conflicting first-order desires." (Frankfurt, 1971: 13). Rather than drawing on language of wilful action and identification, I have simplified the concept of a wanton. The more extreme wanton I depict still has much in common with Frankfurt's prototype, and suffices, as an example, to motivate basic contours of an account of reasoned control.

tion is true, then taking reasoned decisions is valuable not just from an instrumental, but from a non-instrumental perspective as well.

To take a reasoned decision, one must attempt and comparatively evaluate some options. By comparing alternatives with respect to their value, one opens oneself up to the possibility of contradictions in one's reasoning. These contradictions may in turn put oneself in a position prone to indecision. Exercising reason thus is not free of challenges. Yet the point is that we must ultimately confront these challenges. If we continuously avoid them, we run the risk of eventually resembling wantons or randomizers. Since these creatures lack the important agential feature of reasoned control, acting in the way that they do more often than not comes with the possibility that one no longer is capable of an important agential feature.

Exercising reasoned control successfully demands from a person that she devises the appropriate higher-order standards about the reasoning required to justify a given choice. Conditional on an ability to devise such higher-order standards, the kind of indecision which stems from ranking some options cyclically need not unambiguously be viewed as negative. Understanding how one's preferences can lead to a justificatory impasse may allow oneself to more fruitfully exercise reasoned control over time. Seen in this way, instances of indecision can be an important part of becoming a reasoning agent.⁸

Had Sophie not realized that her preferences are cyclical, and that such a cycle complicated the justification for her choice, she may have made a decision that she would ultimately regret. In particular, without awareness of the preference cycle, Sophie may not have sought further information about the philosophy major, and perhaps would never have heard about the philosophy of physics option. Even though her preference cycle initially posed a problem to her, awareness of this problem and an ability to resolve it seems to be an important component of what it is to be a reasoning agent who exercises control over herself.

Certainly, there are different ways Sophie could have gone about her justificatory impasse. In various life situations, it may be more important to levy a less demanding standard on oneself. Sophie's example by itself surely is insufficient to develop a full-blown account of reasoned control. Such a normative conception would have to be spelled out in much greater detail. In an attempt to do so, one would need to pay close attention to the contextual features of agents' choices.

My point here, however, has not been to develop such an account in all its details. Rather, I have sought to sketch some basic features of a normative conception of reasoned control. By doing so, I have illustrated how the indecision interpretation could principally contribute to spelling out such a conception. Since one's ability to exercise reasoned control does not seem reducible to considerations about the outcomes

⁸ Interestingly, Ullmann-Margalit and Morgenbesser (1977: 780) are drawn to a seemingly opposite conclusion: "Children, we say, see differences where we do not see any, or take trifling differences to be relevant – that is, to be sufficient reasons (usually patently ad hoc) for preference. Indeed we generally regard it as a sign of growing up when a child stops "behaving childishly" and is able to take a picking situation proper as just that (rather than fight with his or her little brother over who gets which)." Even though their conclusion diverges from mine, there is a unified point here; namely, that grasping and resolving a justificatory impasse is a central component of developing one's agential capacities.

one brings about through one's actions, the instrumental rationality conception, on its own, would provide insufficient tools to theorize reasoned control.

To hold on to the value of the rationality interpretation in this context, an alternative would be to define rationality in non-instrumental terms. One possibility is to say that ranking some alternatives acyclically is constitutive of having a preference in the first place. On the constitutive view, agents with cyclical preferences would need to be viewed as "preferenceless." Such a perspective, however, seems comparatively unfruitful – both in terms of explaining and evaluating agents' actions.

As Sophie's example illustrates, there are situations when, for good reasons, we prefer some alternatives over others in pairwise comparisons, and yet end up with a cycle when considering all options. In such cases, one may adjust one's justificatory standard, consider only one pairwise comparison, or change one's preferences. But zooming in on any of the initial pairwise comparisons suggests that one had a preference all along. A constitutive formulation of the rationality interpretation would overlook these relevant details.

Framing a person with cyclical preferences as being undecided in light of a justification problem describes an important psychological phenomenon we seem to encounter. And yet, the phenomenon appears not to be adequately capturable by the standard rationality perspective. It can neither be captured by an instrumental formulation of the rationality perspective, according to which such preferences are inconsequential if not acted upon, nor by a constitutive one, according to which an agent fails to have preferences in case she cycles back and forth between some options.

It is important to emphasize that the indecision interpretation ultimately stands as complementary to the familiar view that ties preferences cycles to irrationality. Since the two interpretations come with separate explanatory aims, they need not be in tension with one another. The two interpretations instead have the potential of jointly providing a richer perspective through which to theorize an individual's deliberation process. Seeing the indecision interpretation in this vein illustrates how the rational choice model can serve as a useful foundation to explore a range of explanatory and normative questions.

6 Conclusion

I have argued that a preference cycle can be associated with a problem in justifying a choice. If the justificatory impasse persists, then the agent reveals herself to be undecided. Indecision can arise from multiple sources. Preferences cycles, thus, are only one possible source of indecision. Yet different preferences – cycles, indifference, or incomparability – bear varying relationships toward indecision. Importantly too, indecision which is linked to cyclical preferences is distinct from uncertainty as to which option best tracks some goodness facts.

An interpretation of cyclical preferences in terms of indecision has not been explored systematically in the philosophical literature on rational choice theory. The literature has, rather, focused on the connection between cyclical preferences and instrumental irrationality, as is common in microeconomics. The differing interpretations, which either tie a preference cycle to indecision or to irrationality, may

ultimately complement each other. But the two interpretations also diverge in their explanatory aims. Indecision is a more first-personal phenomenon than rationality. In the case of rationality, one needs to give a third-personal account of the concept to evaluate or guide people's actions. In the case of indecision, the same is true in principle. But, in practice, one may not be as strongly committed to an evaluative or action-guiding standard.

The justificatory perspective encapsulated by the indecision interpretation, moreover, lends itself to formulating an account of reasoned control. Exercising reasoned control, I suggested, does not merely have to do with the outcomes one obtains by choosing in a particular way, but with how one justifies the choice to oneself. Due to its non-instrumental orientation, the indecision interpretation thus provides apt conceptual resources to spell out basic tenants of such a normative conception. An interpretation of preference cycles in terms of indecision ultimately enriches the existing rationality perspective and broadens the scope and applicability of the rational choice model.

Acknowledgements For discussion of related issues and comments on earlier versions of this article, I am grateful to Arthur Applbaum, Selim Berker, Tweedy Flanigan, Dimitrios Halikias, Christian List, Eric Maskin, Eric Nelson, Michael Rosen, Ken Shepsle and Richard Tuck. For written comments and suggestions, I am particularly indebted to Eric Beerbohm, Liya Nahusenay and Amartya Sen. I would also like to thank the journal's two anonymous reviewers as well as audiences at Harvard University and LMU Munich for their helpful comments.

Declarations

Conflict of Interest The author did not receive support from any organization for the submitted work.

References

- Andreou, C. (2016). Cashing out the Money-Pump Argument. *Philosophical Studies*, 173.
- Bradley, R. (2016). *Decision theory with a Human Face*. Cambridge University Press.
- Broome, J. (1993). Can a Humean be moderate? In R. G. Frey, & C. W. Morris (Eds.), *Ethics out of Economics*. Cambridge University Press.
- Chang, R. (1998). Introduction. In R. Chang (Ed.), *Incommensurability, incomparability, and practical reason*. Harvard University Press.
- Frankfurt, H. G. (1971). Freedom of the Will and the Concept of a person. *The Journal of Philosophy*, 68.
- Hampton, J. (1994). The Failure of Expected-Utility Theory as a Theory of Reason. *Economics and Philosophy*, 10.
- Ullmann-Margalit, E., Morgenbesser, S. (1977). Picking and Choosing. *Rationality, Choice, and Morality*, 44.
- Sen, A. K. (1971). Choice Functions and Revealed Preference. *The Review of Economic Studies*, 38.
- Sen, A. K. (1997). Maximization and the Act of Choice. *Econometrica*, 65.
- Thoma, J. (2021). Preference Cycles and the Requirements of Instrumental Rationality. Working paper, available here: <https://johannathoma.files.wordpress.com/2020/04/preference-cycles-april2020.pdf>.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted

manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.