Max Khan Hayward
m.hayward@sheffield.ac.uk

Increasing Marginal Utility
or
How to Plan a Bus Route

Abstract
Some goods, like buses, are more valuable, the more of them you have. A bus route has a value much greater than that of a series of individual buses; a bus network has a value much greater than that of a series of disconnected bus routes. These goods exhibit increasing marginal utility.

When a planner is trying to decide whether it is worthwhile to buy a first bus, how should she assess the value of that bus? Well, it depends on whether she is going to carry on buying more buses. In other words, it depends on what kind of chooser she is. If her route may connect to other bus routes in the city to compose a network, it depends on what kind of choosers the other bus planners are.

I argue that the value of goods with increasing marginal utility cannot be fully realised by choosers who respect two principles of the standard view of strategic rationality - separability and individualism. Rather, I present a familiar but unorthodox view of rationality, "team-reasoning", and develop an analogous model of rationality for resolute intrapersonal diachronic choice, "plan-reasoning", and argue that rational planning in the face of increasing marginal utility should avail itself of both of these styles of choice.

I conclude by briefly sketching how there may be a connection between these models of rationality and wider questions concerning political economy. Plan- and team- reasoning, I suggest, is particularly well suited to the goals of social democracy.

1. The Bus Planner

Here's a story, inspired by my daily life. I like to walk. But I also like to sleep in. So I often find myself rushing to work in the morning. Thus I used to consider it a great advantage that a bus runs from the end of my street right to the Philosophy Department. That bus used to come regularly enough that it was *always* quicker to take the bus than to walk.

No longer. In pattern that has been replicated up and down the country, bus frequencies were cut back after the pandemic. And when local councils urged bus companies to reinstate the old services, they were told that *there was insufficient demand for them*. So it was with me – with the bus

coming more infrequently, I couldn't rely on it, and so I had to take greater care to leave myself enough time to walk – at which point, I no longer had any need to take the bus if it came. And so I took the bus much less often. Clearly other people had come to similar conclusions. Far from increasing frequency of service, the bus company responded by cutting back service still further. Now, car traffic has increased along that artery, and the bus travels even slower than it used to. I now rarely take the bus, and, when the half-empty bus rolls by, I see that others have come to a similar conclusion. In a sense, the company was right – there *is* no demand for the bus – as it now exists. But, of course, this pessimistic judgement was partially self-fulfilling.

Orthodox economic theory disproportionately focuses on goods that exhibit *decreasing marginal utility* – goods, like money, that become less valuable the more you have of them. Ten pounds means more to the homeless man than to the billionaire prime minister serving him in a soup kitchen photo-op. *But not all goods are like this*.

Buses, for example, become *more* valuable the more you have of them – up to a point. Imagine a planner for a municipal authority. She has a certain amount of discretionary funds in every budget cycle. She could use that money to buy a bus. Or she could allocate it to some other valuable cause, like the local food bank, which always has hungry mouths to feed.

But here's the thing. One bus is not very valuable. Think about a bus route. A bus that comes once an hour isn't something passengers will rely on in many parts of their lives. A bus that comes every 15 minutes is useful, but some people will still drive, and most people will need to own cars. If it comes every 10 minutes, then it's going to make a serious difference in terms of reducing traffic, so will run faster. Every 5 minutes, and people won't even need to own cars, and many streets can gain an extra lane for traffic in place of parked cars.

Buses, in other words, exhibit *increasing marginal utility*. Ten buses have far more than ten times the value of one bus. This contrasts with many other goods – like money, food, clothing, shelter – which exhibit decreasing marginal utility. Or, to put it another way, a bus route has a value that isn't straightforwardly decomposable into the value of individual buses, and a bus network has a value over and above that of a series of bus routes.

So how should our planner assess the value of that first bus, when she compares it to allocating the funding to some other worthy cause? Well, it really depends on *her*. If she will *continue* to buy buses in future budget cycles, then that first bus can be figured as part of an eventual bus route, and even network. So its value will be quite high. But if she is likely to devote future funds elsewhere, never completing a network, the value of that first bus will be low. What matters is *how likely she is to continue buying buses*. And that is determined, more than anything (given that her discretionary funds each cycle are fixed) by *what kind of chooser she is*. And, assuming that she is rational, this is in turn determined by the theory of rational choice to which she subscribes.

Now imagine that our planner is only one among many rational planners, creating a network that traverses the various municipalities of a city region. A bus route that connects to *other* bus routes in a unified and structured network is worth far more than one that takes passengers to the edges of a borough and strands them there. So what utility our planners should each place on buses depends in large part on what kind of choosers the *others* are, what style of choice they take to be rational. If the other planners take it as a reason for each of them to invest in bus routes that *they, collectively can create a bus network*, then they are more likely to do so, and so the expected utility of a first bus will be higher for all of them. In other words, the problem of increasing marginal utility is relevant to the choices of teams of agents, as well as the choices of

an agent considered over time. If they are all the kinds of choosers who are likely to work together and stick to a plan to buy buses, then those first buses will be more worthwhile.

Of course, this phenomenon is not restricted to buses. My old football coach used to say "you have to train to train", and it's true: up to a certain point you gain more from an individual workout once you are *already fit*. The value of a weeklong course of antibiotics is much more than seven times the value of one antibiotic pill. You learn more from reading a philosophy paper once you already know some philosophy. Indeed, if that paper is by Bernard Williams, you often gain more from reading it the fourth or fifth time than you did the first. Academic writing is almost tragically like this – if you compare a single day of writing to a day doing something else (especially when you're at the start of the project), the alternatives will frequently look better, when taken pairwise. Devoting today to writing only really makes sense if you know that you will also write tomorrow. More importantly, many *social* goods exhibit increasing marginal utility – bus networks, educational systems, health and social care systems, climate initiatives. In each case, starting on a plan of action seems much more valuable if you anticipate that you will *continue* with it. Increasing marginal utility is all around us – and it means the value of our options depends on the way that we tend to decide.

2. The Standard Theory of Rationality

This observation has important implications for a theory of practical rationality. Here is a standard view of practical rationality, as endorsed by act-utilitarianism and orthodox decision-theory. Rational assessment applies to individual choices. The rational choice for an agent at a time is the choice with the best (expected) utility, out of all those available to her. Even on pictures of practical rationality that are far richer than this sparse account, this is generally at

least understood to apply as an account of what the *strategic* component of rationality amounts to.

There are three, interrelated features of this conception of rationality that I want to draw your attention to. First, "bestness" is a fundamentally comparative notion; an option is only "best" relative to a frame of alternatives. So it is not a neutral matter which alternatives we take to be relevant in assessing the rationality of our choices. What frames of choices we figure as relevant will determine which options our theory of rationality tells us to select.

The answer might seem simple. I should consider only outcomes *that I can bring about by my actions*. But, in fact, things are far from simple. What is the relevant *I* and the relevant *can* under consideration? If I think of myself as a time-slice, that presents one menu of options. If I think of myself as a diachronic person, that gives me another. If I think of myself as part of a team or collective, that presents me with yet another menu of possibilities. Even if rationality were no more than selecting the best "available" option, which menu of options I consider makes a difference. The best option available to a team and the best option available to an individual are not the same – ditto for time-slices and diachronic persons.

Second, the standard theory *does* offer an answer to this question. Because rational assessment applies to each minimal, atomic choice considered individually, the relevant alternatives are those causally available to a time-slice of an individual agent.

For this reason, the orthodox account endorses the following two principles. First is what McClennen (1997) calls the principle of *separability* – each choice should be settled only on the basis of forward-looking considerations of utility; the fact that an agent has made a particular

choice or commitment in the past is, *ipso facto*, no reason at all to stick to it going forward. *Resolute choice*, wherein agents take their past decisions as independent reasons to continue with a course of action, is figured as irrational. Thinking forwards, a rational agent like our bus planner may attempt to *predict* what she will choose; to the extent that she takes her future self to be a rational agent like herself, she will predict that her future self will also respect separability, and thus treat her past choices as *ipso facto* no reason at all to continue with any plan of action. The mere fact that she may now *intend* that her future self continues to buy buses will be irrelevant to her prediction of whether her future self *will* buy buses – so long as her future self is rational (in this sense of "rational"), then at each point all she will take into account in deciding what to do is the relative expected utility of buying the next bus versus the other available courses of action. As such, there is not really any such thing as a rational assessment of the various time slices of this agent considered *together* – there exists no chooser who can choose to build a bus route over time.

The interpersonal analogue of this is *individualism* – for each agent, the rational choice is a matter of *predicting* what other agents will choose – the value of an option that can only be brought about collectively is not directly relevant to the rationality of anyone's choice, since *this* alternative is not a member of any individual's frame of options. The fact that my performing a certain act would constitute me playing my part in a collectively-best plan of action is *ipso facto* no reason at all to take that action, unless I have already independently predicted that every other relevant agent will play *their* part. And if this is what rationality requires, and if I take the others to be rational, I will predict that the others will not take this kind of consideration to be a reason for them either.

Thirdly, the orthodox conception of rationality is fundamentally *reactive*. The options and their utilities are conceived as independent of the choice to be made, and fully determine its rationality; as Bernard Williams says "it is the actual state of the causal nexus that determines the decision" (1973, p.115). First we find out how the world is: this directly fixes the rational response. There is only one rationally permissible function taking us from states of the world to actions – the expected utility calculus. We never, therefore, consider the fact that the *way* in which we choose shapes the options that are open to us, when we assess the rationality of different styles of decision-making.

3. Why I Think These Claims Cannot be Correct

I want to suggest that the problem of increasing marginal utility shows that the separable, reactive conception of rationality cannot be correct. Rational assessment cannot merely be a matter of reacting to the options presented by the causal nexus, because rationality guides other points in the nexus just as much as it guides the choice under consideration. The utility of buying a first bus depends on *what kind of chooser* our bus planner is. This in turn depends (we hope) on what she thinks is a *rational* way to choose. And, with any luck, her *beliefs* about what is a rational way to choose will depend on the *facts* about what rationality requires.

As such, norms of rationality *determine* the facts and utilities to which individual choices attempt to react. And this creates a confusing situation when trying to assess the rationality of individual choices in cases like that of the bus planner, where the utilities of an option are fixed by other rational choices taken at other points in the implementation of the plan – by other planners, or by the same planner at later times.

To show this, let's start by imagining our bus planner in isolation – in this version of the story, she is the only planner in the city, so we can abstract away from the problem of her cooperative interactions with other planners. Suppose that she accepts the standard conception of rationality, including the principle of separability. She confronts each choice anew, and selects her actions on the basis of whatever action has the highest expected utility. Now suppose she is wondering whether or not to buy that first bus. In order to do that, she has to assign an expected utility to the initial bus. But, as we have seen, one bus on its own is not very useful – what matters most to the utility of a bus is whether our planner will continue to buy more buses. So, in order to know whether she *should* buy a bus now, she needs to know whether she *will* buy another bus in the next budget cycle. But, so long as she takes her future self to be rational, then she will assume that she *will* buy another bus in the next cycle only if she rationally *should* buy a bus in the next cycle. But *that* question is almost the same as the question she started with![1] In other words, it looks as though she needs to know what rationality requires in a case just like the one she started with, in order to know what expected utility to attribute to her options; but if rationality tells her nothing more than to select the option with the greatest expected utility, then she is stuck at an impasse. She needs to know what she ought to do in a situation almost exactly like the one she faces *before* she can know what she ought to do in *this* situation.

I'm not quite sure what the standard theory should say in this case. One option faced with such uncertainty might be to apply the game-theoretic *principle of indifference* – when faced with total uncertainty about the choices of another player in a game, the principle tells us to assign equal

[1] It's not identical of course – given increasing marginal utility, a second bus is a bit more useful than a first bus. But it's not *very much* more useful – not so useful that it's likely to be more worthwhile than the alternatives, *unless she is going to buy still more buses*. So the choice in the next budget cycle will also depend on what the bus planner is going to do in the *subsequent* budget cycle. In other words, the choice in the next cycle faces just the same problem – it requires pre-empting a future rational choice in order to generate the expected utility score needed to make a present rational choice.

probabilities to all available options. In this case, of course, the other player is the bus planner's future self. If we take the options to be only two – buy a bus or spend the money or something else, then we are led to the conclusion that the bus planner should assign a 50% chance to her future self continuing to buy buses in the next budget cycle. But of course, unless the non-bus options are particularly dire, this simply resolves the situation against buses – if the chance that our planner will continue to buy buses in the next budget cycle is only 50%, then it will probably be clear that she shouldn't bother buying one in this cycle either.[2] (Things will look even worse if we take her probabilities to be distributed over even more than two options.)

Another option would be to apply backwards induction. The choice of buying a bus in the very next budget cycle might be similar to the choice of buying a bus in this cycle, such that its utility depends on still more distant choices. However, as marginal utilities increase the disanalogy will grow sharper. At some point, perhaps the marginal utility of each bus will be so high that it would be worth buying *whether or not* the planner were going to continue to buy buses. In that case, there will be one decision problem in the chain that the planner can solve without having had to already have solved a decision problem still farther along the chain, and so she can work backward from there in order to break the impasse.

However, even if such solutions will be available in some cases, I don't think they will always be available – and especially not in cases like bus planning. After all, successful bus networks help cities to grow in geographical size and population; they also tend to end up with a greater proportion of the population in ridership as they become more useful. On the other hand, when

[2] Furthermore, this argument risks creating a contradiction analogous to one that Colman and Gold (2020) identify in similar cases of collective action. Our bus planner starts with the hypothesis that her future self has a 50% chance of buying another bus in order to reach the conclusion that her present self must not buy a bus. But her future self at time t+1 will presumably be able to apply just the same reasoning in thinking about *her* future self at time t+2 to decide that she *must not* buy a bus. And so this argument simultaneously supports the belief that the planner at t+1 has a 50% chance of buying a bus, and the belief that she has a 0% chance of buying a bus.

mass transit systems start to get overcrowded or erratic, or don't go to all the places that riders need to go, that can set in motion a negative spiral of *decline*, as passengers start deciding to buy cars and the transit network loses the ridership needed to fund itself – just the sort of decline I mentioned at the start. In other words, in such cases there may be no point at which planning reaches a finished steady state, such that the value of a particular intervention can be assessed independently of future decisions – in cases like infrastructure planning, it'll often be the case that the only options are positive spirals of ongoing growth or negative spirals of decline. In that case, we cannot reasonably foresee that there will be a step in the process where the value of an atomic choice can be suitably assessed without knowing what future choices are likely to be made. The value of my choices today may *always* be dependent on the choices I make tomorrow.

Here is a final option for the standard view. We might think that the foregoing problems have arisen from an illegitimate attempt to graft a game-theoretic approach, of predicting future choices on the basis of their rationality, onto a decision-theoretic framework of inductive inference. Rather than our planner trying to work out whether she *will* continue to buy buses by attempting to determine whether it is *rational* for her future self to buy buses, she might instead simply try to predict whether she will continue to buy buses by performing an induction over her past actions. If she is normally the kind of person who sticks to her plans, then she should expect her future self to continue buying buses, and thus assign a high expected utility to bus-buying in the here and now. If she is not psychologically resolute in this way, then she should not have this expectation, and thus should not assign a high expected utility to buying that first bus. This, again, can dissolve the impasse.

One problem with this approach is essentially philosophical – it requires the bus planner to see her future self as merely a brute force of nature, rather than as the rational agent she is. The whole point of a theory of rationality is to help us determine what to do and how to choose, but this answer holds us hostage to our own past decision-styles in a way that seems to neglect our capacity for rational choice. And there are also more practical problems. She may not have relevant past experience from which to try to draw this inductive inference. After all, resoluteness may not be a domain-general psychological disposition, but plausibly varies based on the situation – just because she changes her mind easily in her private life that doesn't indicate that she will do so *in her capacity as a city planner*. Most worryingly, it's quite likely that what she knows from past experience is that the most significant determinant of whether she makes decisions of this sort is *whether or not rationality recommends them.* If that is so, any attempt to sidestep the ratiocinative frame in favour of an inductive frame fails, because the latter leads us right back to the former.

If these considerations are correct, then the standard conception of rationality simply won't be able to satisfactorily settle the question of whether our bus planner should buy that first bus or not. Either it will give no answer at all, or it will give an excessively pessimistic answer, or it will draw inductive inferences from past behaviour in an inappropriate manner. And if you think, as I do, that decisions of this sort are of vital importance, and *exactly* the sort of thing that a theory of rationality ought to be able to help with, then this should lead you to look beyond the standard theory for a conception of rational choice that will help in these cases.

4. An Alternative Model of Choice

I've argued that it is unclear, on the standard model, what expected utility our bus planner should assign to the option of buying that first bus – if, that is, she respects separability. So perhaps we can sharpen our thoughts by considering two alternatives.

First, let's imagine that our planner does *not* expect her future self to continue buying buses, and so she assigns an expected utility to buying that first bus on the basis of the usefulness of one bus taken in isolation. This will, of course, be a fairly paltry expected utility rating. Unless her other options for spending her money are pretty dire, this means that she *should not* buy that first bus.

Alternatively, let's imagine that the bus planner imagines her future self as *resolutely committed* to the project of bus-buying. In that case, she will be able to assign a rather high utility to buying that first bus, since she could be confident that her future self would continue to buy buses. In this situation, it seems rationally recommended for the planner to buy the first bus.

The point of this comparison is simply to emphasise the claim made earlier – that *the expected utility of buying the first bus depends in part on what kind of chooser our planner is*. This serves to allay a familiar worry about resolute choice – that resoluteness leads us to suboptimal results, because the resolute chooser will stick to her plans even when better options are available, and so resoluteness is irrational. This objection presupposes that there is a fact of the matter about the expected value of options *independent* of the styles of decision-making employed by the agents who are pursuing those options and the wider plans of which they are a part. But that is not the case, because there is a feedback loop between our styles of decision-making and the utility of our options. The more resolute we are, the better the sorts of goods that exhibit increasing

marginal utility look. The more irresolute we are, the comparatively better the expected value of goods that exhibit decreasing marginal utility.

My point is not that a planner must always rationally continue to buy buses. Maybe, in the future, the situation will change, and so in an important sense it may seem rational for her to redirect the funds elsewhere. My point is that the value of buying a bus, *at each point*, is affected by how one's future self, and other rational agents, are going to choose. Because this is a case of increasing marginal utility, the more resolute our planner is, the better an option buying each bus is. The less resolute she is, the more dicey it looks to invest in a single bus – the expected marginal utility will be lower.

Since I am not sure how a separable chooser should solve the problem discussed in the previous section – the problem of second-guessing her future choices in order to assign an expected utility to the first bus – I am not quite sure what expected utility a separable chooser should assign to the first bus. But since separability is defined in contradistinction to resoluteness, it seems clear that however the separable chooser solves this problem, she will assign a lower expected utility to buying the first bus than a resolute chooser, since she will be more uncertain about whether her future self will follow through on creating the entire bus route. An agent who eschewed resoluteness altogether might never bother with buses in the first place – better to focus on goods with decreasing marginal utility, whose value can be assessed on a choice-by-choice basis, without having to second-guess other decisions.

Here is another way of understanding the problem. What I've been arguing is that it makes a difference whether we think about our options in terms of *extended plans of action*, rather than as a series of individual choices. When we compare plans of action against one another, one option

might seem like the "best" one, but that wouldn't necessarily be the choice we would get by repeatedly picking between atomic options. But separable choosers *cannot* choose extended plans of action.

So how can we enjoy the benefits of resolution, whilst also recognising the pitfalls of the classic account of resolute choice? We can start by drawing the following distinction. When we consider the value *of a completed bus route*, the determinants of expected utility are *external* to the decision-making style of the planner involved. To assess the value of a bus route as a whole, we do not need to know how the planner involved makes her decisions. We just need to know how useful the buses are, how likely people are to use them, and so on. It is only when our loci of evaluation are the *individual choices* of the planner, that questions *internal* to her, about her decision-making style, are relevant for assessing expected utility, such that our planner is forced to second-guess herself in order to decide what to do. In other words, it is the *internal* determinants of expected utility that create the problem for rational choice explained above.

Therefore, when facing the prospect of embarking on a plan of bus-buying, I propose that the bus planner shouldn't start by asking whether buying a first bus has a higher expected utility than any other option open to her. Rather, she should start by comparing the expected utility of *buying an entire bus route* to the expected utility of the other options that she currently anticipates having. Then, if the plan of buying an entire bus route has a higher expected utility than that of the other things she foresees being able to do with her budget, she should take the first step in that plan by buying a bus. In other words, she should choose *as though* she could now control all her future decisions, in the manner of a resolute chooser. Or, to put it another way, she should change how she frames her options from a synchronic to a diachronic framing. Rather than asking:

> *Synchronic choice:* What is the best option open to me, now, given what I anticipate that my that my future self will choose?

She should ask:

> *Diachronic choice:* What is the best plan that I can enact over time? How can I play my part in that now?

If the best plan she can achieve over time involves a good with increasing marginal utility, then she should embark upon it, and cease to think of her choices in each budget cycle in separable terms. This way of framing the question *screens off* the internal determinants of the planner's agency. Such an approach allows the agent to match her style of decision-making to the options in front of her, rather than starting with a fixed decision-making style and allowing that to affect the foreseen value of her options.

This change of framing may remind readers of the similar transformation involved in *team-reasoning*.[3] When agents engage in team-reasoning, they ask not:

> *Individual choice*: What is the best thing I can do, given how I predict that everyone else will act?

But instead:

> *Team choice:* What is the best thing that we can achieve together? How do I play my part in that?

This is not coincidental. As Natalie Gold (2013) has suggested, team-reasoning stands to individualistic choice as intrapersonal diachronic choice stands to synchronic choice. In each case, a plan of action requires the coordination of rational choosers at different points – in one case, these rational choosers are distributed across space, in the other, across time. In each case, the change of framing allows the agents to screen off the determinants of expected utility

[3] I am here drawing from Bacharach's influential 2006 account.

that are internal to their decision-making styles, and simply select plans of action as a whole on their own merits, matching their decision-styles to the problem at hand. And, of course, coordination problems like the Stag Hunt and the Hi-Lo Game are the synchronic analogues of cases of diachronic increasing marginal utility – they involve choices (like hunting stags) that are more valuable when more people choose them.[4] Given this analogy, we can call the model of diachronic choice that I propose *plan-reasoning*.

I earlier bracketed the question of coordination between different planners cooperating to create a whole bus network across the different municipalities of a city region. In the one-planner case, the right response to the phenomenon of increasing marginal utility is to re-frame a series of atomic choices into a single diachronic plan, and choose resolutely if the plan with increasing marginal utility is the best one. In the multi-planner case, the planners should *additionally* reframe their choices of plan as a collective choice, and choose cooperatively if the team-plan with increasing marginal utility is better than any of the other plans they could pursue alone.[5] When considering the expected utility of an entire bus *network*, the planners shouldn't start by trying to second-guess one another's choices and their own future choices – they should screen off these internal determinants of expected utility, and assess the value of the entire bus network on its own terms. If this is better than the other options that they anticipate having, they should play their parts as team- and plan- reasoners. In that case, both problematic doctrines of the traditional conception of rationality – separability and individualism – are rejected.

[4]And similarly, as I have argued elsewhere (*citation redacted*), in cases of coordination between ideally rational agents, it is similarly indeterminate what the standard theory of rationality would recommend in such coordination cases, since the expected utility of each agent's choices depends on what she predicts that her identically situated partners are going to do – she cannot know what she should do until she knows what they will do, but since they are rational she cannot know what they will do until she knows what they should do, and she cannot know what they should do without already knowing what she should do, since they are identically situated.

[5] Supplementing team reasoning with resoluteness addresses a problem I discuss in (*citation redacted*).

However, the diachronic intrapersonal case is not an exact mirror of the synchronic interpersonal case. In one way the situation is easier. A significant problem for theories of team reasoning, well-discussed by Mellor (ms), is the problem of *willingness* – selecting actions that are valuable only if others cooperate only seems rational if we know that they *are* willing to cooperate. And we often have no idea whether this is the case.[6] However, with intrapersonal plan-reasoning, we are not flying blind in this way. If a planner chooses resolutely, it is of course possible that her future self will be akratic, or disillusioned, or will simply lose her job as a planner, but she clearly exercises a level of control over the decision-style of her future self that potential teams of cooperators lack over one another's decisions.

However, there is one way in which the case of plan-reasoning over time is more complicated. After all, as time goes on so do situations change, and it would seem irrationally stubborn for the planner to *entirely* rule out the possibility of changing course in the light of a change of circumstance. So we need to find a model that explains the possibility of flexibility without collapsing into the model of separable choice that I have already proposed we reject.

Here's the model I propose.[7] When circumstances change, the plan-reasoner *should not* compare her occurrent options in a separable, atomic manner, on the basis of their expected utility. She shouldn't ask herself whether buying the next bus in the series is better than the other actions that she could perform right now. That would simply revisit the problems mentioned above.

[6] The problem is a bit more complex than this, but we need not go into details here. I attempt to respond to Mellor's framing of the problem in my (*citation redacted*).

[7] This proposal is inspired by Doody's 2022 discussion of certain apparent paradoxes of separable reasoning proposed by Hayward (2020). Doody argues that in the cases identified as problematic by Hayward, what has gone wrong is nothing more than the agent was unlucky – she made a rational assessment of probabilities, but things turned out differently. While my argument has some parallels with Hayward's, it aims to avoid Doody's objections. What Doody overlooks is that there is more than one way to respond to changing circumstances and evidence – such responsiveness does not require separability.

Rather, re-evaluation in the light of new evidence should replicate the original form of reasoning that prompted the planner to commit to the bus route in the first place – she should compare the value that completing the bus route will add to the value of the buses already bought, against the value of the alternatives that she *now* anticipates having. If it looks as though completing the bus route will end up being much less useful than projected, or the alternatives – not just right now, but throughout the future duration of the plan – will be significantly more valuable, then she can rationally change course. When the plan-reasoner re-evaluates in this way, she continues her original policy of screening off the effect of her own decision-making style from her assessment of the utilities open to her, considering only the external determinants of utility – she only compares *entire plans of action* to their alternatives, not atomic options.

This model of choice ensures resolution over time. But it does so in a different way from the standard model of resolute choice, in which a resolute chooser treats the mere fact of her past decision as an additional reason to continue with a course of action going forward. On the proposed model, the plan-chooser does not treat her past decision as a reason to *act*, but as a reason to *continue framing her options* in a plan-choice manner. We've already seen that, when we count the internal determinants of utility, this makes the value of plans of action with increasing marginal utility either indeterminate, or worse than if we screened those off. Therefore, committing to continue screening off these considerations increases (or at least makes determinate) the expected utility of continuing with the plan. So someone who knows that her future self will continue to plan-reason has a good basis to think that her future self will carry on with the plan – after all, she will only deviate from it if it turned out that her original decision to take it on was mistaken anyway, and the plan taken as a whole is not really better than the set of alternatives. Indeed, because the plan exhibits increasing marginal utility, there will even be cases where the initial decision to undertake the plan *was* mistaken, but where it is still the

case, looking forward, that continuing with the plan is better than any of the remaining alternatives (perhaps it was a mistake to start building the bus route, but, now that we have built some of it, finishing it is better than any other option).

However, the *fact* that she is in fact likely to carry on with the plan is *not* part of her reason to commence the plan. After all, if she is engaged in plan-reasoning, then she is *not* attempting to second-guess her future choices. It is precisely by not trying to work out whether she *will* be resolute that our planner ensures her own resolution. And this may seem strange – I consider whether this is an objection to my view in the next section.

This last point makes clear the way in which plan-reasoning outruns the resources of the standard model. Many attempts to reconcile separability with problems of diachronic choice appeal to the notion of *Sophisticated Choice*, in which agents build a prediction of their future decisions into their assessment of present options. Of course, it is precisely this form of second-guessing oneself that generates the problems discussed here.[8] So we might say that plan-reasoning is a form of *Hyper-Sophisticated Choice*, in which agents factor in the effects of their own decision-*styles* on their options. By *screening off* internal determinants of expected utility arising from uncertainty about their own future choices, and committing to doing so in the future, plan-reasoners can rationally compare entire plans of action taken as a whole. Just as team-reasoning transforms the unit of agency from the individual to the team, so plan-reasoning transforms the unit of agency from a time slice of an agent to an agent extended over time. This allows agents to enjoy the fruits of increasing marginal utility without the perplexities of predicting their own future actions.

[8] Here, again, I disagree with Doody (2022). Doody thinks that separable choice can be rescued by sufficient Sophistication. I think that, in these cases, Sophistication is precisely the problem.

5. Objections, Alternatives and Perplexities

The problem of resolute choice has primarily been explored in the context of preference change and temptation. In that context, resolute choice has often been regarded as problematic. Given that plan-reasoning presupposes diachronic resolution, we might worry that similar objections apply here.

One problem is what Holton (2004) calls the *problem of akratic resolution*. In the context of a change of preference, and a concomitant change in judgement[9], it does look subjectively irrational to continue with a past resolution. But I am not discussing resolute choice in the context of *preference* changes – I don't assume that our bus planner changes her goals of creating public goals. Of course, it might be that she *does* change her judgement about whether to continue buying buses. But so long as she changes her mind by employing plan-reasoning, she is permitted to change course. On the other hand, if she changes her mind after thinking about her choices in a separable manner, then she is violating the norms of rationality appropriate for her situation. So I conclude that this problem of akratic resolution does not arise for plan-reasoning.

Another objection is what Bratman (1987) calls the *bootstrapping problem*. On the standard view of resolute choice, agents take their past decisions as additional reasons to continue with a course of action, over and above the future utilities of that choice. The *bootstrapping* objection charges that this constitutes agents *creating* reasons *ex nihilo*. Since plan-reasoning aims to make agents more resolute over time, we might worry that similar problems arise here. However, plan-reasoning does *not* involve agents creating new reasons for themselves *ex nihilo*. Plan-

[9] On Humean conceptions of judgement this judgement shift will be understood as constituted by a change in preferences, but, as Holton points out, even non-Humeans must acknowledge that, as a matter of fact, preference shifts often do occasion judgement shifts.

reasoners will indeed rationally behave in a more diachronically stable manner than separable choosers, but they do this *because* they have changed their style of decision-making. And they adopt a planning style of decision-making because of the kinds of goods that confront them – specifically, because they are confronted by goods with increasing marginal utility. This only looks like bootstrapping if we presuppose the crudely reactive conception of rationality that I have been arguing against, in which the causal nexus straightforwardly determines the rationality of each action. Once we realise that styles of rational decision-making are themselves part of what affect the utilities of choices facing an agent, and that there is no one privileged way for agents to frame their options, it becomes clear how adopting the framing of plan-reasoning can change what is rational for agents to do.

Nevertheless, you might think that the positive goal of predictable diachronic consistency can be achieved without such a radical emendation to the traditional theory of rationality. Two main alternatives can be found in the literature. First, our bus planner might, as in Elster's (1979) case of Ulysses and the sirens, engage in *self-binding*. She might arrange it such that she will suffer some kind of penalty if she deviates from her initial course of action. But while this might be our only option in cases of foreseen *preference change*, it seems inapt as a response to the problem of increasing marginal utility.

Firstly, self-binding incurs costs, of two kinds. Simply arranging things to bind oneself takes effort and often expense. Moreover, it means that the penalty will arise even in the cases where our planner *should* change course due to large changes in external circumstances. If some degree of diachronic consistency can be achieved simply by changing our style of decision-making, as I have proposed, then self-binding looks *unnecessarily* costly.

Secondly, self-binding looks unnecessary in a more fundamental sense. Self-binding in the face of sirens and other temptations might be necessary due to the fragility of human nature – we simply *cannot* stop our preferences from changing. But the case of the bus planner involves no such compulsion. If an separably-choosing bus planner were to be irresolute, she would be doing so on the basis of what she took to be rational considerations. If she had reason – as I propose she does – to bring it about that she does not readily deviate from her course, then rationality alone should be able to prevent such deviation, without the need for crude self-manipulation. The case of increasing marginal utility shows the rational value of different styles of decision-making, and so all we should need to do to respond to this fact is to re-evaluate our conception of rational choice in order to choose in the right way. We might say – *if rationality is the problem, than rationality can also be the solution*.

A second alternative from the literature is Holton's own proposal of *rational nonreconsideration*. On this view, rather than taking our past decisions as reasons to continue with a course of action, we take ourselves to have a reason not to *reconsider* our selected courses of action in certain cases. And, indeed, Holton's approach is somewhat sympathetic to my own. However, again, Holton is responding to a different problem from that to which I am responding. In the case of foreseen temptation, avoiding reconsideration might be the best that we can do, psychologically. But the case of the bus planner is not a case of temptation, in this sense. My model allows the planner to reconsider, so long as she only does so within the plan-reasoning frame. Moreover, it is often part of the institutional structure and democratic nature of such planning that reconsideration is always an option – the bus planner (or her elected superiors) must always be available to present and justify her reasons to the demos or to relevant stakeholders, and to do so in a way that is at least *open* to reconsideration. My proposal allows for a greater democratic and institutional transparency. Rather than simply refusing to consider

changing course, the planner and her colleagues can honestly explain that they will stick to their plans *because it is important for planners to stick to their plans*. If what I have said above is true, then this is a perfectly good justification.

There is one final perplexity, to which I alluded in the previous section. I have argued that committing oneself to plan-reasoning engenders greater diachronic resolution in agents. However, it is entailed by the structure of plan-reasoning that this fact *does not* factor into our planner's decision-making. It is not that she screens off her future decisions, treating an entire plan of action as though it were something that she could select all in one go, *because* she predicts that her future self will be resolute. Rather, it is only true that her future self will be resolute *because* she is committed to screening off her own future decisions. It is because she doesn't second-guess herself that she is in fact likely to stick to her plans.

But, of course, my model allows plan-reasoners to change course when their evidence changes significantly. Thus, even on the assumption that the planner's future self will continue to choose rationally, and in a plan-choice manner, it is still not guaranteed that she will stick to her plan. So how can it make sense for her to screen off this possibility when deciding what to do at the start?

I have two responses to this. First, on my model a planner starts by comparing the value of a completed plan to the value of the set of alternatives that she anticipates having during the period of the plan; she will only commit to plan-reasoning if the former is greater than the latter. Of course, if she changes course then she won't complete the plan, and so it won't achieve the total projected value. But she will only rationally change course if she was in fact *wrong* about the comparison between the plan and the alternatives – if she takes the plan to be worth

choosing, then, *by the same token,* she predicts that she will stick to it. Of course, she might be wrong! But to take that uncertainty into account *both* when assessing the value of the plan versus the alternatives, and then *also* when determining how likely she is to complete the plan (and hence, what expected utility to assign to *trying* to enact the plan), seems to me like double-counting. The planner should take her uncertainties into account fully in comparing the value of the plan to the value of the alternatives; if, having done this, the plan looks better, she need no further concern herself with any uncertainty as to whether she will execute the plan – given her current beliefs, she should suppose that she will.

Still, even if this is correct (which it may not be – I think the response I am about to give is more important), you might think there is something rather weird in screening off certain possibilities in the first place. If it is possible that a planner, even while rational, should fail to complete her plan, then is this not *evidence* that the planner should take into account in assessing the expected utility of her options? Isn't screening off one's future decisions tantamount to ignoring evidence? I want to say that it's not. Rather, when we try to combine *deliberating* about what to do with *predicting* what we will do, we get ordering effects. If I start by predicting what I will do, this may rationalise a particular *decision*; if I start by deliberating about what to do, this may rationalise a different *prediction*. The former ordering cannot be uniquely rational, or else we would be rationally bound to predict our answers to all deliberative predictions before deliberating (after all, if I predict that I will not choose to *phi*, why bother deliberating as to whether to *phi*?).[10] Many philosophers accept that it is rational to screen off our predictions about how we will choose in a particular case while we deliberate about *that very choice* – I am just extending this

[10] Of course, this gets us into difficult questions about whether and why "deliberation crowds out prediction". As may be apparent, I think that it does. But this is a complex issue which I cannot settle here. However, since I think that very many philosophers agree with me, I hope that at least *they* will find my argument persuasive.

thought to encompass suites of actions in addition to atomic actions. We can start by deciding whether or not to enact a plan *before* trying to predict whether or not we will do so.

6. Social and Political Implications

The orthodox theory of rationality is the one presupposed by mainstream free-market economic thought, and thereby employed in much public and social policy. As I've suggested, this is a useful way to think about goods with decreasing marginal utility – the goods that orthodox economic thought has tended to focus on – but an unhelpful pattern of choice when it comes to goods with *increasing* marginal utility.

This is a bit speculative, but I want to suggest that it is no coincidence that free markets, filled with separable individualistic choosers, are excellent at providing the sorts of goods that exhibit decreasing marginal utility, but poor at producing goods with increasing marginal utility. Capitalist economies have done better than others in providing citizens with food, shelter and clothing. And yet, the intrusion of free market structures of decision-making into the provision of public transport, education and social care planning has, I think, been disastrous. This is because these are goods with increasing marginal utility, and so a style of planning that respects separability and individualism will be systematically bad at capturing these benefits.

So I want to suggest that there is a deep connection between the styles of decision-making we deem rational, and the sorts of political economy that we make possible. If we want a society that provides public goods like bus networks – if we want some kind of social democracy – then we should also endorse a style of decision-making that allows public planners to work together over time, resolutely and cooperatively. In other words, flourishing social democracy requires us to see plan-reasoning and team-reasoning as forms of rational choice.

7. Conclusion

It's seductively easy to think that rational choice is merely a matter of looking at how the world is, and reactively responding to the facts as best we can. But one of the most interesting features of the world is that it contains other rational choosers – ourselves at future points, and other agents both now and in the future. Rationality is – we hope – guiding all our actions (or at least many of them). So when we, collectively and diachronically, accept a particular conception of rationality, that *affects* some of the facts to which we are trying to respond. In a world that contains increasing marginal utility, that has certain kinds of goods that appear more valuable when viewed as stages of a plan than as atomic choices, then we need to ensure that we adopt styles of rational decision-making that allow us to make the most of these possibilities. I've argued that both plan-reasoning and team-reasoning are styles of choice that allow us to do just that – to make the most of the opportunities offered us by goods with increasing marginal utility, like buses and reading Bernard Williams. It is my hope that, if this were generally accepted, we might start to do a better job of creating the sorts of goods that free market systems are so often bad at supplying, from infrastructure to education. We might say (with a little embarrassment for such glibness) that a theory of rationality cannot just describe the world – the point is to change it.

Bibliography

Bacharach, M. (2006) *Beyond Individual Choice: Teams and Frames in Game Theory* (Princeton, NJ: Princeton University Press)
Bratman, Michael (1987). Intention, plans, and practical reason. Cambridge: Cambridge, MA: Harvard University Press.
Doody, Ryan (2022). Don't Go Chasing Waterfalls: Against Hayward's "Utility Cascades". Utilitas 34 (2):225-232.
Elster, Jon (ed.) (1979). Ulysses and the Sirens: studies in rationality and irrationality. Paris: Editions de la Maison des sciences de l'homme.
Gold, Natalie (2013). Team Reasoning, Framing and Self-Control: An Aristotelian Account. In Neil Levy (ed.), Addiction and Self Control.

Gold, Natalie & Colman, Andrew M. (2020). Team Reasoning and the Rational Choice of Payoff-Dominant Outcomes in Games. Topoi 39 (2):305-316.
Hayward, Max Khan (2020). Utility cascades. Analysis 80 (3):433-442.
Mellor, Rowan (ms). Why Cooperate? Team Reasoning and Unwillingness.
Mcclennen, Edward F. (1997). Pragmatic Rationality and Rules. Philosophy and Public Affairs 26 (3):210-258.
Smart, J. J. C. & Williams, Bernard (1973). Utilitarianism: For and Against. Cambridge: Cambridge University Press. Edited by Bernard Williams.