Self-Prediction and Self-Control

Martin Peterson and Peter Vallentyne

in *Self-Control, Decision Theory, and Rationality*, edited by Jose Luis Bermudez (Cambridge University Press, 2018), pp. 48-71.

1. Introduction

The aim of this chapter is to examine the conditions of rationally permissible choice for agents that have fully rational preferences and beliefs, are fully aware of their dispositions to make choices under various conditions, but *may have only limited self-control over their future actions.*

In the literature, there are three main conceptions of rational choice for agents facing a sequence of choice situations: (1) the resolute choice conception (e.g., McClennen 1990), according to which agents should adopt a plan with the best prospects, and then simply comply with that plan at each choice situation; (2) the sophisticated choice conception (e.g., Schick 1986), according to which agents should predict how they will choose in future choice situations and then make a current choice that has the best prospects, based on those predictions; and (3) the wise-choice conception (e.g. Rabinowicz 1995), which is the same as the sophisticated conception except that it allows that conformance with a rationally adopted plan can count in favor of choosing an option, whereas the sophisticated choice conception does not. We endorse, develop, and defend the wise-choice conception.

We propose, controversially, that rational agents predict how they will choose in the future by ascribing subjective probabilities to their own future choices. They thus treat future choice nodes as a kind of chance event. We argue that backwards induction is not, as commonly assumed, a necessary element of the sophisticated and wise-choice conceptions. In our view, it is merely a method that rational agents may use, in special cases, to determine what to choose in light of their predictions about their future choices.

2. Limited Self-Control

Our key points concern cases of *limited self-control* (limited willpower), which we understand broadly to include both (1) a *synchronic form* (akrasia), in which the agent chooses something, including resolutions or plans, that she judges to be rationally impermissible (e.g., not best), and (2) a *diachronic form* (lack of resoluteness), in which the agent lacks the full ability to determine, in virtue of a resolution (plan adoption), what choices she will make in future choice situations if she reaches them.[1]

Imperfect *synchronic* self-control (akrasia) arises when an agent's *motivational* *preferences*, which determine her actual choice disposition, are not aligned with her *rational preferences*, which reflect the agent's values (i.e., normative considerations that, for her, speak for or against the various options). Imperfect *diachronic* self-control (lack of resoluteness) arises when an agent does not always choose in accordance with previously adopted plans. This arises, for example, when an agent's future motivational preferences do not give lexical primacy to conformance with adopted plans. If an agent's rational preferences (values) give lexical primacy to such conformance, and if the agent has perfect synchronic self-control (no akrasia) now or in the future, then her future motivational preferences will also give lexical primacy to such conformance, and the agent will be perfectly resolute.

To illustrate the issue, consider the following example from Carlson (2003). Suppose that Alice knows she will get two opportunities for eating a chocolate bar. Suppose further that the best outcome for her is eating a chocolate bar on the first occasion but not on the second (because eating one bar earlier is better than eating one bar later). The second best outcome is eating a chocolate bar on the second occasion but not on the first (because one bar is better than none or two). The third best outcome is eating no chocolate bar (because being hungry is better than eating two bars). The worst outcome is eating a chocolate bar on both occasions (because it is very unhealthy). Suppose that Alice knows all this. Is it rationally permissible for Alice to eat the chocolate bar on the first occasion?

If Alice does not suffer from akrasia, and is fully aware of her choice dispositions, it is, we claim, rationally required that she choose to eat the chocolate bar at the first time (the best outcome). If, however, she suffers from later akrasia and has no resoluteness, and she will eat a second chocolate bar in any case, then it seems plausible that, if she knows this, it is rationally impermissible for her to eat a bar at the first time. Finally, if she suffers from akrasia, but is perfectly resolute, then it seems plausible that it is rationally required that she adopt a plan not to eat the second chocolate bar, eat the first chocolate bar, and then comply with her plan. Or so we shall argue.

3. Parametric Dynamic Rational Choice

We focus on what rational choice requires in the context of parametric dynamic choice situations under risk. These dynamic choice situations are *parametric* in that the outcomes of an agent's choices depend solely on the choices she makes and on chance events (acts of nature). Unlike *strategic* choice situations, outcomes do not depend on choices made by other agents. We focus on *dynamic* choice situations, which are situations in which the agent makes a sequence of choices over time (not merely a single choice). We focus on choice *under risk*, where we assume that the agent assigns probabilities (and not merely possibilities) to chance events. We explore the extent to which rational choice in such contexts is also based on the probabilities the agent assigns to her future choices.

Our aim is to determine what (feasible) options, in any given choice situation, are *rationally permissible*. If there is only one rationally permissible option, then it is *rationally required*. If there are several, then each is *rationally optional*. We shall assume that a choice is rationally permissible if and only if it is a *best feasible option*, i.e., one that is at least as good (relative to her preferences) as any other feasible option. Our core argument can be generalized to cover satisficing theories and to cover cases where the agent's preferences are incomplete (in which case, rational choice requires a feasible option that is not worse than other feasible options). For simplicity, we assume a maximizing conception of rationality with complete preferences.

As noted above, we shall consider three conceptions of rationality: resolute choice, sophisticated choice, and wise choice.[2] In order to introduce these views carefully, we first need to say more about parametric dynamic choice.

A parametric dynamic choice situation is a choice situation in which there is only one agent, who faces a sequence of choice situations, each of which involves a set of feasible choices for the agent. This can be represented by a decision-tree (see, e.g., McClennen 1990, ch. 6), which consists of branching choice nodes (represented by squares) and branching chance nodes (represented by circles). Each branch from a choice node represents a choice the agent might there make and each branch from a chance node represents a "choice" that "nature" might there make. If the "choices" of nature have probabilities, those are specified as part of the decision-tree. Throughout, we shall assume, as is standard, that there is an initial (choice or chance) node, to which all else is connected.

Our chocolate example, in which there are no chance events, can be represented by a decision-tree as follows:

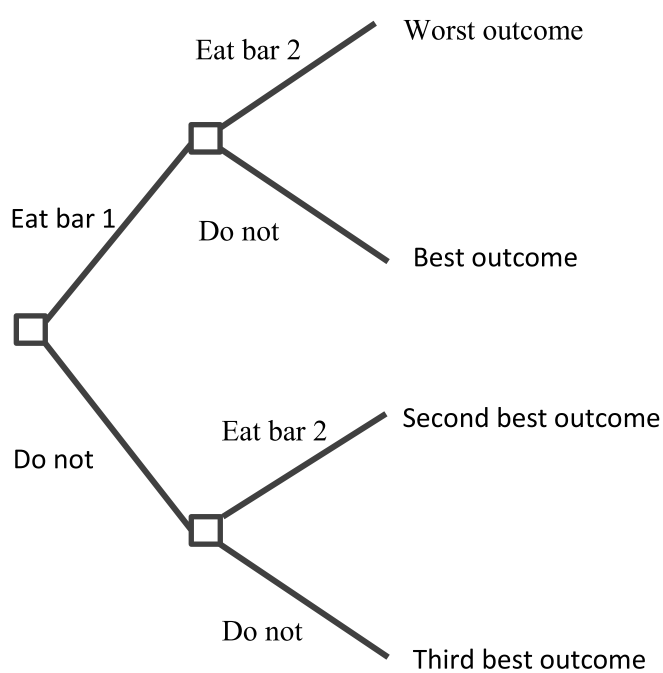

Figure 1. Chocolate: Deterministic Case

In dynamic choice situations, agents may adopt *contingency plans* (hereafter: plans), which specify a choice for each choice node reachable from the initial node by a sequence of choices specified by the plan in question and by chance events.[3] If there are no chance events, a plan is simply a branch from the initial node through various choice nodes. If there are chance events, a plan will be a subtree connecting an initial choice node to later nodes, with one branch-segment (the choice required by the plan) from each choice node and several branch-segments from each chance node (the various "choices" nature might make). Although chance events in a decision-tree need not have associated probabilities (decision-making under uncertainty), for simplicity, we shall assume that they do have associated probabilities. Thus, associated with each plan, there is a *prospect*, which is a probability distribution over the various full branches of the tree and their outcomes. These are the full branches of the tree that can be reached by choices in conformance with the plan, in conjunction with various "choices" of nature. For simplicity, we assume that there are only denumerably many branches. Thus, each such branch has a non-zero probability (as determined by the chance nodes).

With this background, let us start by arguing against the resolute conception of rational choice.

4. Resolute Choice

According to the *resolute* choice conception, rationality requires agents to comply with the plans (or resolutions) they have adopted in the past. More exactly, it does not require you to be “unconditionally committed to execute a chosen plan”. It merely requires that “if on the basis of your preference for outcomes, you [rationally] adopt a given plan, and if unfolding events, including any conditioning circumstances, are as you had expected them to be, then you proceed to execute that plan.” (McClennen 1990; 1997, p. 232). That is, rational choice does not require *unconditional resoluteness*. It only requires *rational resoluteness*, which is a kind of conditional resoluteness. We understand this to be the disposition to comply with adopted plans when (1) it was rationally permissible to adopt the plan at the time of adoption, and (2) the agent has acquired no new unanticipated information that, if available to the agent at the time of the plan’s adoption, would have undermined the rationality of adopting that plan. We here understand the second clause to be violated when the agent is aware that she has failed to comply with the previously adopted plan.

It’s worth noting that perfect resoluteness (unconditional or rational) is compatible with *akrasia* (choosing something that one judges to be rationally impermissible). An agent can akratically adopt a rationally impermissible plan and still be perfectly resolute in her implementation of the plan. Indeed, perfect *unconditional* resoluteness requires such implementation, while perfect *rational* resoluteness is silent about such cases.

Although McClennen is not explicit about this, we assume that the resolute conception of rationality also requires agents to *adopt* plans whenever adoption is rationally required. For, if plans are never rationally adopted, the requirement to comply within rationally adopted plans never has any force. This is relevant, because the resolute theory of rational choice appeals to plans that were *actually rationally adopted*, and not to plans that it *would have been rational to adopt*. In this respect, the theory is like actual consent theories of political obligation, and unlike hypothetical consent theories thereof.

When, then, is it rational to adopt a given plan? The resolute conception of rational choice holds that it is rationally permissible to adopt a given feasible plan if and only if *full implementation* of the plan has prospects that are at least as good as those of any alternative feasible plan. We use feasibility in an objective sense and assume that the set of feasible options (abstractly possible choices for the agent in the situation) is the set of options that the agent could choose, if she were fully informed and had a suitable choice disposition. The resolute conception thus requires that one comply with any plan adopted, if (1) at the time of adoption, full implementation of that plan has, relative to the agent's evidence, prospects that are at least as good as those of full implementation of any alternative feasible plan, and, (2) since then, there has been no unanticipated new information that undermines the previous condition.[4]

We have two objections to this conception of rational choice. First, the rational permissibility of plan adoption is, we claim, based on a *realistic* assessment of how likely the agent is to fully comply with the plan. Appealing to the idealization of full compliance abstracts from crucial features of the choice situation. Few, if any, agents are perfectly disposed to comply with plans that they adopt. In our chocolate example, the best plan based on an *idealized* (full compliance) assessment is to eat the chocolate at the first, but not the second, time. Realistically, however, that is the worst plan to adopt, since it will not be fully executed and will lead to eating chocolate at both times (the worst outcome). We agree with McClennen that for agents with perfect self-control it would be rationally permissible to be resolute. So, we are not directly challenging McClennen's view about what plans should be adopted by agents with perfect self-control. Our point is that, in the general case, where agents may lack perfect self-control, this theory makes little sense. That is, our first objection is, not that McClennen's theory is false, but rather, that its scope is too narrow. The resolute choice conception is not applicable to the type of cases we care about: realistic choice situations in which the agent may lack perfect self-control.

Second, even if an agent is perfectly disposed to comply with whatever plan is adopted, conditional on the adopted plan being a best plan to adopt in the realistic sense, it is not always rational to *comply* with such a plan. In the well-known Toxin Example (Kavka 1983), it may be that adopting the plan to drink the sickness-inducing toxin is the realistically best plan to adopt (since its adoption leads to a reward), but, given that the reward is irrevocably given before drinking the toxin, it can be irrational to drink the toxin, if this is still feasible. The mere fact that it was rational to adopt the plan, and no unanticipated new information undermined the support for the plan, does not ensure that it is rational to comply with the plan.

It is important to note, however, that there is a certain kind of agent for which the resolute conception of rational choice gives correct answers. Suppose that (1) an agent has rational preferences that accord *lexical primacy* to conformance with adopted plans (or more weakly: when such plans have the best prospects under the assumption of full compliance), and (2) she suffers from no akrasia (i.e., her motivational preferences are always in accord with her rational preferences). Call such an agent *perfectly resolute in the evaluation-based sense* (since her resoluteness comes from her lexically primary rational preference to comply with adopted plans). The resolute conception of rationality gives the correct answers with respect to such an agent. For such an agent, a plan has the best prospects under realistic assumptions about compliance if and only if it has best prospects under idealized (perfect compliance) assumptions. Moreover, given her lexically primary rational preference for conformance with adopted plans, it is always better for her to comply with such plans than not to do so. So, for such an agent, the resolute conception of rational choice is correct. The problem, of course, is that no real agent is such an agent, and, indeed, this is a very special kind of idealized agent.

McClennen (1997, pp. 239-42) holds agents can be perfectly resolute, at least when the adopted plans are rationally adopted and not later undermined, even without any preference for compliance with (e.g., rationally) adopted plans. A mere act of the will (e.g., a rational commitment) can make the agent perfectly disposed to comply with (e.g., rationally) adopted plans. Call such agents *perfectly resolute in the commitment-based sense*. We find this type of resoluteness mysterious and psychologically unrealistic. Nonetheless, for such agents, as for the corresponding evaluation-based agents, a plan has the best prospects under idealized (full compliance) assumptions if and only if it has the best prospect under realistic assumptions. So, the resolute conception of rational choice correctly identifies the best plan of such agents to adopt. The problem is that it also entails that compliance with such plans is always rationally required. Given, however, that the agent need have no rational preferences for such compliance, it will sometimes say that compliance is rationally required when some alternative feasible action has better prospects (e.g., drinking the toxin is required, even when one is able not to do so). So, the resolute conception gives the wrong assessments even for perfectly resolute agents, when their resoluteness is commitment-based rather than evaluation-based.

5. Historical Separability and Normal-Form/Extensive-Form Coincidence

Before turning to the assessment of the sophisticated conception of rational choice, it will be useful to discuss briefly two main conditions at issue between the competing conceptions of rational choice, as identified by McClennen.

One condition is *Historical Separability* (McClennen 1990 p. 122), which requires that the rational permissibility, at a given node, of a plan for the future *not depend* on what the past was like.[5] In formulating this condition, McClennen implicitly assumes that agents are fully informed of all facts. If, however, agents are less than fully informed, and if rational permissibility is relative to the agent's beliefs (or at least the beliefs support by her evidence), then the historical separability would need to be reformulated to say that rational permissibility does not depends on the agent's beliefs about the past. For simplicity, we focus on fully informed agents and we thus ignore this complexity. The resolute conception violates this condition because it makes rational permissibility depend on what plans the agent has rationally adopted in the past.[6] Because we believe that rational preferences can be historical (e.g., one can prefer eating a chocolate bar to not eating one, when one has not recently eaten a bar, but have the opposite preference, if one has recently eaten a bar), we agree with McClennen that Historical Separability should be rejected. We further agree that there is nothing irrational about having a pro tanto preference for complying with non-undermined plans that were rationally adopted in the past. Indeed, we think that it need not be irrational (although it's strange) to have preferences that make compliance with previously adopted plans *lexically prior* to all other considerations. So, we agree with McClennen that rational choice need not satisfy Historical Separability (although perhaps for different reasons).

A second condition discussed by McClennen is the *Normal-Form/Extensive-Form Coincidence* condition (McClennen 1990, p.115). This requires that, in dynamic choice situations, a plan is rationally permissible, relative to a given choice node, when represented in *extensive form* (i.e., as a sequence of choices, as represented in a decision-tree) if and only if it is rationally permissible, relative to that node, when represented in *normal form* (i.e., as a *single choice* of feasible sequence of "choices"). As discussed above, this is plausible for agents that are *perfectly* resolute and confront no unanticipated undermining information, but (1) few, if any, agents are so resolute, and (2) the condition, we shall now argue, is implausible for agents that are not.

Consequently, the condition is implausible as a general condition of rational choice.

A plausible theory of dynamic choice should take into account all the information available to the agent at each choice node, and this includes information about how likely she is to make various choices at future choice nodes. The Normal-Form Extensive-Form Equivalence condition precludes such sensitivity, because it treats a sequence of choices as a single choice. For example, in our chocolate example, the best plan open to the akratic agent is to have chocolate on the first occasion but not on the second occasion. The resolute conception of rational choice then says that rationality requires adopting that plan and then complying with it. The agent will comply with the plan at the first choice (by eating chocolate), but, given her akrasia, she will fail to comply with the plan at the second choice node, and the result will be the worst possible result. Rational choice requires sensitivity to *how the agent is disposed to choose* at the various later choice nodes, but the resolute conception is insensitive to such information for agents that are not perfectly resolute.

The resolute conception of rational choice requires ignoring information about the agent's future choice dispositions. As we will show in the coming sections, this problem can be overcome by being sensitive to the probabilities of future choices.

6. Sophisticated Choice

On the sophisticated choice conception of rationality (e.g., Schick 1986), the agent should first predict her final choices and then reason backwards for earlier choices. For example, if she predicts that, due to akrasia, she will eat chocolate at the second time, even if she eats chocolate at the first time, she should then compare the outcome of not eating chocolate at the first time with the outcome of eating chocolate at the first time. Since the former (chocolate only at second time) is, we are assuming, rationally preferable to the latter (chocolate at both times), rational choice requires her not to eat chocolate at the first time. On this view, the agent predicts future choices and then deliberates about her current choices in light of those predictions. This seems roughly correct, but we shall now generalize this approach.

It is generally assumed (e.g., by Schick 1986, McClennen 1990, Rabinowicz 1995) that backwards induction is a necessary element of the sophisticated choice conception, but this is not so. Backwards induction is just a tool the sophisticated agent uses for *making decisions based on predictions about future choices*. In principle, any other tool that does the job can be used by the sophisticated agent. Moreover, backwards induction does not always work. First, if, for some branches, there is no final choice node, then there is no starting point for the backward induction. Second, even if all branches have final choice nodes, if there is more than one option that the agent might choose, and these are not equally valuable (e.g., where a less valuable option might be irrationally chosen), then backwards induction will not get started in this case either.

A more general approach for sophisticated agents is to assign *probabilities* to each of her choices in each future node. These probabilities define the *choice disposition* that the agent predicts she will have at that node. In the simple finite case where, at each choice node, there is only one choice that the agent might make with non-zero probability, backwards induction will be possible, but it is not essential. Instead, associated with each possible choice in the current choice situation, there will be a *prospect*, that is, a probability distribution over branches. The probabilities of the branches will be determined by the probabilities at future *chance nodes* and by the probabilities at future *choice nodes*. Rationally permissible choice requires that an agent choose a feasible option with a best (or at least maximally good) prospect. Here, we leave open what the criteria are for the goodness (relative to rational preferences) of prospects. In particular, we do not assume that they must be risk-neutral or based on expected values. We only assume that prospects are assessed rationally, whatever that requires.

The sophisticated conception of rational choice is thus best understood as requiring that the agent choose an option with rationally best prospects, where prospects are probability distributions over branches, and where the probabilities reflect both the probabilities of future chance events and the probabilities of future choices by the agent.[7]

For agents that are perfectly rationally resolute in the evaluation-based sense, the sophisticated conception agrees with resolute conception, as noted above. If, however, the agent is not perfectly rationally resolute, then conforming to the non-undermined rationally adopted plan may not be rational (given that she may not later comply with it). In our chocolate example, suppose that the agent is sure to eat chocolate at the second time, even if she adopts the plan not to do so (due to akrasia). In that case, the agent faces a choice between eating chocolate at the first time with 100% chance of eating chocolate at the later time and not eating chocolate at the first time with 100% chance of eating chocolate at the later time. Since the latter prospect is rationally better, the rational choice at the first time is to choose not to eat chocolate. Here the resolute conception wrongly entails that she is rationally required to eat chocolate at the first time.

Moreover, the sophisticated conception handles cases where the probabilities of future choices are not zero or one. Suppose that the agent is 50% likely to eat chocolate at the second time, if she eats chocolate at the first time, and she is 100% likely to eat chocolate at the second time, if she does not eat chocolate at the first time. Then rational permissibility of eating chocolate at the first time will depend on the how valuable the different outcomes are. Suppose, for example (see Figure 2 below), that eating chocolate twice has a value of 4 and eating chocolate only at the second time has a value 10. Then, assuming for illustration that the value of prospects is their expected value, (1) if the value of eating chocolate only at the first time is more than 16, then it will be rationally permissible to eat chocolate at the first time (with 50% chance of also eating at the second time), and (2) if that value is less than 16, then it will be rationally impermissible to do so.

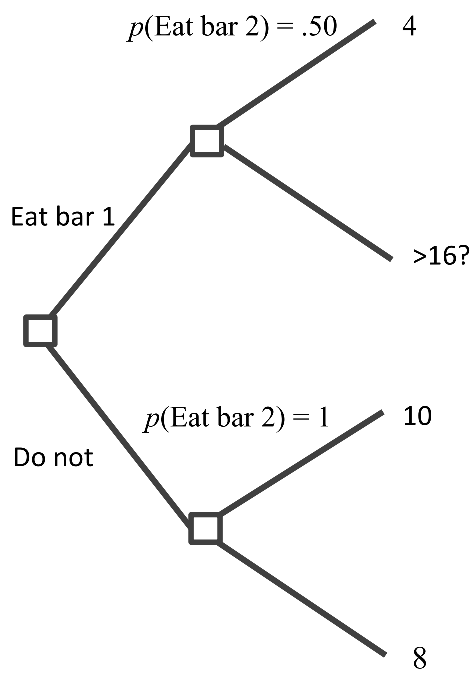

Figure 2. Chocolate: Probabilistic Case

The sophisticated conception of rational choice, as sketched above, is, we believe, essentially correct. A crucial assumption, however, is that agents can rationally assign probabilities to their future choices. In the next section, we defend this assumption. Following that, we address the issue of whether the sophisticated conception is committed to Historical Separability (the irrelevance of the past).

7. Can Probabilities Be Ascribed to Future Choices?

To render the notion of self-predictive probabilities sharp, we distinguish between the *feasibility* of a plan (or a choice) and the objective *probability* that the agent will execute it. As explained in section 4, the set of feasible options (abstractly possible choices for the agent in the situation) is the set of options that the agent could choose, if she were fully informed and had a suitable choice disposition. An agent's *choice-disposition* is the objective probability, for each feasible option, of the agent choosing that option. Feasible options can, in this sense, have zero probability, if there is no chance of the agent choosing it (even if she could). Thus, for example, if a perfectly rational and fully informed agent, with no akrasia, has a choice between a better option and a worse option, the worse option is feasible for her, even though there is no chance that she will choose it.[8] By contrast, choosing to run faster than the speed of light is not feasible for any agent. Obviously, an agent's future choice dispositions need not be fixed. She may be able to alter those dispositions over time. She may, for example, choose to engage in exercises that modify her preferences or reduce her weakness of will.

Agents, we have assumed, can ascribe subjective probabilities to their future choices (e.g., the chance of having a future choice disposition). Our claim is not that agents will be perfect predictors of their future choices. It is rather simply that they can rationally assign subjective probabilities to their future choices in the same way that they can rationally assign subjective probabilities to other future events.

The claim that rational agents can assign subjective probabilities to their future choices is somewhat controversial. Wolfgang Spohn (1977) and Isaac Levi (1989) argue, independently of each other, that a rational agent cannot at the same time deliberate and predict his own behavior. As Levi puts it, "deliberation crowds out prediction." We grant that it may not be possible to deliberate about one's *current* choice, while at the same time predicting what one's *current* choice will be. Fortunately, we do not require this. We hold that agents can, and should, *predict* their choices in *future choice situations*, and *deliberate* about their choices in the *current choice situation.* (For insightful discussion of this issue, see Rabinowicz 2002.)

Still, there is a potential problem, at least if we assume that an agent's probabilities should be reflected in the bets that she is willing to accept. An agent who ascribes subjective probabilities to her future choices must be prepared to accept certain bets on which act she will eventually choose. By accepting such bets, however, it becomes more attractive for the agent to perform the act on which she is betting. Suppose, for instance, that you are disposed to choose, at a future choice node, A with probability 0.6 and B with probability 0.4 and that you predict that you will choose these options with these probabilities. This entails that you should be willing to accept a bet in which, if you reach this choice node, you win 1 unit of value if you choose A and lose 0.60/0.40 = 1.5 units of value if you choose B. (The expected value of A is $1 \cdot .6 = .6$ and that of B is $-1.5 \cdot .4 = -.6$)

Why is this a problem? Once you have accepted the bet described above, you have an incentive to make sure you win the bet, if you reach the choice node. You can easily do this by simply choosing A, if you reach the node. So, once the bet is accepted, the probability of choosing A, should the choice node be reached, is 1.0. This is problematic, because we initially assumed that the agent was disposed to choose A with probability .6. That is, by merely predicting his own choice, the agent's choice disposition increases from .6 to 1. This suggests that self-prediction, even for future choices, is problematic.

Spohn and Levi spell out the technical details of this argument in somewhat different ways, but the key idea in both accounts is that the bets we use for eliciting self-predictive subjective probabilities) interfere with the entity being measured.

The following analogy might be helpful: If you measure the temperature in a hot cup of coffee with a very cold thermometer, then the temperature of the measurement instrument will significantly affect the temperature of the coffee. That is, your measurement instrument interferes with the entity being measured. The only way to make sure that this effect does not occur is to ensure in advance that the thermometer has exactly the same temperature as the coffee, but this is possible only if we know the temperature of the coffee before we measure, which makes the measurement process superfluous. According to a more radical, operationalist version of this argument, we should conclude that because we cannot measure the temperature of the coffee the theoretical term "temperature" has no meaning.

When bets are used for measuring probabilistic choice dispositions, the bets are by no means superfluous, but they influence the entity being measured. So, although the analogy with hot coffee is not perfect, the underlying phenomenon is similar, irrespective of which version of the objection one prefers. (Levi seems to endorse the operationalist version, according to which the impossibility to measure self-predictive probabilities makes the concept meaningless.)

However, in opposition to Spohn and Levi, we think this measurement theoretical effect is not a reason for giving up the idea that rational agents can make probabilistic self-predictions. One might, of course, question whether betting ratios are the proper way of measuring degrees of belief, but here we shall not question this. We shall instead indicate how interference between accepting bets and predictive acts can be controlled. First, the stakes of the bets can be set to be negligible compared to the difference in value of the relevant options. Thus, the agent will not be tempted to win her bet by selecting the less valuable option. This strategy works as long as the agent is not indifferent between all options.[9] Second, we can design the betting mechanism so that the agent sets the odds and the bettor sets the stakes, without the agent knowing whether she will win or lose the bet if the alternative she bets on is chosen. The information available to the agent will be sufficient for measuring her subjective beliefs about her choice disposition, but, because she does not know what choices will make her win or lose the bet, she has no incentive to adjust her choice disposition. (See Rabinowicz 2002 for a discussion of this point.)

The upshot is that Spohn and Levi are right that the agent's betting dispositions will sometimes affect the measurement process used for making accurate self-predictions, but the relevance of this observation should not be exaggerated. If we are aware of the problem, we can adjust the measurement process so that the problem they identify will not arise. The underlying phenomenon we wish to measure—the agent's subjective degree of belief that she makes a given choice—surely exists. So, this is just a matter of performing the measurement in the correct way.

8. Wise Choice

Above, we defended the manner in which the sophisticated conception of rational choice appeals to the *probabilities* of futures choices. This part of the theory is novel. In previous discussions of sophisticated choice, it has been assumed that the agent is able to predict *with certainty* her future choices, except perhaps where there are ties for maximal value. By allowing sophisticated agents to make probabilistic self-predictions, we make the theory applicable to a broader range of choice problems.

The sophisticated conception of sequential choice is often (e.g., McClennen 1990) defined as also satisfying a second feature: Historical Separability. This requires that the rational permissibility of a choice not depend on the history (e.g., what choices were made, or what plans were adopted) prior to that choice node. As indicated in our discussion of the resolute conception, we agree with McClennen that this is an unreasonable requirement. Rational choice can be sensitive to what the past was like in virtue of having rational preferences that are historical (e.g., rationally preferring coffee after a meal but scotch before a meal). Although the sophisticated conception is not always understood to require satisfaction of the separability condition, we will here accept McClennen's definition and therefore reject the sophisticated conception so understood.

Fortunately, there is already a name for the conception of rational of choice that (1) appeals to the agent's predictions about her future choices but (2) does not require satisfaction of Historical Separability. This is the *wise-choice* conception, as introduced and discussed by Rabinowicz (1995, 1997, 2000, 2017). It allows, but does not require, that rational choice may depend on what the past was like, and, in particular, on what plans were rationally adopted. It holds that rational choice in a given choice situation is a matter of choosing an option with best prospects in light of one's choice dispositions in the future. If one's rational preferences are non-historical, then this will be equivalent to the sophisticated conception. If one's rational preferences give lexical primacy to compliance with plans rationally adopted in the past, then, for perfectly rationally resolute agents, this will be equivalent to the resolute conception. Finally, if one's preferences give some finite weight to compliance with plans rationally adopted in the past, then the wise conception will, like the resolute conception, tend to favor compliance with such plans. It will not, however, always favor such compliance, since it will be but one of several competing considerations for what is best.

Like the sophisticated conception, the wise conception has been characterized as involving backwards induction on one's future choice (e.g., Rabinowicz 1995). Our discussion of sophisticated choice, however, makes clear that (1) backwards induction is not always possible, but (2) this is not a problem, since the core idea is captured by reasoning based on the probabilities of one's future choices. So, the wise conception, we believe, should be so understood. Wise agents believe that there is no fundamental difference between probabilities ascribed to one's own future choices and probabilities ascribed to future events.

Rabinowicz (2017) suggests that there is a difference between wise choice and sophisticated choice with respect to how they *predict* future choices when the agent expects her future motivational preferences to become distorted relative to her current rational preferences (e.g., Ulysses' s preferences as he sails close to the sirens, or the preferences of a person about to get drunk). Following Machina (1991), Rabinowicz holds that commitment to a historical separability condition (as in sophisticated choice) requires *predicting future* preferences on the basis of current unconditional preferences. McClennen's Historical Separability condition, however, does not require this. It does not address how to predict future choices. It merely asserts that the rational permissibility of a choice does not depend on what the past was like. Both the sophisticated and the wise-choice conceptions, as we understand them, judge the *rational permissibility* of current and future choices relative to the rational preferences at the time of the assessment (e.g., current rational preferences). Moreover, both hold that *predictions* of future choices are based on the best available evidence about future motivational preferences. So, on our understanding, sophisticated and wise choice differs only with respect to whether a preference for complying with rationally adopted plans can be rational.

The upshot of all this is that the probabilistic account of wise choice that we propose seems able to embrace the best parts of the sophisticated analysis (sensitivity to how the agent is likely to behave in the future) and the best parts of the resolute analysis (sensitivity to what plans were rationally adopted in the past, to the extent the agent's motivational preferences are so sensitive).

9. Two Objections

A standard objection to the sophisticated account is that it can judge a sequence of choices to be rationally permissible, even when dominated by some alternative feasible sequence of choices. This objection could be raised against the wise analysis too, if the agent's rational preferences are historically insensitive. Moreover, the objection applies even if the agents have no akrasia (as we will assume for illustration).

Consider, for example, Figure 3 and assume that (1) where no choice has a sure outcome, the agent rationally prefers the choice with the highest expected monetary payoff, and (2) the agent prefers \$1M for sure to a 10/11 chance of \$5M and 1/11 change of \$0. Sophisticated agents (i.e., non-historical wise agents) are rationally required to go Up at the second choice node (for the certainty of \$1M). At the first node, however, the agent is rationally required (and hence permitted) to choose Down—because (1) she predicts with certainty that she will choose Up, if she reaches the second choice node, (2) at the first choice, node neither Up nor Down has a certain prospect, and (3) Down has higher expected value than UpUp. This violates a standard dominance condition, since going UpDown is better for the agent than Down, no matter what chance events (E or F) occur.

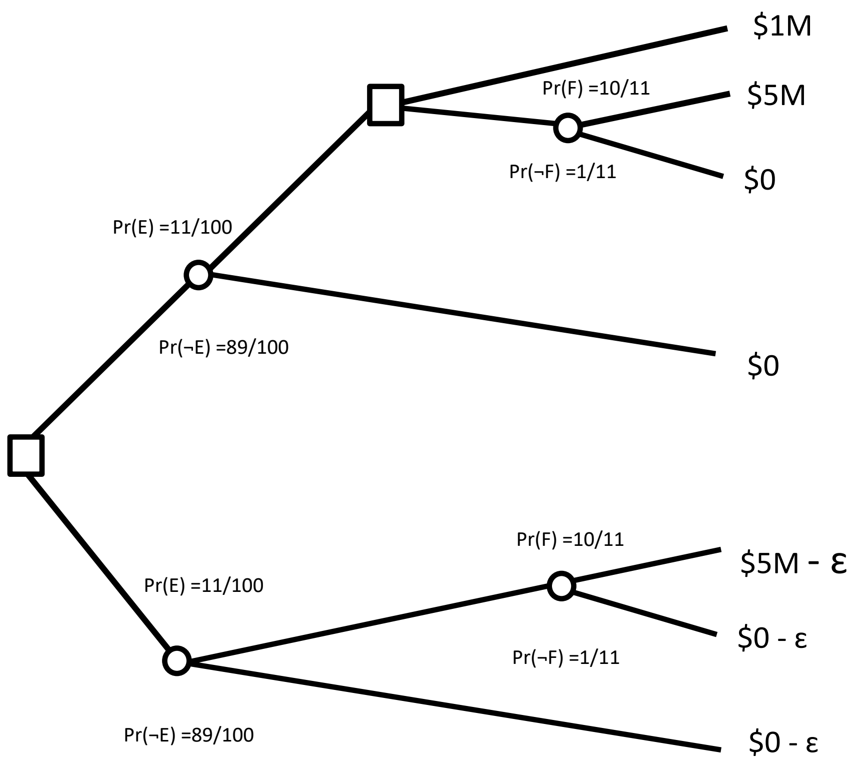

Figure 3 Alleged Failure of Dominance (After Rabinowicz 1995, 2017).

We see no reason to worry about this alleged violation of the dominance principle. For non-dynamic choice (with just one choice situation), the standard dominance principle is sound. For dynamic choice, however, one must distinguish between different dominance principles, only some of which are sound. In its most general sense, (weak) dominance occurs when (1) for all possible events, one option has outcomes that are at least as good as those of a second option, and (2) for some possible events, it has better outcomes. In the non-dynamic case, the only relevant events are the *chance events*, but this is not so in the dynamic case. One must also consider the *future choices* the agent might make. This gives rise to different dominance principles, only some of which are plausible. The above violation of dominance, we claim, is only a violation of an implausible dominance principle. Let us explain.

A very weak dynamic dominance principle holds that, if two options (choices or choice-sequences) are defined on the *same set of chance and future-choice events*, and one dominates the other relative to those events, then the latter is not a rationally permissible choice. This is extremely plausible. The above example, however, is not of this kind, since the choice of Down initially does not lead to the second choice node that the choice of Up leads to. The future choice-events generated by these two initial options are different.

One might strengthen the above dominance principle to apply in certain dynamic cases where future choice-events are not the same for the two current choices. For example, it might apply if, for each current choice, all future choice-events are *certain* (occur with probability one). This seems plausible, since it essentially reduces a dynamic choice to a non-dynamic one. However, this strengthened version of the principle is inapplicable to the above example. If the agent chooses Up at the first choice node, it is not certain that she will reach the second choice node, so the future choice is not certain. There is an intervening chance node.

In the example, comparing the value of outcomes of UpDown and Down for each shared chance event (E and F) would be fine, if the probability that the agent would go Down at the second choice node, given that she goes Up at the first, were one. Given that it is not (it is zero), the comparison incorrectly ignores the probability that the agent may not choose Down after Up. It is precisely because that probability is sufficiently high (and the associated expected value sufficiently lower) that it is rationally required to choose Down at the initial node. The type of dominance that occurs in this example is not normatively relevant.

A second objection to the wise account is that it appears to be *dynamically inconsistent*, in the sense that it can (1) judge a sequence of choices to be rationally *required* relative the first choice node of the sequence but (2) judge continuation of the sequence to be rationally *impermissible* relative to a later choice node of that sequence. For example, for any given decision tree, it is dynamically inconsistent to judge a plan, UpDown, to be a rationally required sequence relative to the starting choice node but to judge Down to be rationally impermissible relative to the second choice node reached after choosing Up at the first node. (See McClennen 1990 Ch. 7.2 and generally for more discussion.)

Consider, for example, Figure 4 below. We assume here that a prospect is better, relative to the agent's rational preferences, just in case it has higher expected monetary value. We further assume that the agent suffers from akrasia and has a disposition to take irrational risky gambles. Suppose that she predicts, at the first choice node, that, if she goes Up, she will go Down at the second node with probability .99 (because she is prone to take irrational gambles). The expected value of going up at the first node is therefore .01·\$3M + .99·1/2· \$5M = \$2.505M, which is more than the expected value of going down (\$2.5M). However, when the agent reaches the second node, the expected value of going Up is higher (\$3M is more than 1/2·\$5M). Thus, despite her previous prediction to go Down, she now decides to go Up.

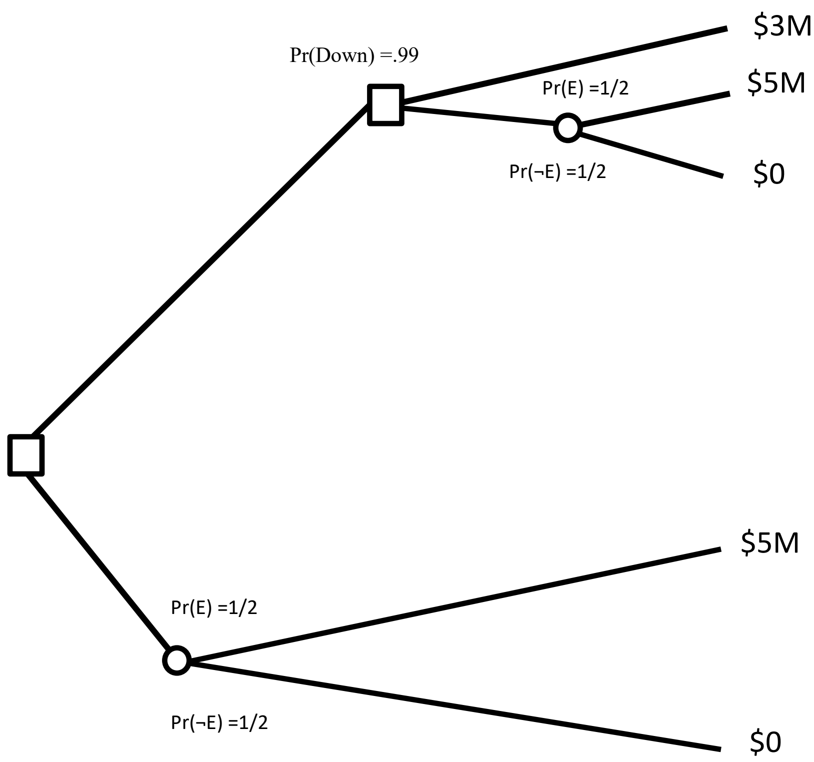

Figure 4. Alleged Dynamic Inconsistency

There is no dynamic inconsistency here. At the first node, the rational sequence of choices is UpUp, and, at the second node, the rational choice is Up. These sequences of choices are clearly consistent. There is of course an "inconsistency" between the *predicted* choice at the second node and the *rationally permissible* choice at the second node. This is, however, not a problem. This is precisely what we would expect for a wise agent predicting akratic behavior. An agent may predict akratic choice, and the rationality of her first choice be based on that prediction, but it's still true that the rational choice at the second node is Up. Stated otherwise, the wise agent predicts Down at the second node, but she is not endorsing the rationality of doing so. Hence, there is no violation of Dynamic Consistency.

10. Self-Control for Wise Agents

Before concluding, we briefly comment on the value of self-control (willpower) in the broad sense) for wise agents. Wise agents need not have perfect self-control in that they may suffer from akrasia or from a lack of resoluteness. We shall address each separately.

Akrasia (choosing an option that one judges rationally impermissible), of course, is irrational. The core question, however, is: What is rationally required of an agent who suffers from akrasia? We claim that this question is to be answered like any other question of rational choice. It all depends on the instrumental value of the various options. Sometimes it will be rationally required to spend resources on the development of willpower to overcome temptations (e.g., by doing various mental exercises). Sometimes, however, the costs will be too great and the rationally required response is simply to anticipate cases of akrasia and minimize the damage (as in not eating the first chocolate bar). We suspect that rarely, if ever, is it rationally permissible to *maximize* one's will power to overcome temptations. It's probably just too costly.

With respect to lack of resoluteness, our general assessment is the same, but there are a few additional complexities. As pointed out above, there is a distinction between *unconditional* resoluteness and *rational* (conditional) resoluteness. Unconditional resoluteness is a disposition to comply with adopted plans, no matter what. Rational resoluteness is a disposition to comply with plans that were rationally adopted and for which the rational permissibility of the adoption has not been undermined by new unanticipated information. It's rarely rationally permissible to be perfectly *unconditionally* resolute, since plans can be adopted irrationally and unanticipated information can undermine plans that were rationally adopted. Indeed, it's probably rarely rationally permissible to be *strongly* unconditionally resolute. The exceptions are bizarre cases where the world rewards such dispositions. Rational resoluteness, on the other hand, is, to the extent it is feasible, often rationally permissible. So, we shall focus on it.

Rational resoluteness comes in varying degrees, which can be understood in terms of the probability that the agent will comply with non-undermined rationally adopted plans in light of various benefits of non-compliance. *Perfect* rational resoluteness is the case where the agent is certain to so comply. In general, it is instrumentally valuable to develop one's rational resoluteness, since this helps one to overcome: (1) future akrasia; (2) future choices that are rational relative to one's future values but are irrational relative to one's present values, because (a) of anticipated changes in basic values (e.g., with age), or (b) no changes in basic values, but basic values, at a time, are (i) *insensitive to the past, but (ii) sensitive to the temporal order of future events*[10]; and (3) future choices that fail to adequately promote one's values because of a deterioration in the completeness or truth of one's beliefs (e.g. due to declining mental powers),

How, then, can one increase one's rational resolution?[11] One way is to increase one's capacity and disposition *not to reconsider* non-undermined rationally adopted plans, and thus allow them to be implemented. Second, one may wish to create, or strengthen, *a rational desire* to conformance with non-undermined rationally adopted plans. To the extent that one's motivating preferences match one rational preferences, one will be more inclined to be rationally resolute. Third, some have argued (mainly McClennen 1990, 1997) that mere act of "rational commitment" (an act of will) can make one rationally resolute, or at least increase such resoluteness.[12] We are skeptical of this possibility, but we here leave this open. No doubt, there are other possibilities.

If some degree of rational resoluteness is feasible for an agent, it will often be instrumentally rational to cultivate such resoluteness. Of course, it does not follow that *perfect* rational resoluteness is feasible for real agents. Indeed, we are skeptical, but we leave this too open. Moreover, even if it is feasible, it may not be rationally permissible to develop (because of the costs involved).

In short, on the wise conception of rational choice, the degree of rational resoluteness that an agent is rationally required to have will be determined by instrumental considerations: the feasibility of developing it, the costs and benefits of doing so, etc. Because we doubt that it will be feasible and instrumentally rational for any (current) human agent to become perfectly rationally resolute, we doubt that the resolute theory of rational choice applies to any (current) humans.

11. Conclusion

We have proposed a theory of rationally permissible sequential choice that covers both agents that have perfect self-control, those who have partial self-control, and those who have none. On this view, agents should ascribe probabilities to their own future choices (reflecting their dispositions for self-control) and at each point in time do whatever maximizes expected utility. This theory (1) agrees with the resolute conception of rational choice for agents that perfectly rationally resolute in the evaluation-based sense (of having a lexically primary rational preference for compliance with non-undermined rationally adopted plans), (2) agrees with the sophisticated conception of rational choice for agents that have non-historical motivational preferences (and hence no rational resoluteness), and (3) covers agents with imperfect rational resoluteness and/or akrasia.[13]

Bibliography

Bratman, Michael (2012). "Time, Rationality, and Self-Governance," *Philosophical Issues*, 22: 73-88.

Carlson, Erik (2003). "Dynamic Inconsistency and Performable Plans," *Philosophical Studies*, 113: 181-200.

Hammond, Peter J. (1988). "Consequentialist foundations for expected utility," *Theory and Decision*, 25: 25 - 78.

Holton, Richard (2009). *Willing, Wanting, Waiting* (Oxford University Press).

Kavka, Gregory (1983). "The Toxin Puzzle,*" Analysis,* 43: 33–36.

Levi, Isaac (1989). "Rationality, Prediction, and Autonomous Choice," *Canadian Journal of Philosophy*, 19 (1989): 339-362 [reprinted in Isaac Levi, *The Covenant of Reason* (Cambridge: Cambridge University Press, 1997), pp. 19-39].

Machina, Mark (1991). "Dynamic consistency and non-expected utility," in Michael Bacharach and Susan Hurley, eds., *Foundations of Decision Theory—Issues and Advances*, Blackwell, Oxford: 39 - 91.

McClennen, Edward F. (1990). *Rationality and Dynamic Choice* (Cambridge: Cambridge University Press).

McClennen, Edward F. (1997). "Pragmatic rationality and rules," *Philosophy and Public Affairs,* 26*:* 210-58.

Mele, Alfred R. (1992). *Irrationality: An Essay on Akrasia, Self-Deception, and Self-Control*. Oxford: Oxford University Press.

Peterson, Martin (2006). "Indeterminate preferences," *Philosophical studies*, 130: 297-320.

Rabinowicz, Wlodek (1995). "To Have One's Cake and Eat It, Too: Sequential Choices and Expected-Utility Violations," *The Journal of Philosophy* 92: 586-620.

Rabinowicz, Wlodek (1997). "Wise choice: On Dynamic decision-Making without Independence," in E. Ejerhed and S. Lindström, eds., *Logic, Action, and Cognition*, Kluwer, Dordrecht: 97-112.

Rabinowicz, Wlodek (2000). "Money Pump with Foresight" in *Imperceptible Harms and Benefits*, Michael J. Almeida, ed., Dordrecht: Kluwer Academic Publishers: 123-54.

Rabinowicz, Wlodek (2002). "Does Practical Deliberation Crowd Out Self-Prediction?" *Erkenntnis*, 57: 91-122.

Rabinowicz, Wlodek (2017). "Between Sophistication and Resolution–Wise Choice" forthcoming in *Handbook of Practical Reason*, Kurt Sylvan, ed., Routledge Press.

Schick, Frederic (1986). "Dutch Books and Money Pumps," *Journal of Philosophy*, 83: 112-19.

Spohn, Wolfgang (1977). "Where Luce and Krantz Do Really Generalize Savage's Decision Model," *Erkenntnis* 11: 113-134.

[1] Holton (2009) understands self-control and willpower narrowly to involve only the diachronic sense (resoluteness) and not the synchronic sense (lack of akrasia). In this chapter, we use the terms broadly to cover both and make no claim about what the most common usage of this term might be. Also, we understand akrasia as making a choice that the agent judges to be rationally impermissible (or not sufficiently good) relative to specified values of the agent (moral, prudential, or all relevant values). Standard conception of akrasia appeal to all relevant values of the agent (practical reasons).

[2] We shall not discuss the myopic conception of rational choice, since it is rejected by all.

[3] Contingency plans do not specify how to choose, should the agent fail to comply with the plan at earlier choice nodes. A more general approach would focus on *strategies*, which tell the agent how to choose at every possible choice. For simplicity, we focus on contingency plans.

[4] McClennen (1990, Ch. 12.6) further suggests that resoluteness is feasible only where all temporal selves (of the agent) can reasonably expected to benefit from resoluteness. For simplicity, we ignore this issue.

[5] McClennen (1990) calls this condition "Separability", but we label it "Historical Separability" in order to make explicit that the issue is separability *over times*, as opposed to over people or states of nature. Also, here and below, we give intuitive versions of McClennen's conditions, but these are meant to capture McClennen's more formal formulations.

[6] A fully adequate decision-tree representation of plan adoption would have to include choice nodes for the choices to adopt plans, as well as "regular" choice nodes (see Rabinowicz 2017 for discussion). For simplicity, we ignore this. Also, note that, for historically-sensitive rational preferences, the ranking of branches will be node-relative *and not fixed for the tree*. In our chocolate example, for example, the branch of eating a chocolate bar at time two is ranked higher than the branch of not eating one, if the agent did not eat a bar a time one, but the opposite ranking holds if she did then eat a bar.

[7] See McClennen (1990) for discussion of prospects and of probabilistic choice.

[8] We assume that choices are mental acts and that the physical acts, such as pushing a button, are outcomes of mental acts. Thus, where (as is typical) agents have imperfect control of their physical actions, the chance of the physical action being executed, given that it is chosen, will be less than one. We treat this simply as one of the many ways in which outcomes may not fully

determined by the choice of the agent.

[9] See Peterson (2006) for a detailed discussion.

[10] For example, suppose that (1) one rationally prefers a future with eating chocolate followed by drinking coffee to one with the reverse order and prefers the latter to coffee at both times, but (2) if there is only one consumption event in the future, one rationally prefers eating chocolate to drinking coffee. One might now rationally decide to drink coffee first because one anticipates that one will later rationally decide to eat chocolate (since the past will be irrelevant at that point).

[11] See Holton (2009) and Bratman (2012) for insightful discussion of how resoluteness might be developed.

[12] Rational commitment is distinct from precommitment, which involves making some future choice infeasible (e.g., Ulysses tying himself to the mast) or altering the *external* costs of benefits of various choices (e.g., signing a contract to pay large sums of money if one smokes). Rational commitment leaves feasibility and external costs and benefits in place but is somehow supposed to be capable of modifying choice behavior.

[13] For helpful comments, we thank Wlodek Rabinowicz, Johanna Thoma, and Paul Weirich.