Theory and Decision (2026) 101:591–608
https://doi.org/10.1007/s11238-026-10137-4

Some Mistakes Are Irreducibly Diachronic

Johanna Thoma[1]

Received: 28 October 2024 / Accepted: 11 April 2026 / Published online: 12 June 2026
© The Author(s) 2026

Abstract
The idea that people make mistakes in how they pursue their own best interests, and that we can identify and correct for these mistakes has been central to much recent work in behavioural economics, and the 'nudge' approach to public policy grounded in it. The focus in this literature has been on individual choices that are mistaken. Agreeing with, and building on the criticism that this literature has been too quick to identify individual choices as mistaken, I argue that it has also overlooked a kind of mistake that is potentially more significant: irreducibly diachronic mistakes, which occur when series of choices over time do not serve our interests well, even though no individual choice is mistaken independently of being part of the series. I argue for the claim that people make such mistakes, and reflect on its significance for welfare economics.

1 Introduction

The idea that people make mistakes in how they pursue their own best interests, and that we can identify and correct for these mistakes has been central to much recent work in behavioural economics, and the 'nudge' approach to public policy grounded in it (see Thaler and Sunstein 2008, Beshears et al. 2008, Camerer et al. 2003, Beshears and Kosowsky 2020). The focus in this literature has been on individual choices that are mistaken. Agreeing with, and building on the criticism that this literature has been too quick to identify individual choices as mistaken, I argue that it has also overlooked a kind of mistake that is potentially more significant: irreducibly diachronic mistakes, which occur when series of choices over time do not serve our interests well, even though no individual choice is mistaken independently of being part of the series.

Two ideas have traditionally been central to how economists approach questions of welfare and policy evaluation. First, an anti-paternalist commitment: When evaluating the welfare of individuals, we should be deferential to people's interests as they themselves see them (henceforth their 'subjective' interests). And second, the claim that these interests are typically revealed in people's observed choice behaviour, at

✉ Johanna Thoma
johanna.thoma@uni-bayreuth.de

[1] Universität Bayreuth, Bayreuth 95447, Germany

least insofar as their choices concern only themselves. These claims together ground a methodology whereby decision theoretic models can be used to measure welfare. Specifically, the orthodox approach assumes that people make choices that reveal preferences in line with the axioms of utility theory. Moreover, it assumes that these preferences are context-independent, that is, they do not change when the choice context is changed in ways that we can reasonably assume are not relevant as far as the agent's subjective interests are concerned. This seems sensible if we hold that utility theory coupled with context-independence is both descriptively and normatively adequate. Descriptively, so actual choices can be fitted to these models. And normatively, so these models are a guide to people's subjective interests. This would be so if utility theory coupled with context-independence captures how rational agents pursue their subjective interests, take the best means to their subjective ends – a broadly Humean conception of practical rationality.

In the face of an abundance of evidence that real-world agents frequently violate the axioms of utility theory or choose in context-dependent ways, something has to give. We could give up the idea that utility theory coupled with context-independence is the correct normative theory; We could give up the anti-paternalist ideal; Or we could give up the idea that subjective interest is always revealed in people's choices, while finding alternative ways to defer to it. Much of the literature on welfare in the "behavioural" tradition has opted for the latter approach. The core idea is that when agents violate the standard framework, they make mistakes. But we can identify these mistakes and correct for them. We do this by identifying people's 'true preferences', reflecting their subjective interests and assumed to be in line with the standard framework, and using those as a welfare standard. This is thought to still be anti-paternalist in spirit.

On this 'true preference' or 'preference purification' approach to welfare economics, described by Sugden (2018) as a 'New Consensus' in behavioural welfare economics, the mistakes in question are individual choices: Whenever actual choices are out of line with the 'true' preferences we ascribe to them, the agent is making a mistake. In the following, after introducing some cases of inconsistent or context-dependent behaviour (Section 2), I will first outline a compelling line of criticism of the preference purification approach, according to which it has been too quick to identify individual choices as mistaken, even if we grant the normative correctness of the orthodox framework. And I will suggest that approaches that allow for indeterminacy avoid the problems in question (Section 3). I will then argue that following this line of criticism, we should also accept the claim that people sometimes make mistakes that are irreducibly diachronic (Sections 4-6). Finally, I will explore the implications of this claim for welfare economics (Section 7).

2 Inconsistent and Context-Dependent Preferences

Traditional welfare economics is challenged by cases where real world agents do not act in line with the presuppositions of that framework – for instance because they violate the core axioms of utility theory (and, in the case of risk and uncertainty, expected utility theory) or because their preferences are sensitive to irrelevant context. Take the following three cases:

At the till. A train station kiosk sells healthy and unhealthy snacks. When the unhealthy snacks are put close to the till, most consumers choose those. When the healthy snacks are placed close to the till, many of the same consumers go for those instead.
Casino. Guests at a casino are asked to choose between bet A, offering a 99% chance to win $30 and a 1% chance to lose $7.50 and bet B, offering a 33% chance to win $120 and a 67% chance to lose $15. Many pick bet A. But many of the same people are willing to pay more for bet B than for bet A when asked to bid for the two bets individually.
Shoe shopping. You need new shoes. All of the available shoes in the shop you are in have advantages and disadvantages. You have narrowed it down to three pairs. When comparing Pair 1 and Pair 2, the most salient difference is that Pair 2 is more comfortable, and so you strictly prefer Pair 2. When comparing Pair 2 and Pair 3, the most salient difference is that Pair 3 is more weather-proof, so you strictly prefer Pair 3. And when comparing Pair 3 and Pair 1, the most salient difference is that Pair 1 is more fashionable, and so you strictly prefer Pair 1.

At the till is a classic case of context-dependence, variations of which are discussed at length in the literature on nudges (see Thaler and Sunstein 2008). Placement of snacks is presumably not something that is in itself relevant for whether a healthy or an unhealthy snack best serves people's subjective interests. Nevertheless, people's choices reliably depend on this context. *Casino*, based on a study by Lichtenstein and Slovic (1971), involves risky choice and, under some plausible assumptions, also features dependence on context, in this case regarding whether agents directly compare two bets, or whether they are being asked to bid on them individually. If people strictly prefer more money to less, and if transitivity holds, then we should be able to infer from higher willingness to pay for one bet rather than another that it is strictly preferred. Yet, the bidding context and the binary choice context induce different preferences, despite this being a difference that most consumers can be assumed to not intrinsically care about. Finally, *shoe shopping* features cyclical preference, in violation of the acyclicity implied by the orthodox framework.

As suggested above, confronted with such cases, something has to give in welfare economics. Either we have to accommodate inconsistency and context-dependence as rationally permissible, or we have to give up on the anti-paternalist ideal, or we need to abandon deference to at least some of the choices we observe. The dominant framework in what has come to be known as behavioural welfare economics takes the latter route. The thought here is that when people exhibit context-dependence or inconsistency, they are making mistakes in the pursuit of their own subjective interests. Rational pursuit of their interests requires consistency and context-independence. And so insofar as their choices deviate from the standard framework, people are irrational. This 'true preference' research programme then aims to correct for such mistakes, 'purify' preferences to arrive at a unique and consistent preference relation reflecting people's 'true' interests, and potentially intervene in order to help people act in line with these true interests (see Beshears et al. 2008 and Kosowsky 2020 for an expression, and Sugden 2018 for a critical exposition of this paradigm).

At the till is a popular case in point used to argue for the merits of this kind of approach. The context-dependence in fact creates an opportunity for policy-makers to incentivise healthy choices through small interventions in choice context, or what is usually called the 'choice architecture'. See Kroese et al. (2016) on the effectiveness of such a 'nudge' in this particular scenario. Moreover, it is argued, such an intervention still at least partly lives up to the anti-paternalist ambitions of welfare economics: For one, interventions on choice architecture seem minimally invasive – they neither constrain choice nor alter pay-offs. Moreover, it is usually claimed by proponents of such interventions that we have good grounds for assuming that what people really want, what they have a 'true preference' for is the healthy snack.

There are two justifications frequently offered for the idea that the healthy choice is what is in people's true interests. One appeals to the thought that most people actually value their health, that this is a near universal good (see, for instance Hausman 2016). The strength of this consideration when considering an individual nudge concerning a station kiosk is not entirely clear. Occasional indulgences are entirely consistent with leading a healthy life. Perhaps, on reflection, most people actually take train journeys to be the perfect setting for the occasional indulgence, and if they consider any of the potential context-induced snack choices to be mistaken, it is in fact being nudged to eat a healthy snack when what they 'really' wanted was to indulge. A more specific justification comes from studies that ask people how they feel about a particular behavioural intervention. Kroese et al. (2016) did this in the case of the train station kiosk intervention and found an overwhelmingly positive response. This kind of reflective endorsement offers a strong vindication of the nudge, albeit one that places a high empirical burden on any proposed behavioural intervention.

In the case of *casino*, neither of the context-induced preferences in favour of bet A or bet B seems to have any intuitive pull as the one accurately representing people's 'true' preference. Neither is there a plausible case that one and not the other choice context induces a cognitive bias distorting choices in identifiable ways. We face a similar difficulty when looking at *shoe shopping*. If your 'true' preferences are supposed to be acyclical, then at least one of your strictly cyclical actual preferences must diverge from what those true preferences are. But nothing in the description of the case, at least, singles one of them out as the one most likely to be a mistake. The most plausible explanation for your choices involves the different degrees of salience of different kinds of reasons applying to your series of shoe choices (as modelled, for instance, in Dietrich and List (2016). Consequently, one major line of criticism of the dominant 'true preference' approach is that it must involve ascription of a mistake in many cases where we have no good grounds for identifying any particular choice as mistaken (see, e.g. Infante et al. 2016 and Rizzo and Whitman 2020). The next section outlines a thesis according to which this is often not only an empirical, but an in principle problem – there is sometimes no unique true preference to be found.

3 Non-uniqueness

One highly plausible thesis, which I have argued for in greater detail elsewhere Thoma (2026), is that even if the standard framework is normatively correct and

people ought to have and act in accordance with consistent and context-independent preferences to count as rational, there may sometimes be multiple different consistent and context-independent preference relations that an agent may permissibly adopt on the basis of her subjective interests. Call this the non-uniqueness thesis: There may be more than one consistent preference relation that captures how your subjective interests may best be served. In the following, I will indeed grant that the standard framework is normatively correct: People rationally ought to have consistent and context-independent preferences. But I will argue that non-uniqueness often holds, and helps make sense of at least some cases like the ones I introduced in the previous section.

The non-uniqueness thesis is made plausible by two kinds of consideration. The first is that, on a purely formal level, the requirements of utility theory, as well as the requirement of context-independence are *structural* requirements (see also Beck 2023 on this point). They rule out certain kinds of combinations of preferences as irrational. But typically, there are multiple ways of changing your preferences to arrive at a permissible combination. To avoid context-dependence in *at the till* or *casino*, you can either always choose the healthy snack or bet A respectively, or always choose the unhealthy snack and lottery B. In *shoe shopping*, to ensure acyclicity, you must abandon one of your strict preferences – but the axioms of utility theory do not dictate which.

The second kind of consideration is more substantive, and notes that there is also sometimes nothing in the content of people's existing subjective attitudes that determines one uniquely permissible way to resolve these kinds of inconsistency – nothing that settles, for instance, whether bet A should be preferred to bet B or vice versa across the contexts of interest, or which of the strict shoe preferences should be abandoned. Perhaps our subjective attitudes – our desires, wants, values, tastes and commitments – are simply not precise enough to always settle such questions.

In Thoma (2026), I explore three kinds of reason why this might be. For one, our desires for things like health, indulgence or style might be vague and thus allow for borderline cases. This sometimes makes it unclear which of two options should be preferred in a way that indifference cannot consistently capture either. Second, there may be indeterminacy in how different considerations should be weighed off against each other in multi-dimensional decision problems, such as *shoe shopping*. It has been suggested that in such situations, we often aggregate on the spot (see Lichtenstein and Slovic 2006, Bernheim 2016), and that we may well end up with different results on different occasions in ways that are sensitive to the most salient differences, as it was the case in *shoe shopping*. But each occasion may represent a permissible way of aggregating the agent's various concerns. And finally, in situations under risk, agents may simply lack a consistent risk attitude of the type that expected utility theory can accommodate. In *casino*, for instance, bet A is the more risk averse choice, and so there is a more and less risk averse way of resolving the context-dependence – but perhaps there is no psychological fact about the agent that settles which one is the more authentic choice for her.[1]

[1] I am assuming here that there is no single right way to pursue a given set of desirable outcomes in the context of risk, but that an agent's attitudes to risk at least partly settle how she best serves her interests

To be clear: The cases I introduced in the last section need not necessarily exhibit non-uniqueness – I only argue that they plausibly could under some realistic ways of filling in the psychological details.[2] What is important for my argument is the plausible *possibility* of non-uniqueness. The possibility of non-uniqueness challenges the 'true preference' approach in behavioural welfare economics insofar as it aims to identify unique true preferences for all agents as a welfare criterion, as I argue in greater detail in Thoma (2026). In the face of non-uniqueness, there are no such unique true preferences to appeal to. In what follows, I will argue that non-uniqueness also leads to the possibility of purely diachronic mistakes.

Before turning to that argument, note that non-uniqueness appears to support switching to an alternative framework that allows for indeterminacy for the purposes of welfare economics. If – as we granted – rationality demands individuals to adopt consistent and context-independent preferences, it demands that they determinately settle questions like whether bet A is preferred to bet B or vice versa, even if only for practical purposes. But when agents have not actually done so, and nothing in their existing attitudes determines how they should do so, then the welfare economist cannot legitimately ascribe determinate preferences to defer to in welfare measurement. It may remain indeterminate, for instance, which of bet A or bet B serves an inconsistent agent's interests best, or how she should order the shoes on offer in *shoe shopping*. Importantly, this indeterminacy in the welfare criterion is justified even though we granted that the inconsistency or context-dependence in preference is irrational: Agents should rationally settle their inconsistencies; But if they haven't done so, the welfare economist has no basis on which to settle the inconsistencies for them.

One prominent framework in the literature that allows for this kind of indeterminacy, and which Bernheim (2016) has recently motivated on similar grounds is Bernheim and Rangel (2009) choice-theoretic framework. This framework first distinguishes between choices that merit deference and those that do not. Choices that do not merit deference stem from straightforwardly identifiable mistakes, such as false beliefs about the nature of the options available, so-called 'characterisation failure'. Note that this step also involves the notion of individual choices as mistakes, but places a high empirical burden on the identification of any individual choice as mistaken. The framework then turns to welfare analysis on the basis of only those choices that merit deference. But this set may still contain inconsistencies. These inconsistencies are accommodated by allowing for indeterminacy in welfare. Determinate welfare rankings are only ascribed when choice of one option over another is unambiguous, in the sense that an agent would never choose the latter when the former is also available. Context-dependent choice that is not due to a straightforwardly identifiable mistake such as characterisation failure is typically accommodated as indeterminacy in this framework. In the context of non-uniqueness, I take this to be the approach that is more faithful to people's subjective interests than the 'true preference' approach.

under risk. The potential source of non-uniqueness I have in mind here is then the absence of the relevant attitude, that is, an attitude to risk that would result in a consistent set of preferences.

[2] Neither am I suggesting that vagueness, indeterminacy in aggregation and absence of consistent risk attitude are always the best explanation of the inconsistency and context-dependence in these examples respectively. The inconsistency in *casino*, for instance, could also be due to imprecision in how the agent values money.

What I want to argue in what follows is that non-uniqueness creates the possibility of mistakes that are irreducibly diachronic: Agents end up unambiguously worse off by their own lights, making their series of choices mistaken, while at the same time still, no individual choice is mistaken in any sense other than that it is part of a mistaken series. Though the possibility of this kind of mistake has important consequences for welfare economics, it has not been appreciated as such in the literature. While the kinds of examples I will draw on are familiar from the dynamic choice literature, their potential interpretation in terms of irreducibly diachronic mistakes that spell trouble for welfare analysis is novel.

4 Dynamic Trouble

It is well-known that agents with inconsistent or context-dependent preferences can get into dynamic trouble. Consider the following two dynamic choice problems, which are dynamic versions of *casino* and *shoe shopping*, respectively.

> **Dynamic casino.** You enter the casino with $30 in your pocket. With that money, you can buy bet B (offering a 33% chance to win $120 and a 67% chance to lose $15). You think that is very good value for money (and that at this rate the casino will go out of business). You then consider the opportunity to swap bet B for bet A (offering a 99% chance to win $30 and a 1% chance to lose $7.50). In comparison, that seems like the safer ticket to you, so you go ahead. The casino then offers to buy bet A back from you for $29. A no-brainer, you think. You leave with $1 less in your pocket than you came in.
> **Dynamic shoe shopping.** After browsing and narrowing things down to the three pairs of shoes, all of which cost $80, the shop assistant asks whether you would rather have Pair 1 or Pair 2, suggesting Pair 3 is on hold for somebody else. Between these, you strictly prefer 2, and so you head to the till. Just as you handed over the money for Pair 2, the shop assistant comes back and says Pair 3 is available after all (you had expected this all along). Between 2 and 3, you prefer 3, and so go and exchange them. As you turn to head out, the shop assistant reminds you that you could still have Pair 1, but it is $81 now, given what a tedious customer you have been. Seeing you strictly prefer Pair 1 to Pair 3, you go for that.

The decision trees in Figures 1 and 2 illustrate the two choice problems respectively, where the square nodes are decision nodes at which agents can decide to go 'up' or 'down' at three subsequent points in time t_1, t_2, and t_3. Let us assume that the sequence of choices they will be offered is perfectly transparent to the agents, so no new information is gained in the course of the dynamic choice problem. In both cases, our agents go 'down' three times. And in both cases, they end up with less money than they started out with, but outcomes that are otherwise just like ones they could have had without losing that money. Assuming that it serves their interests to have more money rather than less, this means they end up unambiguously worse off by their own lights as a result of their series of choices. It seems like something has gone wrong

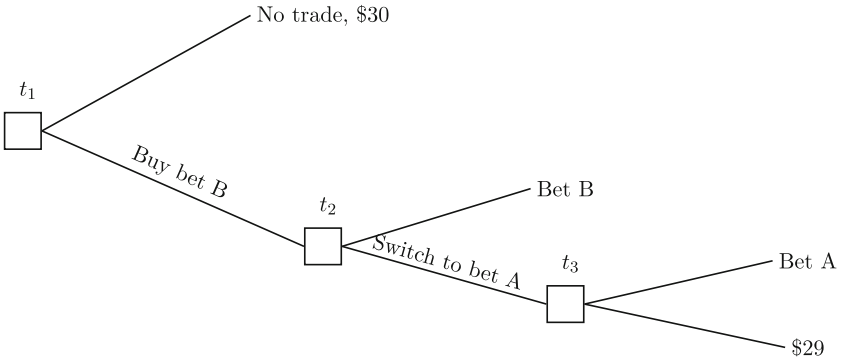


Fig. 1 Dynamic casino.

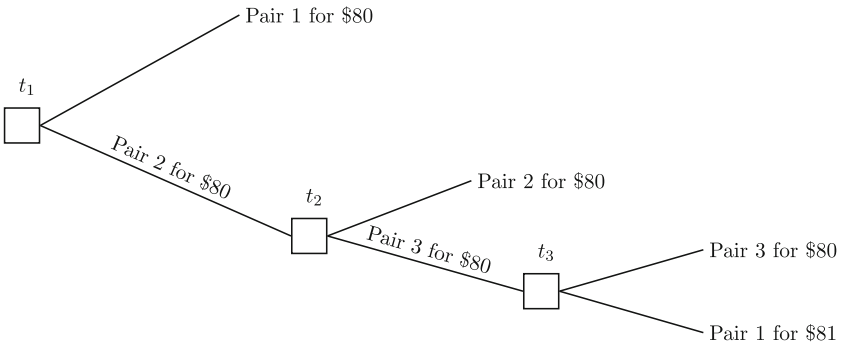


Fig. 2 Dynamic shoe shopping.

for them, and it can't be explained by not having had all the relevant information from the start. This much is granted by the dynamic choice literature. What I want to argue is that when, and *only* when, such Money Pump as well as Dutch Book scenarios also involve non-uniqueness, they are instances of irreducibly diachronic mistakes.

As I presented the examples here, there is one clear rational failing the agents can be accused of: They are myopic, in that they fail to anticipate their own future choices. If they did, according to at least one prominent dynamic choice rule, discussed in more detail below, they would in fact go 'up' at the second to last node and avoid sure loss. It can be shown that, in similar but slightly more complicated scenarios, even agents who act with this kind of foresight can end up making a sure loss (see Rabinowicz 2000). For simplicity, I will stick with my examples featuring myopia. I will argue below in Section 6 that my agents' myopia does not undermine the argument for irreducibly diachronic mistakes.

Discussion of these kinds of examples often focuses on the question of whether the fact that something has gone wrong for these agents shows that their inconsistent preferences are irrational or not. We have actually already granted that the standard framework is normatively correct – that is, we granted that the agents in these examples are irrational by having preferences that are context-dependent or violate the axioms

of (expected) utility theory. My focus is on the implications of dynamic trouble like this for welfare analysis once we acknowledge the possibility of non-uniqueness.

Proponents of the true preference account hold, like us, that the agents in these examples are irrational in virtue of their preferences. Since, according to the true preference theorist, an agent's true preferences are consistent and context-independent, at least one of the choices made by our inconsistent agents facing dynamic trouble must be mistaken in a straightforward sense: it diverges from the agent's true preferences. That choice is mistaken quite independently of what dynamic sequence it is a part of. It is mistaken in virtue of being out of line with true preference, and would be so even if it were faced in isolation. We thus don't even need to point to the tragic outcomes of sure losses to convince the true preference theorist that a mistake has happened. If uniqueness holds, Dutch Book and Money Pump scenarios are not instances of irreducibly diachronic mistakes. We can pinpoint precisely the mistaken choices: They are all those that diverge from the unique and consistent preference relation the agent ought to exhibit.

But once we assume non-uniqueness, this analysis is not open to us anymore. Let us suppose that our cases of dynamic trouble are also cases of non-uniqueness against a backdrop of stable underlying values. That is, nothing changes about how the agents evaluate chances of money, stylishness, comfort and weather-proofness respectively as they move through the dynamic choice problem. But there are several permissible ways of expressing these stable values as a consistent preference relation, due to the different sources of non-uniqueness discussed above. There are then several permissible ways for the agents to resolve the context-dependence or inconsistency of their actual preferences. In particular, let us assume that in *casino*, the agent could either prefer A to B and also assign it higher monetary value when considered individually, or vice versa. In *shoe shopping*, let us assume any transitive ordering of the three pairs would be permissible. Now we can no longer say that any of the choices made by our dynamically troubled agents are mistaken in virtue of diverging from true preference, or that they would be mistaken even if they were to occur outside of the dynamic scenario. If that is not what is going on, then what is the problem?

5 Irreducibly Diachronic Mistakes

In the following, I want to argue that in our two dynamic scenarios above and cases like it involving non-uniqueness, agents make a mistake that is irreducibly diachronic. What do I mean by this? A mistake is a choice, action or activity that you should not have performed, in light of the values and information you had at the time of action/each stage of the activity. Within the broadly Humean framework, the kinds of mistakes we are specifically interested in are choices that do not serve your subjective interests, or your ends. Sometimes, individual choices don't serve your interests well. Perhaps choosing an unhealthy snack when it is close to the till is such a case: Your health is more important to you, but you ignored it against your better judgement. Doing the same thing repeatedly will also not serve your interests well. Then your series of choices is also mistaken. More generally, there is no difficulty in thinking of extended activities as mistaken – they are also things that you do, just over time

rather than at one particular time, and that we may want to say you should not have done. But in this case, we can potentially fully explain what is wrong with the series of choices in terms of the individual mistake you make on each occasion.

By an irreducibly diachronic mistake, I mean a case where a series of choices is mistaken, in the sense of not serving your interests well, and in a way that is not fully explained by the mistakes, if any, involved in the individual choices an agent is making. The clearest kind of case of this would be one where the individual choices are not mistakes at all except in the sense that they are part of a mistaken series of choices. No choice is a mistake considered in isolation, yet the series of choices is mistaken. I think that, in the way we have spelled them out, our problematic dynamic choice scenarios involve irreducibly diachronic mistakes of this kind, and in what follows I will be referring only to such 'pure' cases.

Before making this case, a clarification is in order regarding individual choices considered "in isolation". You might say that all of our choices are only ever mistaken or not mistaken against the background of all the other choices we have made and expect to make in the future. After all, past and future choices to a large extent determine what kinds of options are open to us at any one moment. And so, you might say, it makes no sense to speak of individual choices being mistakes in isolation anyway. But I think the distinction between choices that are mistakes considered individually, and choices that are only mistakes because they are part of a mistaken series can be made sense of. In particular, it can be made sense of when we individuate choices in terms of all of the options that are available to the agent at the time of choice, which are themselves described in terms of all the consequences and intrinsic attributes that the agent cares about.

Past and future choices often make a relevant difference to what options are available to us at a given moment. But this in itself does not make just any mistake irreducibly diachronic in the sense I have in mind. Take, for instance, a situation where you are hungry and want to eat, know that you previously bought some bread, yet choose not to eat. This is clearly a mistake. At the same time, it is true that if you choose not to eat in a situation where there is actually no food in the house and it would be very costly for you to get any, you would not be making a mistake. So is this a purely diachronic mistake? After all, choosing not to eat seems to be only a mistake against the background of a series of choices that involved you buying bread earlier. No, because choosing not to eat is really a different choice when there is a very low effort food option available compared to when it is not.

For a true irreducibly diachronic mistake in the sense I have in mind, a choice with all the same options and consequences the agent cares about (which could of course be partly determined by previous or expected future choices) would not be mistaken if it were not for the series of choices that it is part of. I think that, under some plausible additional assumptions about our agent, this is in fact the way we should think about the problematic dynamic choice scenarios we looked at in the previous section. Suppose non-uniqueness holds in the way we posited, and the agent cares only about chances of money. Then paying \$30 for bet B, choosing bet A over bet B, and taking \$29 for bet A are all not mistakes considered in isolation. Similarly, picking Pair 2 over Pair 1, Pair 3 over Pair 2, and Pair 1 - \$1 over Pair 3 are also all not mistakes considered in isolation if we captured everything the agent cares about in the way we described

the options. If these very same options are the ones open to agents in the dynamic choice scenarios (that is, if the dynamic context does not change these options in a way that the agent cares about), and nevertheless there is a mistake, that mistake must be irreducibly diachronic.

My argument in favour of irreducibly diachronic mistakes is based simply on two claims. (1) The agents in our dynamic scenarios are making a mistake, since they end up unambiguously worse off than they could have been by their own lights, even though their values remained constant and they gained no new information. And (2), it is often not psychologically unrealistic to assume that non-uniqueness holds *and* the dynamic context does not change the nature of the options on offer in a way that the agent cares about. When there is non-uniqueness and the dynamic context doesn't change the nature of the options, no individual choice in the series would be mistaken in isolation. So the mistake that claim (1) grants there is must be irreducibly diachronic. To support the latter part of (2), it does seem psychologically plausible that an agent exhibiting the choice behaviour in *casino* would continue to care only about money and her chances of winning when we go to the dynamic case. Similarly, it is plausible that an agent with cyclical preferences in *shoe shopping* may continue to only care about comfort, weather-proofness, style and money when moving to the dynamic context.

A final clarification about irreducibly diachronic mistakes: Note that in our examples, it is the last choice that creates the problematic outcome: It brings it about that the agent makes a sure loss. Doesn't that make that choice a mistake in isolation? I do not think so. The final choice is the choice that turns the series of choices into a mistake. But this is compatible with saying that the individual choice is not a mistake considered in isolation. The only sense in which it is mistaken is that it is part of a series of mistaken choices. And in that, it is no different from the previous choices in the series. The problematic outcome of making a sure loss could have been avoided by changing any one of them.

6 Potential Objections

My claim that there are irreducibly diachronic mistakes can be resisted in broadly one of two ways. One is to hold that there simply is no mistake at all in the dynamic scenarios as I outlined them. And the other is to argue that we can explain what goes wrong for the agent in terms of individually mistaken choices after all. I will consider these in turn and find them both to be unconvincing. A first potential ground for thinking that no mistake was in fact made, despite appearances, would be to appeal to what is revealed not only by the agent's individual choices, but also by the sum of her choices. In *dynamic casino* the sum of her choices results in her walking out of the casino having lost \$1, and in *dynamic shoe shopping*, the sum of her choices results in her paying \$81 for a pair of shoes she could have had for \$80. If this is what she chose (albeit through a series of choices), perhaps we should just defer to that, and accept that losing this money is really in her subjective interest – it's what she really wanted, and we can't argue with tastes.

I think this analysis is unconvincing, for two reasons. First, it is highly likely that any agent in these scenarios will in fact, with hindsight, regret her series of choices. And there is a simple explanation for this: She loses something most people care to have more of – money – without making any gains regarding anything else she cares about (e.g. quality of shoe). And secondly, there is another set of preferences that could be revealed in an alternative synchronic choice experiment: In the case of *dynamic shoe shopping*, one where the agent gets to choose between pairs 1-3 for \$80, and Pair 1 for \$81 outright. While she may end up picking any of pairs 1-3 out of that set, for the reasons just given, she would never choose Pair 1 for \$81, given Pair 1 for \$80 is also available. And this revealed preference seems to be the better guide to the agent's actual subjective interests and thus more worthy of deference by the welfare economist.

Another ground for thinking that no mistake was in fact made may appeal directly to our arguments regarding non-uniqueness. If non-uniqueness means that none of the choices in *dynamic casino* and *dynamic shoe shopping* are mistakes when considered in isolation, shouldn't we just conclude that no mistake was in fact made? However, for one, this is intuitively unsatisfactory, since sure losses are disconcerting. And moreover, this response is only compelling if we think that there simply is no such thing as an irreducibly diachronic mistake. There is in fact a view in the literature according to which all principles of rationality apply only to persons at a particular moment in time, so called time slices of a person, rather than over time. Advocates of 'time-slice rationality', such as Hedden (2015), can say that agents with inconsistent preferences are irrational at a point in time in virtue of having these irrational preferences at that point in time. They can also say that a choice at a particular point in time is a rational mistake for not serving the agent's interests at that time. But they cannot say that agents can make irreducibly diachronic mistakes.

There is no space here to engage in detail with the arguments in favour of time-slice rationality (see, however, Tenenbaum 2020 and Andreou 2022 for the alternative view that principles of instrumental rationality also apply to temporally extended actions or series of choices). But I would like to raise two considerations that I think support the case for diachronic mistakes at least in the specific context of non-uniqueness I have in mind. The first is that time-slice rationality can't offer a satisfactory alternative explanation of what goes wrong in the kinds of example we considered. As I just mentioned, its proponents *can* say, in the case of *dynamic shoe shopping*, that the agent is irrational at every point in time, because she has cyclical preferences at every point in time. You might think that this already gets us the result that something has gone wrong. However, this synchronic irrationality can't be the whole story.

Suppose that we point out the preference cycle to the agent and she makes an effort to make her preferences acyclical at every point in time. But she adopts different acyclical preference relations at different points in time, shuffling between permissible consistent preference relations. Acting in accordance with her preferences at every point in time, she makes exactly the same decisions as in the original description of the case. She still ends up making a sure loss, that still seems problematic, but this time we cannot explain this in terms of irrational preference at a time. We have de facto turned a problem of inconsistency into a problem of context-dependence across time.

And time-slice rationality can't condemn violations of context-independence across time.[3]

The second consideration against time-slice rationality is that the kind of diachronic rationality principle that would rationalise the judgement that there has been an irreducibly diachronic mistake in our cases is actually very weak. The upshot of arguments for time-slice rationality is often that we can think of time slices of a person in analogy to how we think of different people interacting with each other. Since, for instance, Pareto inefficient outcomes could be the result of the strategic interaction of different perfectly rational agents in a Prisoner's Dilemma-type situation, we should also not generally be concerned about "self-"defeating outcomes in dynamic choice scenarios for a succession of time slices of the same person. However, even if we accept the analogy, note that our cases lack two features that tends to make Pareto inefficient outcomes likely and unsurprising in interpersonal contexts, namely conflicts of interest or asymmetric information. The fundamental concerns of the agents in our examples do not change during the course of the dynamic choice problem. The agents just aggregate them in inconsistent or context-dependent ways. Neither do they gain new information. The analogous case for groups of people would be one where there is a common interest shared by all members of a group, and they all share the same information. We claimed that a mistake is made when an agent who does not change her interests over time and learns nothing new ends up with an unambiguously worse outcome than she could have had. The interpersonal analogy to that is saying that a mistake is made when a group all of whose members share the same interests and information ends up making a decision the outcome of which is judged strictly worse than another by everybody. A principle condemning such mistakes seems like a plausible minimal principle of *group* rationality.[4] By the analogy granted, we should thus accept the analogous principle for time slices of the same person.

The other general strategy for avoiding the conclusion that there are irreducibly diachronic mistakes is to argue that we can after all identify mistakes at specific points in time in our examples. Apart from falling back to the idea of unique "true preferences" the strongest case for this I can think of is to suggest that there are specific strategies that the agents should have used in our examples in order to avoid sure loss, and insofar as they did not use those, we can identify where they went wrong. Remember that as we described them above, the agents in our examples behaved in a myopic way – considering only their preferences between the options immediately before them, without foresight as to the future choices they might make. There is general agreement that myopia is irrational, even if there is no agreement on what rationality does demand in dynamic choice problems.

While I can't go into a full discussion of different dynamic choice rules, let me discuss the two most prominent, before offering a general argument why appeal to dynamic choice principles does not undermine the case for irreducibly diachronic

[3] At best, it can find alternative explanations of why we find such examples troubling. See Doody (2019).

[4] Granted, it is not a trivial principle, given we are dealing with shared interests that can be expressed in terms of different, conflicting preference relations. Coordination by the different agents is thus required in order to abide by the principle. But given there is no conflict at the level of the values to be pursued, I take it to be compelling that when there are no informational or practical barriers to coordination, group rationality demands it.

mistakes. First, take resoluteness, as discussed and defended in a restricted way by McClennen (1990) (see also DesRoches 2020 who defends a form of resoluteness in the context of identifying correction-worthy mistakes in behavioural welfare economics). Resoluteness requires agents to consider different potential paths they might take through a decision problem, then commit to one, and in fact stick to that plan throughout the dynamic choice problem. This would indeed allow the agents in our examples to avoid the problematic outcome, as they would surely not make a plan that leads to sure loss. However, if we are convinced that this is what those agents should have done (instead of being myopic), this does not show that the mistake made by the actual agents is not irreducibly diachronic. If non-uniqueness holds, then there are in fact several permissible plans that they could have made and stuck to if they were resolute. For instance, in both cases, they could have gone 'up' at t_1, t_2 or t_3. In the absence of an actual plan, we can't identify which of the actual choices was a mistake.

The other prominent dynamic choice rule I want to comment on is sophistication (again discussed in detail in McClennen 1990). This rule involves agents always treating 'continuation trees' in dynamic choice problems – those they face from a particular node onwards – as if they were new trees, and correctly anticipating that they will do so in future. They then solve dynamic choice problems in a process of backward induction. They consider what they would do at the last node, choose at the second-to-last node taking that prediction into account, and so on. In our examples, anticipating that they would act in accordance with their preferences at t_3, sophisticated agents would assume they will go 'down' there. This turns the choice at t_2 into, effectively, a choice between bet B and \$29 and a choice between Pair 2 for \$80 and Pair 1 for \$81 respectively. The agents would choose to go 'up' then, 'down' at the first node, and effectively avoid making a sure loss. Note that this choice rule is in fact compatible with time-slice rationality. And it does seem to identify a unique choice as the mistake – that of trading bet B for bet A and Pair 3 for Pair 2 respectively. The problem with this, however, is that, as Rabinowicz (2000) has shown, there are alternative forms of this type of dynamic choice scenario where sophistication does not help an agent avoid making a sure loss. My argument can be put in terms of these slightly more complicated scenarios instead.

There is a more general problem for any attempt to appeal to dynamic choice rules in order to identify individually mistaken choices: Given our presumption that inconsistent preferences are irrational, and that non-uniqueness holds, there are always additional alternatives open to our agents to avoid sure loss. Namely, they can adopt consistent and context-independent preferences instead, in which case even a myopic approach would avoid sure loss. Given non-uniqueness, there are several such preference relations that they could permissibly adopt. In the absence of them actually having done so, we again cannot identify any individual choice as mistaken. I thus conclude that sometimes, agents make mistakes that are irreducibly diachronic. What are the implications of this for welfare economics?

7 Implications and Conclusion

Behavioural welfare economics aims to correct for mistakes agents make given that they do not always conform to the requirements of the classic framework. In so doing, it has been focused on individual mistakes, ways in which individual choices diverge from what is really in an agent's interests. The upshot of my argument is that there is another type of mistake that should be acknowledged by welfare economics, namely irreducibly diachronic mistakes. The possibility of this kind of mistake arises from one ground for criticising the dominant 'true preference' approach, namely non-uniqueness. Yet, it has not been recognised as such by the literature critical of this approach. I want to close by arguing that even alternative frameworks would do well to explicitly acknowledge the possibility of irreducibly diachronic mistakes.

Irreducibly diachronic mistakes are missed by welfare economists if they focus exclusively on identifying mistaken individual choices. And this worry applies to the dominant, 'true preference' framework and alternative frameworks that allow for indeterminacy alike. For the reasons given above, approaches that allow for indeterminacy in welfare actually more faithfully represent agents in the context of non-uniqueness, while the 'true preference' approach tends to overidentify *individual* choices as mistaken. But if we apply such frameworks in a way that still presupposes that only individual choices can be mistaken, we will miss the kinds of mistakes that the agents in *dynamic casino* or *dynamic shoe shopping* make. Context-dependence and inconsistency may often not be practically problematic in one-off choices. But they *are* problematic when they make us end up with unambiguously bad outcomes over time.

To be clear, the progress made by such alternatives is that we no longer treat just any inconsistency or violation of an axiom of the standard framework as a mistake to be corrected, and I am not proposing we should go back on that. Indeed, one of the advantages of such alternative frameworks is that they refocus attention away from the plausibility of the various axioms of the standard framework. Many cases of context-dependence or inconsistency never result in an irreducibly diachronic mistake, since they never lead to sure loss. Permissiveness seems like the right response here. Irreducibly diachronic mistakes only happen when a series of choices leads an agent down a clearly bad path. The small losses in our examples seem trivial, and we can imagine many instances of irreducibly diachronic mistakes remain transitory, potentially leading to corrections by the agent herself. But repeated diachronically mistaken choice behaviour can have significantly bad consequences.

My hypothesis is that irreducibly diachronic mistakes with more serious consequences are indeed quite common. There are many things, for instance, that we take to be bad for us if we do them too often, but for which we do not take total abstinence to be required. I do want to generally eat healthily; but I also do not want to live a cake-less life. I make a mistake when I indulge in several pieces of cake a day; but at the same time, there is plausibly no individual cake-eating instance that is a mistake in itself, independently of it being an instance of an unhealthy habit. It simply does not make enough of a difference; and for any such cake-eating occasion it might well be true that there is an alternative healthy habit that would allow me to indulge on just that situation. Another kind of common phenomenon that would fit in this mould is wavering consumer behaviour, where people shuffle back and forth between

options in a way that is cumulatively quite costly (e.g. in terms of time and effort of exchanging goods), while each option is in fact a perfectly permissible way of serving their goals. Something clearly goes wrong, but we can't identify any individual choice as mistaken. And it seems there is an opportunity here for behavioural economics to help, in particular by using choice architecture to encourage self-correction. Still, even alternative frameworks may fail to detect the relevant mistakes even when they are significant.

Take Bernheim and Rangel (2009) framework as defended in Bernheim (2016). In the first step, the approach only discards data on clearly mistaken individual choices as not worthy of deference. And suppose that in the second step, we proceed to only apply the framework to individual decisions separately. In the case of *dynamic casino*, taking into account choice data about many contexts in which the agent can pay $30 for bet B, you may find she unambiguously chooses to do so. Likewise for the choice of A over B, and for the choice of paying $29 for bet A. You may conclude no mistake was made by our agent at any time. However, if you were to consider choice data including choices between all of bet A, bet B, $29 and $30 – all the outcomes you could potentially reach in your series of choices – you would find that the agent unambiguously chooses $30 over $29. This would enable you to detect the mistake that happens in the dynamic choice scenario – if you actually look for it, that is, by considering the results of entire series of choices agents make. In order to detect irreducibly diachronic mistakes, we need to make sure to apply the framework with the right temporal scope.

Bernheim and Rangel (2009) explore the theoretical possibility and merits of basing welfare analysis on choices in contexts where agents face the potential outcomes to be reached in a series of choices in a single choice (pp. 86-87). But they provide no general guidance or criteria for how to pick the right scope in practice.[5] The concept of an irreducibly diachronic mistake as I have characterised it is helpful here. To spot such mistakes, we need to make the scope large. But we can't make it so large that we have grounds for thinking an agent's underlying values in fact change in the course of the sequence of choices.

One practical cost of extending the temporal scope to look for irreducibly diachronic mistakes is that the empirical burden of identifying such mistakes is higher, and that we may need to rely more on hypothetical choice data than choice data from the field. For instance, most people never face one choice of whether to have a daily piece of cake on the train home from work over the course of a year. Yet having a daily piece of cake is both a potential outcome of the dynamic decision problem that is one's daily snack choice, and something one unambiguously would never choose when a strategy of, e.g., having cake only once or twice a week is also available. An alternative to eliciting this choice data would be to abandon strict adherence to revealed preference theory and rely on the kinds of assumptions about what people generally care about that adherents of the 'true preference' account often make: That people tend to value their health, and are willing to make at least some sacrifices to attain better health.

[5] They in fact suggest we should be guided by finding a scope at which there is greater coherence (while acknowledging we may not find complete coherence). But where inconsistency stems from non-uniqueness, there may be no coherence to be found, and it may be unhelpful to strive for it – as the authors acknowledge, we will always find coherence when we consider only single choices, and so there is a temptation to go too narrow.

When it comes to longer term habits, such an appeal is also much more compelling than it is concerning snack choices on individual occasions (which could always be accommodated as permissible exceptions). But as we said – not all applications allow for appeal to such generally accepted values.

Despite these difficulties, refocusing attention on a larger temporal scope seems to be a very worthwhile endeavour. For one, it detracts away from the individual contexts where there is a danger of overidentifying mistakes. And, whether they are irreducibly diachronic or not, chances are that the mistakes we will identify in this way are the more significant ones, as stakes tend to become higher when we consider the outcomes of entire patterns of choice.

Funding Open Access funding enabled and organized by Projekt DEAL.

Data Availability Data sharing is not applicable to this article as no new data were created or analysed.

Declarations

Conflicts of Interest The author has no relevant financial or non-financial interests to disclose.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

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

Beck, L. (2023). The Econ within or the Econ above? On the plausibility of preference purification. *Economics and Philosophy,39*(3), 423–445.

Bernheim, D. (2016). The good, the bad, and the ugly: A unified approach to behavioural welfare economics. *Journal of Benefit-Cost Analysis, 7*(1), 12–68. https://doi.org/10.1017/bca.2016.5

Bernheim, D., & Rangel, A. (2009). Beyond revealed preference: Choice-theoretic foundations for behavioral welfare economics. *Quarterly Journal of Economics, 124*(1), 51–104.

Beshears, J., Choi, J. J., Laibson, D., & Madrian, B. C. (2008). How are preferences revealed? *Journal of Public Economics, 92*(8-9), 1787–1794.

Beshears, J., & Kosowsky, H. (2020). Nudging: Progress to date and future directions. *Organizational Behavior and Human Decision Processes, 161*(Supplement), 3–19.

Camerer, C., Issacharoff, S., Loewenstein, G., O'Donoghue, T., & Rabin, M. (2003). Regulation for conservatives: Behavioral economics and the case for 'asymmetric paternalism'. *University of Pennsylvania Law Review, 151*, 1211–1254. https://doi.org/10.2307/3312889

DesRoches, C. T. (2020). Value commitment, resolute choice, and the normative foundations of behavioural welfare economics. *Journal of Applied Philosophy, 37*(4), 562–577.

Dietrich, F., & List, C. (2016). Reason-based choice and context-dependence: An explanatory framework. *Economics and Philosophy, 32*(2), 175–229.

Doody, R. (2019). If there are no diachronic norms of rationality, why does it seem like there are? *Res Philosophica, 96*(2), 141–173.

Hausman, D. (2016). On the Econ within. *Journal of Economic Methodology,23*(1), 26–32.

Hedden, B. (2015). *Reasons without Persons: Rationality, Identity, and Time*. Oxford University Press.
Infante, G., Lecouteux, G., & Sugden, R. (2016). Preference purification and the inner rational agent: a critique of the conventional wisdom of behavioural welfare economics. *Journal of Economic Methodology, 23*(1), 1–25. https://doi.org/10.1080/1350178X.2015.1070527
Kroese, F. M., Marchiori, D. R., & de Ridder, D. T. D. (2016). Nudging healthy food choices: a field experiment at the train station. *Journal of Public Health, 38*(2), e133–e137.
Lichtenstein, S., & Slovic, P. (1971). Reversals of preference between bids and choices in gambling decisions. *Journal of Experimental Psychology, 89*(1), 46–55.
Lichtenstein, S., & Slovic, P. (2006). The construction of preference: An overview. In S. Lichtenstein & P. Slovic (Eds.), *The Construction of Preference* (pp. 1–40). Cambridge University Press.
McClennen, E. (1990). *Rationality and Dynamic Choice: Foundational Explorations*. New York: Cambridge University Press.
Rabinowicz, W. (2000). Money pump with foresight. In M. Almeida (Ed.), *Imperceptible Harms and Benefits* (pp. 123–154). Kluwer Academic Publishers.
Rizzo, M. J., & Whitman, G. (2020). Escaping Paternalism: Rationality, Behavioral Economics, and Public Policy. *Cambridge University Press*. https://doi.org/10.1017/9781139061810
Sugden, R. (2018). *The Community of Advantage: A Behavioural Economist's Defence of the Market*. Oxford University Press. https://doi.org/10.1093/oso/9780198825142.001.0001
Tenenbaum, S. (2020). *Rational Powers in Action: Instrumental Rationality and Extended Agency*. New York: Oxford University Press.
Thaler, R., & Sunstein, C. R. (2008). *Nudge: Improving Decisions about Health. Wealth and Happiness*. Yale University Press.
Thoma J. (2026). The precision fallacy: On the futility of preference purification. *Studies in History and Philosophy of Science*, 117.

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