Rational Commitment and Legal Reason

BRUCE CHAPMAN*

TABLE OF CONTENTS

I. INTRODUCTION

Is it rational to do something that you have no reason to do? Let us press the point: Could it be rational to do something that, on balance, you have reason *not* to do? On the view that practical rationality simply *is* acting for reasons, this would appear to be impossible. If there is no space between what you ought rationally to do and what reasons tell you to do, then the possibility of acting rationally, but contrary to the balance of reasons, is closed off. Thus, John Gardner and Timothy Macklem conclude their recent analysis of this topic: "rationality . . . is simply the capacity and propensity to act (think, feel, etc.) *only and always* for undefeated reasons."[1]

---

* Faculty of Law, University of Toronto, bruce.chapman@utoronto.ca. The Author is grateful to Shachar Lifshitz, Joe Mintoff, Oren Perez, and Wlodek Rabinowicz for comments on an earlier draft.

1. John Gardner & Timothy Macklem, *Reasons*, *in* THE OXFORD HANDBOOK OF

The theory of rational choice also seems to have this view about how reasons relate, structurally, to rational conduct, although it is not a theory that devotes much effort to analyzing reasons as such. According to rational choice theory, reasons (which, it should be emphasized, may be self-interested or other-regarding, consequentialist or deontological, objective or subjective) ultimately give rise to a preference for doing *x* rather than *y*, and rational choice consists in following that preference. It would be irrational, in other words, to act contrary to a preference, or contrary to the reason that lies behind it. Now, this idea does commit the rational choice theorist to requiring something else, namely, that the preference relation, which is essentially binary, satisfy certain minimal consistency conditions when more than two alternative choices are involved. For example, the preference relation must be transitive, or at least not cyclical.[2] For if an agent, for whatever reason, preferred *x* to *y*, *y* to *z*, and *z* to *x*, then it would not be possible for the agent to choose any of these three alternatives without choosing contrary to some preference or the requirements of some reason. Thus, the basis for imposing this formal condition of rationality, one that appears to connect different possible choices, is really only to meet the same fundamental concern identified by many theorists as essential to rationality, namely, that in every choice an agent must act only and always for undefeated reasons.

However, if practical rationality consists of something more than acting for reasons, then it might be possible, rationally, to do something that you have no reason to do, or even that you have reason not to do. Suppose, for example, that practical rationality, at least in part, consisted of doing what "makes sense," a point recently suggested by David Velleman.[3] An action that does not make sense certainly looks like it

JURISPRUDENCE AND PHILOSOPHY OF LAW 440, 474 (Jules Coleman & Scott Shapiro eds., 2002) (emphasis added). We should be careful in our interpretation of this summary that Gardner and Macklem provide of their position on rationality and reasons. For example, for them, rationality goes only to a general capacity, not a particular action. Thus, it may not be that an action itself must accord with reasons if it is to be rational. Further, in their view, reason-based choice need not be choice *consciously* guided by reason. They provide examples of this being counterproductive. Nevertheless, theirs is an analysis of rationality that does put reasons at the center. For another prominent theorist of rational decisionmaking who seems to collapse rationality into action according to reasons, see JOSEPH RAZ, ENGAGING REASON 1 (1999) ("Being rational is being capable of acting intentionally, that is, for reasons . . . ."). *See also id.* at 68 ("An account of rationality is an account of the capacity to perceive reasons and to conform to them . . . ."). Of course, Raz is also well known for allowing the possibility that rational choice can be choice when certain (sorts of) reasons for action are excluded. See his discussion of exclusionary reasons in JOSEPH RAZ, PRACTICAL REASON AND NORMS 35–48 (1975).

2. AMARTYA K. SEN, COLLECTIVE CHOICE AND SOCIAL WELFARE 16 (1970).

3. *See* J. David Velleman, *The Self as Narrator*, *available at* http://www-personal.umich.edu/~velleman/Work/Dennett.pdf (last visited Mar. 2, 2005); J. David

might be a promising candidate for irrationality. Velleman seems to have in mind the idea that some actions might make less sense than others (or not make sense at all) for an agent because they are less coherent with other actions that the agent has already chosen. The agent's life, or at least this part of the agent's life, would hang together less well as a coherent narrative for the agent if these misaligned actions were now the ones that the agent chose to perform. Of course, independent of the prior narrative, there may be every reason to do these misaligned actions, and no reason not to. But it is Velleman's view that the agent, as narrator of a coherent life, will—sometimes, at least—feel the rational pull of these prior actions. Therefore, it seems possible that, at these later moments of choice, an agent could rationally do what she has no reason to do, and even, perhaps, what (on balance) she has reason not to do.

It will be objected, of course, that the agent's prior decisions and choices simply provide reasons for the agent to carry on in a way that is coherent with them. Thus, Velleman's account of practical rationality is not at all inconsistent with the account that reduces rationality to acting for (undefeated) reasons. But, as some recent work by John Broome makes clear,[4] this objection confuses reasons with the normative requirements of practical rationality. Unlike reasons, the normative requirements of practical rationality do not detach from the elements they hold together (for example, a series of decisions). Thus, they do not give you reason to have any one of those elements (or make any one decision) in particular. They only require that *if* you have one of those elements, *then* you must have some other one on pain of irrationality if you do not.

This difference between reasons and the normative requirements of

Velleman, *Narrative Explanation*, *available at* http://www-personal.umich.edu/~velleman/Work/narrative.pdf (last visited Mar. 2, 2005).

4. John Broome, *Normative Requirements*, 12 RATIO 398 (1999) [hereinafter Broome, *Requirements*]; John Broome, *Are Intentions Reasons? And How Should We Cope with Incommensurable Values?*, *in* PRACTICAL RATIONALITY AND PREFERENCE 98, 98–120 (Christopher W. Morris & Arthur Ripstein eds., 2001) [hereinafter Broome, *Are Intentions Reasons?*]; John Broome & Christian Piller, *Normative Practical Reasoning*, *in* Supp. 75 THE ARISTOTELIAN SOCIETY 175, 175–93 (2001) [hereinafter Broome, *Normative*]; John Broome, *Practical Reasoning*, *in* REASON AND NATURE 85, 85–111 (José Luis Bermúdez & Alan Millar eds., 2002) [hereinafter Broome, *Practical Reasoning*]; John Broome, *Reasons*, *in* REASON AND VALUE: ESSAYS ON THE MORAL PHILOSOPHY OF JOSEPH RAZ (R. Jay Wallace et al. eds., forthcoming 2005) [hereinafter Broome, *Reasons*].

practical rationality is crucially important for what practical rationality can achieve. For only if the different reasons for action can be separated from each other by something that is itself not a reason, but is nevertheless a normative requirement of practical rationality, will it be rationally possible for an agent to follow through on rational commitments, rationally made. This, we shall see, is a source of great advantage for an agent, although securing this advantage cannot be the agent's reason for action. Thus, the agent must secure the advantage rationally, but without reason.

This Article develops this argument more fully as follows. Part II begins by distinguishing reasons from reasoning, and introduces the possibility of having a reason to choose to do something that you have reason not to do. This possibility is related to some quite conventional problems that the rational choice theorist faces in the theory of rational commitment. As we shall see, these problems arise because the rational choice theorist reduces practical rationality to action according to reason. Part III argues that practical rationality, in addition to requiring that action accord with reasons, also requires that action meet certain normative requirements and outlines the logical difference between the two. It argues that the special conceptual space that is occupied by normative requirements prevents the different reasons that animate distinct moments of decisionmaking from collapsing into one another to the disadvantage of the agent. Again, the analysis, at this point, is related to the special difficulty of rational commitment that confronts the rational choice theorist. In Part IV, I argue that the more robust model of rational commitment that is made possible by the idea of normative requirements of practical rationality should be familiar to legal theorists. After all, it is an idea manifested constantly in common law decisionmaking, where defeasible legal rules both determine cases (as a matter of normative requirement) and are determined by them (as a matter of reason), apparently simultaneously. Thus, the distinction between reasons and normative requirements of practical rationality can be used both to prescribe a solution for a problem in rational choice—namely, the problem of rational commitment—and to provide structural understanding for what is rational in legal reason and the method of common law adjudication. Part V provides some concluding remarks.

II. REASONING, REASONS, AND RATIONAL COMMITMENTS

All reasoning starts from an existing state of mind and concludes in a new one.[5] Theoretical reasoning, for example, takes us from one beginning

5. Broome, *Practical Reasoning*, *supra* note 4, at 110.

state of belief to another. If you begin by believing that Frankfurt is in Germany and Germany is in Europe, then theoretical reason would have you conclude by believing that Frankfurt is in Europe.

Practical reasoning is said to differ from theoretical reasoning in that, while it might proceed partially by way of beliefs, it concludes in an action rather than a belief. But that is not quite right.[6] An action (at least a physical action) requires physical ability as well as the ability to reason. More generally, we might say that an action requires opportunity. Thus, the most that practical reasoning can do is to take us from some existing state of mind to a *decision* or an *intention* to act—that is, another state of mind, albeit not one of mere belief. The action itself is something that carries out that decision or intention and lies beyond what reasoning alone can do for us.

This separation between what we decide or intend to do and what we actually do seems to allow for the following interesting question: Can we have reason to decide or intend to do something that we have no reason actually to do? Notice that this is not the same question with which this Article began. There the contrast was between what rationality and reasons might demand of us; here it is between what reasons themselves might demand of us at the two moments in a decision process that are opened up by the possibility that practical reasoning can only conclude in a state of mind (say, an intention) that falls short of an action.

It should be clear to any rational choice theorist who has anguished over the problem of rational commitment that we can have reasons to decide or intend to do something that we have no reason actually to do. More strongly, it seems that we can have reasons to decide or intend to do something that we have reason *not* to do. Indeed, more specifically, we can have *reason R* to decide to do something that we have *reason R not* to do. In other words, the *same* reason *R* can provide a rational basis both for choosing to do *x* and for not actually doing it.

A familiar and problematic example includes my promising someone to do *x* in exchange for that person doing *y* (where my promise is sincere at the time I make it) and yet, just as rationally, not doing *x* when the time comes to execute on the promise after the other party has done *y*. The reason for me to promise to do *x* is that I am better off with *y* being done (even after I incur the costs of doing *x*), and my promise helps to accomplish that; the reason not to do *x* is that I am (again) better off not

6. *Id.* at 85.

doing something that it is costly for me to do if there is no further benefit to be secured by actually doing it. Similarly, there can be self-interested reasons for me to threaten to do *x* should others do *y*, but (again) the same self-interested reasons not to actually do *x* when (despite my threat) they do *y*. Again, the reason to make the threat is that I am better off if they do not do *y*, and my threat helps to accomplish that; but I may be better off actually not to carry out my threat if they do *y*.

This much points to a kind of dynamic inconsistency[7] in the reasons that we can have at different moments within a single decision process. But more striking than the inconsistency itself is the manner in which the rational choice theorist resolves the inconsistency. An inconsistency between two apparently conflicting reasons can be resolved by relaxing the force of one or the other, and nothing in the argument so far points us to any particular resolution in this respect. However, as we shall now see, the rational choice theorist is inclined to give a priority to the reasons that agents have for particular choices over the reasons they might have had for a broader set of (more categorical) commitments.[8] This, I suggest, follows from combining (1) the idea, familiar by now, that rational action is action according to reasons with (2) a mode of reasoning that is essentially inductive—that is, that begins with the rationality of a particular choice and, with that choice in place, goes on to build an understanding of more general rationality requirements as an aggregation of similar such choices. What is rational for the general category, therefore, is built out of what is rational for the particular case. This inductive buildup from the particular to the general has the effect, I will argue, of displacing from our understanding of practical rationality anything that is different from, and which *begins* at a more general level than, the rationality of a particular action as one done in accordance with a reason. In other words, inductive reasoning helps to fill in *all* of practical rationality with action done according to particular reasons. Thus, it displaces from practical rationality the very possibility of having anything that is conceptually distinct from reasons, such as normative requirements.

To see how this works, consider again the example of promising. As we have seen, the rational choice theorist is committed to the general idea that practical rationality consists in acting according to reasons. Of course, for

7. R.H. Strotz, *Myopia and Inconsistency in Dynamic Utility Maximization*, 23 REV. OF ECON. STUD. 165–80 (1955–1956); Peter J. Hammond, *Changing Tastes and Coherent Dynamic Choice*, 43 REV. OF ECON. STUD. 159–73 (1976). For a philosopher's discussion, see generally EDWARD F. MCCLENNEN, RATIONALITY AND DYNAMIC CHOICE (1990).

8. For further discussion of the "particularity" of rational choice theory, and how this approach contrasts with the more "categorical" approach that characterizes an alternative tradition of rationality, see Bruce Chapman, *Rational Choice and Categorical Reason*, 151 U. PA. L. REV. 1169, 1169–210 (2002).

the rational choice theorist, reasons come mediated by preferences, and preferences must be transitive (or at least acyclic) if the idea of acting according to reasons is to be generally realized. But this does not alter the basic point that rational choice cannot be choice contrary to an undefeated reason. Thus, in the context of promises, it cannot be rational to carry out the promise if to do so at the time is contrary to preference or reason. Now, this much allows us to know what the promisor might *do* having already made the promise. But the rational choice theorist also has a way of determining what the promisor will *choose to do* at the prior moment when it is possible to make the promise. As this situation has an interpersonal aspect, this argument appears to require that the rational choice theorist make a fairly sophisticated assumption about what the promisor knows or believes about the promisee and, more particularly, about what the promisor knows or believes about what the promisee knows or believes about the promisor. Specifically, in these situations, the rational choice theorist assumes not only that all agents are rational, in the sense that they choose according to preference (and reason), but also that there is common knowledge of this rationality—namely, that each player knows that each is rational, and, further, that each knows that each knows this, and that each knows that each knows that each knows this, and so on. Exactly how sophisticated this last assumption is, and how, by way of induction, it helps the rational choice theorist to resolve the problem of dynamic inconsistency, can better be seen with the help of a more detailed example.

Imagine the following situation, which rational choice theorists commonly refer to as a "centipede game."[9] The bank has put out one hundred coins on a table. Two players, Art and Bart, are to take turns removing either one or two coins from the table, each keeping all the

9. This game seems to originate with Robert W. Rosenthal, *Games of Perfect Information, Predatory Pricing and the Chain-Store Paradox*, 25 J. ECON. THEORY 92, 92–100 (1981). For a detailed discussion of the game which reproduces some of the analysis presented here, see Bruce Chapman, *Legal Analysis of Economics: Solving the Problem of Rational Commitment*, 79 CHI.-KENT L. REV. 471, 475–82 (2004). For other recent discussion, see Robert J. Aumann, Note, *On the Centipede Game*, 23 GAMES & ECON. BEHAV. 97, 97–105 (1998); John Broome & Wlodek Rabinowicz, *Backwards Induction in the Centipede Game*, 59 ANALYSIS 237, 237–42 (1999); Wlodek Rabinowicz, *Grappling with the Centipede: Defence of Backward Induction for Bi-Terminating Games*, 14 ECON. & PHIL. 95, 95–126 (1998). The term "centipede" is used because when the game is represented as a decision tree (in extensive form), the tree consists of a long horizontal line segment (representing the players moving through the game as they take only one coin) with many short downward lines (representing the player taking two coins and ending the game at that point), i.e., a picture of a long centipede with many short legs.

coins that he removes. The game stops as soon as either player removes two coins, and at that point all the coins (and only those coins) still remaining on the table are returned to the bank. However, so long as each player takes only one coin, the game continues until all the coins are removed. Potentially, therefore, each player could take one coin at each turn and end up with fifty coins.

We are to imagine that Art and Bart are both rational in the sense that each wants to maximize his own monetary payoff from playing the game. Thus, each will not choose an option, or develop a strategy, if there is some other option or strategy that he could choose that will give him more money. Moreover, this rationality is common knowledge in the game in the way described above.

Suppose Art has the first move. The rational choice theorist's standard argument, based on backwards induction, is that Art will rationally choose to take two coins and the game will end. Of course, this seems a little problematic, even for Art; he might like to think that the game could have gone on a little longer so that he (and, incidentally, Bart too) could have picked up a few more of the one hundred coins that were available. But, unfortunately, that thought has no survival value under the assumptions of rationality and common knowledge of rationality.

To see why, imagine Art thinking ahead to where there are only two coins left on the table. This means that, up to this point in the game, each player has taken only one coin and has forty-nine in his possession. But now Art can either end the game by taking the two coins that remain or take only one and allow the game to end with Bart taking the only coin that is left. Clearly, the first option provides a higher payoff for Art and, therefore, is the rational option for him. So that is the option he chooses on this play of the game and the game ends.

But now consider Bart thinking ahead to where there are three coins on the table—that is, to the penultimate play in the game just before the one imagined by Art in the previous paragraph.[10] Since, under the assumption of common knowledge of rationality, Bart knows that Art is rational, he knows what Art will do in the next play of the game should Bart choose only one coin and the game move on to that next (and ultimate) stage. But Bart can do better than that by taking two coins at this penultimate play, thereby stopping the game. So, being rational, that is what Bart chooses to do.[11]

10. Or, more accurately, consider Art thinking this about Bart. For all of this thinking is really going to an explanation of why Art, who has the first move in the game, will choose to take two coins on the first move. So it is really a question of what Art is thinking about what Bart is thinking (about what Art is thinking, etc.). All this is made possible by common knowledge of rationality.

11. There is of course a problem here that more than a few commentators have

Now, of course, this last choice by Bart is perfectly predictable by Art (again, given common knowledge of rationality), and so Art will anticipate at the pre-penultimate play of the game, when there are four coins still on the table, that Bart will end the game at the next penultimate play. So, given that he is rational, Art will choose to do better by taking two coins rather than one at this pre-penultimate point, thus ending the game. And so on. We must conclude, therefore, that under this sort of inductive reasoning, and these assumptions, the game will end on the first play when Art takes two coins, leaving the other ninety-eight to be returned to the bank.

Does anything change if each player, at the beginning of the game, promises the other to only take one coin throughout the game? We can certainly see that each player has a reason for wishing that he could make such a sincere and credible promise (i.e., a promise that the other player could rationally believe). After all, each would be so much better off if, by promising, each could induce the other to behave according to their respective promises; each would have fifty coins rather than Art having two and Bart having none. But the backwards induction argument, based on the assumptions that each player is rational and that this rationality is common knowledge, prevents the promise from being

noticed. For the players to reach this point in the game, where there are only three coins remaining on the table, each player must have chosen not to terminate the game; that is, each must have chosen to remove only one coin at all the prior turns. But, as the backwards induction argument goes on to show (under the assumptions of rationality and common knowledge of rationality, assumptions that the players themselves can use to generate the argument), to remove only one coin on one's turn is not rational. Thus, at the point when there are only three remaining coins on the table, for Bart to hypothesize that Art will remove two coins on his next move (should Bart take only one coin and allow the game to continue on to that next move) is for Bart to hypothesize that Art is rational on this next move even though, also by hypothesis, Art has shown no such rationality in the game so far. Is it plausible for Bart to have, or to hypothesize having, such a resilient (i.e., contrary to fact) belief in Art's rationality? Indeed, is it plausible for Bart to anticipate acting rationally on his own turn having himself acted irrationally in the game so far? More generally, is it plausible to argue or hypothesize, at *any* turn in the game, that the player (whose turn it is) will either act rationally at this turn, or believe the other player will act rationally on the next turn, if this turn could not have been reached except through irrational play either by himself or the other player (or both) at some point earlier in the game? For good discussion of this difficulty in the backwards induction argument, see Philip Pettit & Robert Sugden, *The Backward Induction Paradox*, 86 J. PHIL. 169, 169–82 (1989). For a reconstruction of the argument that cleverly avoids this problem, at least at a formal level (by building in the assumption that all players believe that any turn in the game, *if it is actually reached*, must have been reached only by way of rational choices), see Broome & Rabinowicz, *supra* note 9, at 238–39. Also, for a similar argument, see Aumann, *supra* note 9, at 103.

credible. Each player knows, under these assumptions and regardless of what has been promised by the other player, that it is rational for the other player to end the game on the next move should he, himself, choose, according to his promise, not to end the game by only taking one coin. Thus, it is pointless for each player to believe the other's promise, and just as pointless, therefore, for each player to make it.

The rational choice theorist, therefore, resolves the dynamic inconsistency of having a reason to make a sincere promise that one has no reason to perform by denying the feasibility of making such a promise at all.[12] Without any real opportunity to make such a promise, there is nothing to which reason can attach. And, therefore, there is nothing to which a reason for not performing the promise can attach either. Now, by obliterating *both* of the inconsistent elements that make it up, this might appear more to *dissolve* the possibility of dynamic inconsistency altogether, rather than resolve it in some particular way by privileging *one* of the inconsistent elements. But it is clear from seeing how the backwards induction argument actually works that the dissolution is driven by (1) beginning with the reason that attaches to removing two coins at any particular choice in the game, particularly the last possible choice, (2) holding constant to the rationality of that choice (and each player's knowledge of its rationality) as one considers other prior choices, and then (3) generalizing to all prior (like) choices the same rationality that requires the agent to choose according to reason on the last possible choice. Thus, the dissolution of the dynamic inconsistency is clearly based on privileging the reason that attaches to not performing the promise, showing then (under the common knowledge assumption) that the prior making of the promise is without reason, *and*, *only then*, showing that there is nothing to which the reason for nonperformance can actually attach. This effectively resolves the dynamic inconsistency by privileging the reason not to perform the promise over the reason that one originally had to make it.

Because, under the assumptions of rationality and common knowledge of rationality, the players do so much worse for themselves as compared to how they might otherwise have done, the backwards induction argument has been thought to be somewhat paradoxical.[13] Apparently acceptable assumptions, combined with an apparently acceptable argument, have led us to an apparently unacceptable conclusion in terms

12. MCCLENNEN, *supra* note 7, at 200–18.

13. Even prominent game theorists concede this. *See* Reinhard Selten, *The Chain Store Paradox*, 9 THEORY & DECISION 127, 138 (1978) (arguing that the backwards induction argument provides a "game theoretically correct" answer for how rationally to play the game, but conceding that other ways of playing seem "to be the better guide to practical behavior").

of the payoffs that individually rational players secure. Moreover, there seems every reason to think that, in fact, two rational players in such a game would not actually play the game in the way that the backwards induction argument suggests. To accommodate this last point, the rational choice theorist's typical response has been to change the common knowledge of rationality assumption.[14] That is, we will see the players play this game longer, and more profitably, the argument goes, because it cannot be assumed that each player knows that each player is rational, or that each knows that each knows that each is rational, etc. This change in the common knowledge of rationality assumption will allow Art (or Bart) to entertain, at least, the thought that at some point in the game he should take only one coin because the other player will not necessarily respond by taking two coins and end the game in the next round of play. Predicting at what point exactly the game might end depends on the precise details of how the common knowledge assumption is relaxed, and need not detain us here. The important point is that a relaxation of this assumption allows us to comprehend the thought that the players might play the game more profitably than they do under the strictest version of the backwards induction argument that is implied by assuming common knowledge of rationality.

Moreover, as an empirical matter, it does seem implausible to think that the players would actually have common knowledge of rationality. After all, such an assumption requires each player to know a great deal about the other player's rationality and, further, about the other player's knowledge about one's own rationality. Indeed, it requires a player to know about the other player's knowledge about one's own knowledge about that player's rationality (and so on)! As the demands of common knowledge grow through these different levels, the assumption that there could actually be the sort of interpersonal transparency that is required seems more and more strained. And so, it seems reasonable to the rational choice theorist to relax the common knowledge of rationality assumption.[15]

---

14. *See, e.g.*, David M. Kreps et al., *Rational Cooperation in the Finitely Repeated Prisoners' Dilemma*, 27 J. ECON. THEORY 245, 246 (1982) ("[W]e are able to show that certain kinds of informational asymmetries *must* yield a significant measure of cooperation in equilibrium, and that other plausible asymmetries may produce cooperation as well.").

15. The economist seeks to relax the common knowledge assumption to explain the fact that players do not play the game in the way that the backwards induction argument suggests. Thus, while the argument might not apply as a contingent matter of

But I want to suggest now that the backwards induction argument does not depend so essentially on this sort of interpersonal knowledge. The argument, for all intents and purposes, will go through just as well if an agent is only required to have a sound knowledge of *his own* rationality and, in particular, if it is assumed that an agent knows that he cannot rationally intend or plan to do what he knows he will not rationally do (when the occasion arrives for him to act on that intention). To see this, consider the following variation on the centipede game.[16] Suppose Perfectly Reliable Bart makes the following offer to Art: that at any point *n* in the game where it is Bart's turn he (Bart) will take only one coin so long as Art can form the intention *at n* to take only one coin on the next play of the game, *n+1*, when it is Art's turn. Bart is assumed here to be perfectly reliable in the sense that he always takes one coin at *n* on observing that Art has formed the requisite intention at *n*. Thus, there is no question here of Art having to make any difficult assumptions about Bart's rationality, let alone any higher level assumptions about Bart's knowledge of Art's rationality or, further, about Bart's knowledge about Art's knowledge about Bart's rationality. And, likewise, Bart does not need to know any of this about Art, although, for the purposes of the argument, Bart does need to be able to observe Art's intentions at any play of the game.[17]

Consider again the problem from Art's point of view. A new offer from Bart is only worthwhile to Art if he can form the requisite intention at that point to take only one coin on the next play of the game. But, at Art's last possible move in the game, when there are only two coins left on the table, Art knows he will take both of them (after all, there is no possibility at this point of getting any new offers from Bart, and rational behavior, we assume, consists of maximizing one's monetary payoff).

---

fact, it is not as if they think there is any thing problematic with the argument, as such. Philosophers confronting the backwards induction argument are more inclined to think that there is something necessarily (not just contingently) wrong with the argument itself. Graham Priest has also noted that there is this difference in approach between philosophers and game theorists more generally. Graham Priest, *The Logic of Backwards Inductions*, 16 ECON. & PHIL. 267, 267–68 (2000). For a good review of the broad range of philosophical arguments dealing with backwards induction, most of them dealing with so-called surprise exam paradox, see generally ROY A. SORENSEN, BLINDSPOTS (1988).

16. This is a version of the variation introduced by SORENSEN, *supra* note 15, at 337. It builds on a problem about intentions introduced by Gregory Kavka. Gregory S. Kavka, *The Toxin Puzzle*, 43 ANALYSIS 33, 33–36 (1983).

17. Of course, there might appear to be something implausible about assuming that one person can observe another's state of mind, e.g., another person's intentions. But not even this is really necessary. What is needed is only that Art *believes* this about Bart. However, as I hope now to suggest with this variation on the original example, the real implausibility of the backwards induction argument does not seem to turn on the particular version of interpersonal transparency that is used. The real problem appears to be in the notion of individual rationality that is being assumed.

Thus, he knows, by assumption, that he cannot form the requisite intention at the move before this, when there are three coins on the table (and where it is Bart's move), to take only one coin on the next move. Thus, he knows that an offer from Bart at this point is worthless to him. But then, he asks himself, why not take two coins on the move (his second last possible move) just before this move by Bart? Art would only not take two coins if, by taking one coin instead, he could again get Bart to make a worthwhile offer to him at the next move. But Art has already concluded that such an offer is worthless to him because he cannot form the requisite intention to make it worthwhile. So Art knows that it is pointless to take only one coin on his second to last move; he should take two. But then, of course, he cannot form the requisite intention at Bart's immediately prior move to take only one coin. And so Bart's offer to him at that point is also worthless. But why then, he asks himself again, should he not take two coins at his third last possible move? To take only one coin at this point only generates another worthless offer. In like manner, it can be shown that all the prior offers that Bart might make to Art are worthless and that, as a consequence, Art will take two coins on the first move of the game. And none of this argument makes any general demands on Art's knowledge of Bart's rationality or vice versa. All that is required is that Art know that he cannot form an intention to do something that he knows he will not rationally do.

This last requirement seems acceptable in general, but particular interpretations of it might not be. The real force of the requirement is in the idea that a rational agent cannot intend to do what he knows he will not do. But how does he know that he will not do it? Because, the argument goes, he knows that it will be *irrational* for him to do it. So far, so good; this much also seems acceptable. The difficulty arises on the interpretation of practical rationality that is used. If practical rationality means, simply, acting for (undefeated) reasons (and in rational choice theory this means acting according to reasons as manifested in preferences, all things considered), then the requirement reduces to the idea that a rational agent cannot intend to do what he knows he has reason not to do. For then he knows he will not do it, and this contradicts the real force of the requirement. But suppose that there was more to practical rationality than acting for reasons. Then it would be possible for a rational agent to intend to do something that he had reason not to do. Why? Because then he might *not* know that he would not rationally do it even though he knew he had reason not to do it. And without the knowledge that he might not rationally do it, he could intend to do it in a way that is consistent with the real force of the requirement.

Thus, the possibility that there is more to practical rationality than acting for reasons opens up the further possibility that an agent can intend to do what he has reason not to do. Again, it is worth emphasizing that this is not the same as saying that he can have a *reason* to intend to do something that he has reason not to do. That was the possibility with which we began our investigation in this Part of the Article. And we saw fairly quickly that an agent could have such countervailing reasons; the examples of the centipede game and of promising seem to establish this point in a practically important way. What was problematic for the agent, however, was whether the reasons that he had for his prior intentions or promises could ever be made effective: could he actually form these intentions, or make these promises, if he had reason actually not to do as he intended or promised? The backwards induction argument, as applied to intentions (in the *intra*personal knowledge case) and promises (in the *inter*personal knowledge case), suggested not. But now we can see that this argument turns on the same assumption that we have been questioning all along—namely, that practical rationality consists only in acting for reasons. For only then does the real force of the more general requirement—that an agent cannot rationally intend or plan to do what he knows he will not rationally do—reduce to the more particular idea that an agent cannot rationally intend or plan to do what he knows that he will have reason not to do.[18]

The suggestion here is that we should accept the real force of the general requirement, but not the particular interpretation of that idea that drives the backwards induction. That is because there is something more

18. In some very helpful comments on an earlier version of this Article, Wlodek Rabinowicz questioned whether it was plausible to impose this general requirement (viz., that an agent cannot rationally intend or plan to do what he knows he will not rationally do). He suggested that even if an agent knew that he would not do x rationally when the time came actually to do it, the agent could nevertheless rationally intend or plan to do x if it was thought that forming the intention or plan would make it more likely that x would actually be done (albeit not rationally). It may even be that Ulysses binding himself to the mast to overcome (nonrationally) the lure of the Sirens provides us with a classic example of such an effective and rational plan. However, in this sort of example, it seems that the physical restraint rather than the intention itself is doing the work to hold the agent to the plan. If Rabinowicz means to suggest that the mere fact of having adopted the intention or the plan, without more (such as using physical restraints, giving up hostages, etc., measures which either avoid the influence of reasons or change their balance at the moment of acting) can make it more likely that the act will be done, then he is closer to the structure of the problem being analyzed here. But then, as this Article will go on to argue in the next section, I am inclined to say that an act carried out under the normative requirements of an adopted intention or plan is rational rather than irrational.

to practical rationality than acting for reasons. I hope that this Part of the Article has given us some indication of why it might be important that there is something more. The next Part of the Article will tell us more specifically what that something more is.

III. THE NORMATIVE REQUIREMENTS OF PRACTICAL RATIONALITY

Let us begin by reconsidering our earlier example of theoretical reasoning. Theoretical reasoning, it is said, takes us from one belief state to another. Thus, if you begin by believing the proposition that Frankfurt is in Germany (FG) and the proposition that Germany is in Europe (GE), then theoretical reason would have you conclude by believing the proposition that Frankfurt is in Europe (FE). Suppose that you do in fact believe FG and GE. Does this mean that you have a reason to believe FE? You may have reason to believe this (as it happens you do!), but not because of your beliefs about FG and GE. In fact, you might have no reason at all to believe FE or only have reasons not to believe FE. Thus, while it is true that *if* you believe FG and GE, you should *then* believe FE, there is nothing in this that gives you any reason to believe FE.

To see why, consider this alternative example. Suppose that you believe the proposition that Toronto is in Germany (TG) and the proposition that Germany is in Europe (GE). Then, theoretical reason would have you conclude by believing the proposition that Toronto is in Europe (TE). But you have no reason, based on these beliefs, to believe TE. Indeed, you have many other reasons, independent of these beliefs, *not* to believe TE. And it is not that these other reasons, based on independent beliefs, simply prevail over, or outweigh, the reason you have to believe TE based on your beliefs in TG and GE. Rather, it is that there simply is no such reason to believe TE at all. *Any* independent reason not to believe TE would be enough to provide an all-things-considered reason not to believe TE, at least if the only "reason" that you claimed for believing TE was your belief in TG and GE. This suggests that the weighing of conflicting reasons simply has no application here. The beliefs in TG and GE add nothing to the balance of reasons for believing TE.

But there does seem to be some sort of normative connection between believing TG (or FG) and GE and believing TE (or FE). What is that connection if it is not that believing the first two propositions provides a "reason" for why you should believe the third? John Broome provides an answer.[19] Although your beliefs in the first two propositions provide no reason for you to believe the third, they do *normatively require* you to believe the third. Normative requirements differ from reasons, says Broome, in that they are *strict* and *relative*. They are strict because, in the context of theoretical reasoning, they really do *require* or obligate you to the conditional that *if* you believe TG and GE, *then* you should believe TE. If you believe TG and GE, but do not believe TE, then you are not entirely as you should be; in particular, you have failed to meet the normative requirements of rationality (here, the requirements of good theoretical reasoning). But these requirements, while strict, are relative because they do not detach from the conditional proposition "if . . . , then . . ." and, therefore, do not give you any reason to believe TE *tout court*.

Reasons, on the other hand, are not relative in this way; they do detach and do give *independent* reasons, say, to believe TE (e.g., perhaps a very reputable geographer told you that TE). But these reasons are not strict; they are only *pro tanto*. That is, while you might have this independent reason to believe TE, it still might be that you do not believe it, perhaps, because you have some other independent *stronger* reason for *not* believing it (e.g., that TE goes against everything you were taught in school). However, because reasons are not strict, not believing what you have a reason to believe is quite consistent with being entirely as you ought to be. While there might be *a* reason to believe TE, the *balance* of independent or detached *pro tanto* reasons might be such that you do *not* believe TE. That is no problem.

Reasons, therefore, are weaker than normative requirements in being only *pro tanto* and not being strict. But they are stronger than normative requirements in being independent rather than relative. These are differences that go to the very logical structure of each. We are now ready to see how these important logical differences are relevant to practical reasoning and what they can do for an agent.

Practical reasoning, as I have already said, differs from theoretical reasoning in that it concludes in a state of mind that involves a decision or intention (usually, to act) rather than a belief.[20] Here is an example:

(1) I intend that (I will visit Heidelberg); and
(2) I believe that (to visit Heidelberg I need to fly to Germany); and so
(3) I intend that (I will fly to Germany).

19. Broome, *Requirements*, *supra* note 4, at 401; Broome, *Are Intentions Reasons?*, *supra* note 4, at 105.
20. Broome, *Normative*, *supra* note 4, at 175.

The bracketed propositions provide the content for the different statements and the prior nonbracketed terms reveal my state of mind, or attitude, with respect to each of the propositions. The logic of the reasoning is contained in the propositions themselves.[21] We can see this if we think of these same three propositions under the aspect of theoretical reasoning, where only belief states of mind apply. If I believe the bracketed proposition in (1) and the bracketed proposition in (2), then the "and so" logic of theoretical reasoning will have me conclude that I believe the bracketed proposition in (3). In the practical reasoning that is described by the above example, the same "and so" logic applies, although now it takes us from an intention state of mind in (1) and the belief state of mind in (2) to the concluding intention state of mind in (3).

We can now pose questions about practical reasoning that are fully analogous to the ones that we posed earlier about theoretical reasoning. Does my prior intention in (1) together with my belief in (2) give me any *reason* for my final derivative intention in (3)? No, not any more than the same logic applied to the following three statements would give you any analogous reason to have the derivative intention in (6) in this example:

(4) I intend that (I will visit Heidelberg); and
(5) I believe that (to visit Heidelberg I need to fly to Canada); and so
(6) I intend that (I will fly to Canada).

The prior intention in (4) together with the belief in (5) gives me no reason to have the derivative intention in (6).

Nor is any reason that I might have for my prior intention in (1) (or in (4)) transferred by the logic of practical reasoning into a reason for me to have the final intention in (3) (or in (6)). I may have independent *pro tanto* reasons to intend to fly to Canada or not to fly to Canada, and what I have most reason to do in that respect will be determined by the balance of these independent reasons. However, the fact that I have a reason to have the intention in (1) (or (4)) will add nothing to the balance.

But it is true that I am *normatively required* to have the intention in (3) (or in (6)) *if* I have the intention in (1) (or in (4)) and the belief in (2)

21. Broome, *Practical Reasoning*, *supra* note 4, at 89.

(or in (5)). While relative in this way, this normative requirement of practical rationality is, as all such normative requirements are, strict. In other words, if I do have the intention in (1) (or in (4)) and the belief in (2) (or in (5)), then, if I do not have the intention in (3) (or in (6)), I am not entirely as I should be. In particular, I have failed to meet the normative requirements of practical rationality.

These are, by now, familiar enough points. So let us add a little conflict into the mix. Suppose that I do have the intention in (1) and the belief in (2). Then, I am normatively required to have the intention in (3). If I don't, I am not entirely as I should be. But suppose that I have an independent reason *not* to have the intention in (3) and, further, no independent reason to have it (perhaps there is a strike by air traffic controllers in Germany, making any flight to Germany less safe). Then the strict normative requirements of practical rationality are in conflict with my independent *pro tanto* reasons. Am I still entirely as I should be? It seems not. Something is wrong here and needs sorting out.

Here is where the *relative* quality of normative requirements of practical rationality can be useful. The strict quality of these normative requirements obligates me to have the derivative intention in (3), but only *if* I have the prior intention in (1) and the belief in (2). Thus, I can satisfy these strict requirements either by accepting the antecedent conditions of the conditional *and* accepting the consequent (*modus ponens*), or by rejecting the consequent *and* rejecting one or other (or both) of the antecedent conditions that require the consequent (*modus tollens*). The fact that I have an independent reason for rejecting the consequent seems to provide me with some motivation for the second method of satisfying the normative requirements of practical rationality. Then, I could satisfy both my independent reason for not having the intention in (3) *and* the strict normative requirements of practical rationality. And, after this adjustment, I would be entirely as I should be.

Suppose, as seems reasonable, I cannot adjust my beliefs in (2).[22] Then, to make the necessary adjustment, I would need to change or repudiate my prior intention in (1).[23] But that does not seem problematic, at least on the argument so far. So far I have not provided any reason for

22. On the difficulty of deciding to believe, see BERNARD WILLIAMS, *Deciding to Believe*, *in* PROBLEMS OF THE SELF 136, 136–51 (1973) (describing the dilemma of whether belief can be related to decision and will).

23. Broome, *Are Intentions Reasons?*, *supra* note 4, at 112. Note that, for Broome, repudiation is more than merely ceasing to have the prior intention, but it might not require a reason either. For suppose there was no reason for the prior intention. Why, then, should it take a reason to give it up? Broome requires repudiation to be deliberative, but not necessarily with reason, something that is a little mysterious.

my prior intention in (1); there is only the fact that I have it. But it seems implausible that the mere fact of having this prior intention could count for much if I have an independent reason not to have the derivative intention in (3). This is consistent with the insight that a prior intention in (1), together with the belief in (2), gives me no reason to have the derivative intention in (3). Thus, while the normative requirements of practical rationality strictly require me to have the intention in (3) if, as a matter of fact, I have the intention in (1), they do not provide much normative resistance against my changing that fact by repudiating the intention in (1).

What if you had no reason to adopt the prior intention and no reason not to follow through on it by adopting the derivative intention? Does this mean that you ought to satisfy the normative requirements of practical rationality by accepting the antecedent conditions of the conditional and accepting the consequent? John Broome thinks not; you are still at liberty to repudiate the prior intention and deny the consequent. If there was no reason to adopt the prior intention in the first place, there is no reason not to repudiate it.[24] Yet he provides an interesting example that, ironically, goes some part of the way towards challenging the rationality of his approach.[25] While the example is somewhat special, it sets the stage, I believe, for thinking that there might be something irrational in always repudiating the prior intention and, more particularly, in repudiating it in those cases where one has a reason in favor of a prior intention and a reason against the derivative intention. The latter, of course, are the cases most analogous to those we saw earlier when we discussed whether you might have a reason to choose to do something that you have reason not to do—namely, the sorts of situations captured by the centipede game and promising.

Broome's example, borrowed from Kierkegaard's *Fear and Trembling*, turns on the idea that certain values are incommensurable. We are to imagine the situation where God has told Abraham to take his son Isaac to the mountain and sacrifice him there. The options at the moment of prior intention are either to intend to obey, thereby showing one's submission to God, or intend to disobey, thereby saving Isaac's life and preserving one's relationship with him. Broome argues that the values here are incommensurable, something that does make sense of these

24. *Id.* at 118.
25. *Id.* at 114.

situations (and other such tragic choices[26]) as posing a moral dilemma for the protagonist. Neither option is better than the other, nor are they equally good. For some theorists, this is simply what incommensurability means.[27]

Yet, Abraham must decide. Because the options are incommensurable, there is no reason to decide in favor of either one of the options rather than the other.[28] As we know from the story, Abraham decides to obey God and sets out for the mountain. But now, suppose that, at any point—say, at the foot of the mountain—he can change his mind and repudiate his prior intention. Is there any reason not to? If the values continue to be incommensurable, the answer is presumably not. So, this is a situation where, first, there is no reason to adopt the prior intention to obey God rather than save Isaac and, second, no reason to carry out that prior intention as a derivative intention rather than not carry it out under a repudiation of the prior intention. Because of the incommensurability of the values, the balance of *pro tanto* reasons has no role here.

Yet, simply because one has already formed the prior intention, there might be something problematic in repudiating it, a point Broome explicitly recognizes.[29] For by adopting the prior intention and then repudiating it, Abraham has ended up taking a course of action that is worse than another course of action that he could have adopted by never adopting the prior intention in the first place. Had he never adopted the prior intention to obey God, he would, of course, have given up the value that was contained in that option. But, at least, he would have preserved *all* that was valuable in his relationship with Isaac. However, by adopting the prior intention and then repudiating it at the foot of the mountain, Abraham has both given up the value of obeying God *and* sacrificed *something* of his relationship with Isaac. Thus, although neither of his decisions (for one option rather than the other) was wrong, or against reason (again, because these values are incommensurable), the

26. *See* GUIDO CALABRESI & PHILIP BOBBITT, TRAGIC CHOICES 18 (1978) (pointing to the inability to reconcile deeply incommensurable values). Great tragedy more generally makes much of these moments of choice between the seemingly incommensurable—e.g., Antigone's choice between the obligations of citizenship and the obligations to her dead brother or Agamemnon's choice between his daughter and his obligations as a military leader, etc. *See* SOPHOCLES, ANTIGONE 56 (Richard Emil Braun trans., 1973); AESCHYLUS, AGAMEMNON 26–29 (Hugh Lloyd-Jones trans., 1970).

27. *See* JOSEPH RAZ, THE MORALITY OF FREEDOM 322 (1986).

28. This is *not* to say that there are no reasons for choosing one *or* the other. Indeed, that there are such reasons is the source of the dilemma. However, while there are reasons for choosing the one option *and* reasons for choosing the other, because of the incommensurability there are no reasons for choosing the one *rather than* the other. I am grateful both to Shachar Lifshitz and to Wlodek Rabinowicz for encouraging me to clarify this point.

29. Broome, *Are Intentions Reasons?*, *supra* note 4, at 116–17.

two decisions together add up to a course of action that was worse for him (and, incidentally, for Isaac) than one he could have chosen. And this last assessment is one we can make despite the incommensurability of the values; the course of action finally chosen is no better for the value of obeying God and worse for the value of preserving the relationship with Isaac.

Had Abraham not repudiated his prior intention, then he would have at least achieved the value of obeying God (which he had no reason not to do). Thus, a resolute commitment to carry out his prior intention allows Abraham to avoid a bad outcome in a way that his repudiation of that prior intention does not. Broome concedes this, but does not consider it *irrational* for Abraham to repudiate the prior intention. Why? Because, he says, there was no reason, at either of the two points where a decision had to be made, not to make the particular decision that was made. Thus, since no particular decision was ever made contrary to reason, there was no irrationality in the overall course of action.

This comes very close to reducing practical rationality to action according to reason, something that Broome has been very careful to avoid. Of course, Broome is not, strictly speaking, guilty of this reduction. After all, at the base of the mountain, when Abraham decides to turn back from the sacrifice, Abraham is also obliged to repudiate his prior intention so that the normative requirements of practical rationality are met. Nevertheless, Broome's willingness to count repudiation as *a course of action* that is as rational as being resolute and following through on the prior intention—despite the fact that values are less well served in the former—is troubling, and it seems to depend on the idea that there cannot be less rationality in the former if its *particular* choices or actions are no more contrary to reasons than actions in the latter.

Perhaps, under repudiation, the problem is the way that the rationality of all the particular choices, as choices never contrary to reason, are generalized to a broader and more categorical assessment of the rationality of an overall course of action. This, it will be recalled, was the problem we confronted earlier in our discussion of backwards induction. While Broome's careful analysis of normative requirements allows him to claim that there is more to practical rationality than acting according to (the balance of *pro tanto*) reasons, his unwillingness to privilege any particular method for resolving violations of these normative requirements exposes him to the possibility that a resolution can proceed just as rationally from back to front, as in the repudiation of a prior intention, as front to back, when the prior intention is actually carried out under a later derivative intention. The result, I think, is no less paradoxical than what we saw under backwards induction. One might have hoped that all this structural analysis of practical rationality would do more for us than this.

And I think it can. However, before showing what this might be, we need first to appreciate why Broome might be concerned about labeling the repudiation of a prior intention as irrational. Consider the following sort of example, which Broome discusses briefly in another context before he developed his analysis of the difference between reasons and the normative requirements of practical rationality.[30] Suppose that you have good reason to choose to drop in on a friend, Mrs. Silstein, on the way home from work. She is an elderly widow, on her own, and would love a visit from you. On the other hand, you really do not want to spend the whole evening with her (perhaps there is something on television that you do not want to miss). So the best plan of action, a plan that most satisfies all your reasons for acting, is to form the intention to stay only a short time and be resolute in carrying out this intention. Call this course of action Resolute (R). However, you know that, once there with Mrs. Silstein, you will want to stay and keep her company, with the result that you will give up the whole evening (and your television show). Call this course of action Stay (S). So, on the way to seeing her, you repudiate your prior intention to visit with her and head straight for home. Call this course of action Home (H). This appears to be the best that you can do for yourself given your current preferences and what you think is possible for you to do rationally.

Broome recognizes that it is tempting to label the resolute course of action R as more rational than the course of action H involving repudiation. By being resolute, you can achieve everything of value in being home *and* something of value in being there for your friend. The repudiation of the prior intention sacrifices the latter possibility completely, something which is contrary to the reasons that you have at the moment of the prior intention (when you rank the different courses of action in the order, from left to right, RHS) *and* at the moment when you are actually visiting (when you would rank them SRH).[31] So, while

30. John Broome, Book Review, 102 ETHICS 666, 666–67 (1992) (reviewing EDWARD F. MCCLENNEN, RATIONALITY AND DYNAMIC CHOICE (1990)).

31. Of course, when you are actually visiting your friend, course of action H is not actually available any more. But it is still sensible to say, at that point, that you rank the course of action that repudiates the possibility of visiting as worst of all ("How awful it would have been not to have had this time together!"). It is useful to compare these rankings (RHS and SRH) with those presented in the problem posed by Amartya Sen in his theorem on the "impossibility of a Paretian Liberal." *See* Amartya Sen, *The Impossibility of a Paretian Liberal*, 78 J. POL. ECON. 152, 152–57 (1970). Sen showed

reasons and commensurable values are in play here in a way that they were not in the Abraham and Isaac example, there seems to be the same kind of irrationality in repudiating one's prior intention here as there was in that example. But this is where Broome offers a powerful objection. While Broome felt that, in the face of incommensurable values and the absence of reasons to choose one incommensurable option rather than another, there was no irrationality in Abraham either resolutely holding to his prior intention or repudiating it, in the Mrs. Silstein example he argues that there would be irrationality in being resolute and *not* repudiating the prior intention. Consider what he says:

> I think that resolute action is irrational. Having dropped in on Mrs. Silstein, what reason can you possibly have for leaving? You do not want to, and nothing would be gained by doing so. If you have a reason, it must be that your original decision to leave early, made before you entered the house, supplies one by itself. The reason you had for making this decision—that at the time you preferred leaving early to staying—is defunct, now that you have changed your preferences. If resolute action is rational, then, the bare fact of having decided to do something must in itself be a reason for doing it. But it is plainly not. Suppose that your original decision resulted from a false belief: suppose that, on entering the house, you find that Mrs. Silstein needs your company more than you expected. Or suppose that, when you decided, you did not realize that you would later change your preferences; you simply change your mind in an ordinary way. In both these cases, you should stay with Mrs. Silstein, and in neither would the bare fact of having made the opposite decision constitute even the weakest reason for leaving early. No more does it in the original example.[32]

Much in this objection anticipates Broome's later analysis, already discussed here, that the mere fact of prior intentions, and even the reasons that we might have for forming them, cannot provide any reason, as opposed to normative requirements, for forming the corresponding derivative intentions and actually following through on

---

that, for some configurations of preferences, the assignment of even the most minimal powers of decisiveness (or rights) to two (or more) different individuals might mean the violation of the Pareto principle (or lead to an outcome where all the individuals are worse off than they could otherwise have been). The same sort of problem arises in the Mrs. Silstein example under dynamic choice where, instead of two different individuals, we are dealing with only one individual, but one who has an early self and a later self with different preferences. The later self (visiting with Mrs. Silstein) has the power (and a preference) to choose course of action S (staying on with Mrs. Silstein) over course of action R (resolutely leaving early); the early self, anticipating this choice by the later self, has the power (and a preference) to choose course of action H (repudiating any visit with Mrs. Silstein) over course of action S. The result is course of action H, a course of action that both the early and later selves consider inferior to course of action R. Thus, we end up with a Pareto inferior outcome, for some a (socially) irrational result.

32. Broome, *supra* note 30, at 668.

them with actions. As Broome suggests, without any reason not to, you are always free to change your mind. And given his more recent analysis of the normative requirements of practical rationality, you change your mind deliberatively and rationally by repudiating your prior intention.

However, despite what Broome says in the last sentence of this passage, there are important differences between the original example and the close variations on it that he discusses in the course of making his point in favor of the rationality of repudiation. The differences turn out to be important once one has taken on the idea that normative requirements, as well as reasons, play a role in practical rationality. Consider again the original example. In that example, at the time you form your prior intention to be resolute (adopting course of action R), you anticipate and deliberate over the possibility that, when you are visiting with Mrs. Silstein at the later time, you will want to stay rather than leave early. Yet, in the face of that fully anticipated possibility, you, nevertheless, form the intention or plan to be resolute. Of course, you do so with good reason (namely, that resolute action seems to allow you to satisfy both your prior preferences for an evening at home and something of your preferences, before and after this point, for spending some time with Mrs. Silstein), but the rationality of now following through on this prior intention, in the face of contrary reasons that you had fully anticipated would arise, is not reason-based. That it *could not* be is Broome's constant point in emphasizing the logical distinction between reasons and the normative requirements of practical rationality. However, following through on the prior intention could be normatively required.

Now, Broome will say that the normative requirements of practical rationality, being relative, do not require any follow through in the consequent derivative intention (and action under that derivative intention). It is just as consistent with these normative requirements, and, therefore, just as practically rational, to abandon any consequent derivative intention (and action) so long as one repudiates the antecedent prior intention. But what reason is there to do the latter? Broome's *variations* on the original example suggest *new* considerations that might provide a reason. For example, "you find Mrs. Silstein needs your company more than you expected,"[33] or "you did not realize you would later change your preferences [and] you simply change your mind in an ordinary way."[34] Of course, on these variations, the changes in the situation might well give you a reason to repudiate your prior intention.

33. *Id.*
34. *Id.*

You did not think of these possibilities in advance, and it would be almost thoughtlessly mechanical to go ahead with the prior intention without allowing these new possibilities to have some sort of rational impact on what you should do. But in the *original* example there are no such surprises for you. The situation, complete with the reasons that you now have for wanting to stay longer with Mrs. Silstein, has unfolded exactly as you anticipated it would when you formed your prior intention or plan. How, therefore, can there be any reason, not already contemplated *and accounted for* under that prior intention or plan, to change your course of action? Surely, there is only the normative requirement to carry out the course of action, as it was planned or intended, for exactly the sort of circumstances (including the reasons that you now have for staying) that arose.

Attending to this difference between the original example and its variations *does* require you to attend to some of the content of what you intended and, more specifically, what you contemplated under the prior intention. You need to do this to know whether the prior intention unambiguously *applies* to the situation as it arises at the time of forming the derivative intention and acting on it. This may even have you considering what was thought to be the balance of *reasons* at the moment of this prior intention—for example, that on balance you thought they favored leaving Mrs. Silstein early even though you knew that, at the time of the visit, you would want to stay. But, when this sort of content for your prior intention is brought forward into the situation where carrying out the prior intention is at issue, it is not that the reasons *qua reasons* are being carried forward to that point. Again, that would be to fail to appreciate Broome's essential point about the difference between reasons and the normative requirements of rationality. Rather, when the content of the prior intention, and even the balance of reasons that make up this content, is brought forward into the actual choice situation, it is only to apply correctly, and accurately, the normative requirements of practical rationality, not the original reasons for the prior intention.

This is an important difference made possible by Broome's logical distinction between reasons and the normative requirements of practical rationality. For if (contrary to Broome's argument) the reasons themselves were brought forward, then they would be the sort of thing that was accessible to the claim that those reasons are now defunct or no longer apply as reasons in exactly the way that Broome suggests in the above passage. However, if they come forward as a content for the application of a prior intention, *not yet repudiated* (or defunct) and, given their content, if they account already for all reasons that might otherwise have called for such a repudiation, then all that remains unaccounted for is the satisfaction of the normative requirements of rationality in the actual doing of what one has intended—even, it should be noted, if one has (an anticipated) reason *not* to do what one has intended.

This last point is the crucial one for the final application of this analysis to the problems uncovered in the last section when we looked into the centipede game and the problem of promising. For there the promisor, faced with playing the centipede game, seemed to be caught by the fact that, while he had a reason to make the promise to take only one coin, and the same reason to *intend* to carry out that promise, he also had a reason not to carry it out. And, if he knew that he had a reason not to carry it out, he would know, on the argument that reduced rationality to never acting contrary to (the balance of) reasons, that he could not intend to carry it out. This seemed to render the whole practice of promising worthless to him.

But suppose, as this Part of the Article (following Broome) has argued, that there is more to practical rationality than acting in accordance with (the balance of) reasons. Suppose in particular that there is the practical rationality of acting under normative requirements as well. Then there would be the possibility of forming an intention to do what one has reason not to do. More specifically, one could form the intention to carry out a promise that one has reason not to carry out. But now suppose further (contrary to what Broome seems to allow) that the *content* of this intended promise was such that it anticipated and accounted for the very consideration that is proposed as a reason for not carrying it out as intended. The promisor, in other words, is well aware, at the moment that he forms his prior intention, that he will have (some particular balance of) reasons to renege on the promise that he makes. And yet he makes the promise, with the full intention of carrying it out. Then, on the analysis provided in this Part, the (balance of) reasons not to carry out the promise, having already been accounted for under the intention to form the promise, would only have a place under the normative requirements of rationality to carry out the promise that one has intended (at least, *if* one has, in fact, formed the prior intention to carry out the promise, and only so long as one has not repudiated this prior intention). These reasons would not, in other words, act as further *independent* reasons for *not* carrying out the promise because these reasons have already been anticipated and comprehended within the prior intention that forms the basis for applying the normative requirements of practical rationality. But without any independent reason not to do as one has intended, and without any independent reason to repudiate that prior intention, there is *only* the normative requirement to carry out the promise as intended. Thus, there can be practical rationality in carrying out a promise that one has made, and might have had reason to make, even if one has (a balance of) reasons not to carry it out. The key is that the (balance of) reasons for not carrying out the promise must be anticipated and accounted for under the prior intention.[35]

Broome is correct, however, to challenge the rationality of not repudiating one's prior intention in the face of new *unanticipated* reasons for not doing as one has intended to do. This is the force of the variations that he provides on the original Mrs. Silstein example. To feel obliged, under the normative requirements of rationality, to do something that one has formed the prior intention to do, simply because as a matter of fact one has intended it, and even though *new*

35. It would also be important that the prior intention satisfy some sort of rationality requirement at the time it was formed. An obviously irrational intention or plan, for example, should not oblige the agent to carry it out as a matter of normative requirement even if the circumstances ex post were exactly as they were anticipated ex ante. Nor would such a rationality requirement on the prior intention be limited to the idea that the agent be better off if circumstances unfold as intended, or as anticipated, than she would be if no such intention or plan been adopted at all. Such a requirement might be either too restrictive or not restrictive enough, depending on what set of anticipated circumstances the agent considered for the purposes of this comparison. For example, under a threat of mutual annihilation, carrying out the threat seems irrational, even though, strictly, it might be normatively required if the threatened party simply acts in the very way that the threat contemplated. The source of the irrationality, however, is in the making of such an extreme threat initially, even though, had the threat been successful in at least one of the ways the agent must have anticipated as possible, the agent (not having had to carry out the costly threat) would have been better off than she was not having made any threat at all. Thus, this last welfare comparison is not restrictive enough. One might be tempted to say in the alternative, therefore, that it is not rational to form an intention or plan, and to carry it out as normatively required, if carrying it out makes one worse off than would have been had the intention or plan not been adopted at all. But this is too restrictive a comparison to make since it seems to preclude making some perfectly rational (more moderate) threats, at least if the carrying out of those threats can make the agent worse off than she was before making the threat. It is obviously crucial, therefore, for the overall theory of rational commitment to have an account of what intentions or plans it is rational to adopt. For excellent analysis of this point, see Joe Mintoff, *Rational Cooperation, Intention, and Reconsideration*, 107 ETHICS 612, 635–42 (1997). The purpose of this Article, however, is only to emphasize that once such rational plans and intentions are adopted, it is also rational, as a matter of normative requirement, to carry them out, at least if they are not repudiated for some unanticipated *pro tanto* reason, and even if doing so is contrary to the balance of fully anticipated *pro tanto* reasons.

considerations have arisen which suggest that there are now reasons *not* to do as one intended, does seem irrational. Indeed, it seems closed minded and mechanical rather than rational. Someone who has formed a prior intention to do something after a careful consideration of certain consequences might well feel rational in following through on this intention, even if the already considered consequences are of the kind that now give reasons not to do as one has intended. That is simply to feel the normative requirements of practical rationality. However, someone who allows the prior formation of such an intention to blind him to any new reasons that might countervail the prior intention is blocking out another component of practical rationality—namely, the *independent* reasons that one has for action at any point in time. While it is a mistake to think that all of practical rationality is action according to the balance of reasons (or the balance of preference based on reasons), a common mistake that Broome corrects with his analysis of normative requirements, it is also a mistake to think that (at any moment calling for action) all of practical rationality is action according to normative requirements. A *full* account of practical rationality must comprehend *both* the rationality of following through on one's prior intentions, even though there may arise (fully anticipated) reasons for abandoning those intentions, *and* the rationality of always being open to the repudiation of these prior intentions in the face of new, truly independent (unanticipated) reasons. This is the full account of rationality that is allowed for by Broome's analysis, and it is an account of rationality that, even in the face of countervailing reasons, allows a rational agent to follow through on rational commitments, or promises, rationally made.[36]

IV. DEFEASIBLE LEGAL RULES

The argument to this point has been complicated by the presence of some quite subtle theoretical distinctions. For example, there appears to be an important difference between reasons and practical rationality. While reasons are an important part of practical rationality, they are not its whole. There are also the normative requirements of practical rationality

36. While it is not the purpose of this paper to develop this point in any detail, it is tempting to speculate that when something unexpected comes up that provides the agent with an independent *pro tanto* reason against carrying out some prior intention, then the agent should weigh these new reasons against the balance of reasons that she had for adopting the prior intention and carrying it out in the first place. Of course, this does mean that in this unexpected new circumstance, previously accounted for reasons *are* being reconsidered. What is not permitted under a rational commitment is the reconsideration of previously accounted for reasons when circumstances develop exactly as expected. Again, I am grateful to Wlodek Rabinowicz for raising this point and encouraging me to consider it.

to be accounted for, and these are logically distinct from reasons. Normative requirements are strict, but relative, whereas reasons hold only *pro tanto*, but are independent. Further, for reasons themselves, there is an important separation between the different moments in which we can have reasons within a single decision process. We can (on balance) have reasons to choose or intend to do something that (on balance) we have reason not to do.

These subtleties, while interesting theoretically, might make us question whether there really could be any such form of rational decisionmaking in practice. However, in this Part of the Article, I want to argue that this form of rational decisionmaking is more than a mere theoretical construct. The institutionalization of this form of decisionmaking should be familiar to legal theorists as something that they study every day. The conjunction of reasons and the normative requirements of practical rationality, I suggest, is to be found in any system of defeasible legal rules. I will focus specifically on common law adjudication to make my point.

A system of common law is more than a mere list of all the decisions that judges have chosen to impose upon litigants. It is also comprised of the legal rules which are said to bring order to these different results. Of course, within the common law method of adjudication, these rules do not typically appear in some preexisting authoritative text, like the rules of a tax code. Rather, they develop over time in the cases—sometimes abruptly, more often gradually—as the general rules of, for example, tort or contract law. This might suggest that the general rules are mere descriptions of the behavioral regularities of judges. After all, if they do not preexist the cases, then the only other option seems to be that the rules come into place as rationalizations for what has actually been done for some independent reason in the particular case.

Now, it is certainly true that, like the laws of science, rules of law bring order and understanding to a legal reality which is independently laid down and which can be the object of external and scientific observation, the stuff of induction. But, more than this, rules of law are also said to bring order to a judge's self-conscious understanding of what she does and, further, of what she feels she ought to do. Rules, it is said, help to provide particular justifications for the legal result that she comes to in a case. Rules, therefore, can be said to order the law from both an external ex post point of view (the point of view of the scientific observer) and from an internal ex ante point of view (the point of view of a committed participant or judge guided by rules in the legal decisions she makes).[37]

However, that legal rules have this double aspect suggests that there will be some ambiguity as to what the proper relationship should be between these rules of law and their apparent instantiations in the cases. Under the more descriptive, scientific account, the particular case sets the standard for the rule. A rule will fail as a description, or fail to provide a proper understanding, insofar as it is an inaccurate representation of what is going on in the case. Although a limited number of exceptions can sometimes be said to prove the rule (since that is what the very idea of an exception presupposes), too many will be fatal to its descriptive claim.

On the other hand, under the more prescriptive account—where rules are said to provide reasons or justifications for judges to decide cases one way rather than another—the relationship between a case and the applicable rule is reversed. Now, the rule sets the standard for the case. Moreover, because the rule has this seeming autonomy from the cases, it can pronounce almost any number of them as wrongly decided. The number of such decisions only attests to the frequency of judicial error, leaving the legal rule intact and still perfectly capable of governing other cases.

To accommodate this dual aspect of common law rules, what is required is an account that allows the rules to be strong enough to guide judicial decisionmaking in particular cases, but not so strong that it does not allow for the possibility that these same rules might require revision in the light of these same cases. This might suggest that what we are looking for is something quite banal—namely, an account of rules that merely *weighs* the good that they do as rules (say, in securing general expectations, making life more predictable, controlling harmful judicial discretion, etc.) against the independent good that can be done by revising or abandoning the rule in some particular case. But I want to press the intuitive point that a rule, and the following of a rule, is more strict (or "rule-like") than this balancing or weighing metaphor allows. Borrowing from the above analysis, I want to now argue that a rule can *normatively require* a particular result in some case without regard to the good that is achieved in, or frustrated by, that result so long as this good (and its possible frustration) has already been anticipated by the rule and is accounted for in it. But a good that is frustrated by that result can also be an *independent reason*, even a decisive independent reason, for *not* following the normative requirements of a rule. A reason would be independent in the required sense if it involved a good that was not

37. For discussion of the importance of this internal point of view for the committed participant in law, see H.L.A. HART, THE CONCEPT OF LAW 55 (1961).

anticipated by the rule and accounted for in it. Thus, the integration of normative requirements of practical rationality and independent reasons promises to provide the right combination of rule respect and rule denial that is required if we are properly to accommodate the essential roles that rules and cases each play within a theory of common law decisionmaking.

This way of integrating the normative requirements of practical rationality with independent reasons is familiar enough to those who understand the common law as a system of defeasible rules. Mention of defeasibility, of course, reminds us of H.L.A. Hart, as Hart was influential in introducing the idea of defeasibility into legal theory.[38] Borrowing the idea from the law of property, Hart noted that "a legal interest . . . is subject to termination or '*defeat*' in a number of different contingencies but remains intact if no such contingencies mature."[39] Although he believed that this idea, or more particularly the dual structure of this idea, had wide application in the law, Hart developed the idea most explicitly with reference to the concept of a contract. He might equally have referred to a rule of contract formation. For Hart, as much as for other legal scholars, there is the usual list of positive conditions required for the existence of a valid contract (e.g., two parties, an offer by one, its acceptance by the other, consideration on both sides). However, knowledge of these conditions does not, according to Hart, give us a full understanding of the concept of contract nor of the rule of contract formation. We also need to know the various ways in which the claim that there is a contract (under the concept or the rule) can be defeated. Such defenses to the claim might include, for example, that there was fraudulent misrepresentation, duress, or lunacy. Hart argued, therefore, that the concept or rule of contract formation was best captured structurally as a list of conditions that are *normally* necessary and sufficient for the existence of a valid contract, together with a series of "unless" clauses that spell out the conditions under which this existence claim is defeated. However, Hart recognized that the list of unless clauses could not, in all likelihood, be exhaustively specified. And so, such a rule would often end only (and perhaps only implicitly) with the word "unless. . . ." However, Hart was clear: "A rule that ends with the word 'unless . . .' is still a rule."[40] Specifically, it is a defeasible legal rule.

38. H.L.A. Hart, *The Ascription of Responsibilities and Rights*, *in* ESSAYS ON LOGIC AND LANGUAGE 145, 145–66 (Antony Flew ed., 1951).
39. *Id*. at 148.
40. HART, *supra* note 37, at 136.

Now Hart's idea of a defeasible legal rule meets either with enthusiastic acceptance (amongst legal realists[41]) or outright skepticism (amongst legal formalists[42]) because of the flexibility that it appears to allow around the rigidity of rules. Here, I want to focus on the skeptical view, best articulated by Frederick Schauer.[43] Schauer argues effectively that the best interpretations of Hart's defeasibility claim either reduce his unless clauses to components of the rule, in which case there is only a relatively straightforward rule-like application of what is now a more complicated rule, or they allow the independent force of the unless clauses to modify the rule in light of some background justifications of the rule, in which case the background justifications are all that really apply and the force of the rule *qua* rule simply disappears.[44] This is to reduce a defeasible legal rule either to a rule without defeasibility or to defeasibility without a rule.

To make his point, Schauer considers seven possible interpretations of defeasibility. The first four advance variations on the way in which the unless clause in the rule can be incorporated into the rule, albeit with varying degrees of difficulty. For example, in the most simple case, it may only be that an unless clause is used expressly because some more convenient technical term or phrase (one that simply defines the conduct and incorporates the limiting idea within the definition without any use of the word unless) is not available.[45] Schauer rightly dismisses this as a "trivial linguistic point;"[46] there is no extensional difference in the application of two rules that are merely being expressed in these two different ways. Likewise, Schauer argues that a version of defeasibility that simply recognizes that any rule or principle, including legal rules, might be subject to some sort of override in the face of an overwhelming moral obligation cannot be what Hart was suggesting. Again, these overriding concerns could be quite conventionally incorporated into the rule by adding a closed-ended list of the relevant factors. Of course, it might be that these factors are only specifiable in the rule in some quite general way, as some broad type of consideration rather than something very particular. But this inability to pre-specify (under a more fully articulated rule) the full extension of the potentially overriding

41. Richard A. Posner, *The Jurisprudence of Skepticism*, 86 MICH. L. REV. 827, 834–35 (1988).

42. Frederick Schauer, *On the Supposed Defeasibility of Legal Rules*, 51 CURRENT LEGAL PROBS. 223, 223–40 (1998).

43. *Id.*

44. *Id.* at 226–27; *see also* FREDERICK SCHAUER, PLAYING BY THE RULES 212–15 (1991) (discussing same).

45. Schauer, *supra* note 42, at 227–28.

46. *Id.* at 227.

conditions (something Schauer calls "weak non-specifiability"[47]) is really no different from the inability to pre-specify the primary prescription under the rule for lack of an available technical term.

The first four interpretations by Schauer of Hart's defeasibility claim are the ones, therefore, that attempt to reduce defeasible legal rules to rules (albeit more complicated rules) without any significant defeasibility. If these were all that Hart meant to capture by advancing his claim, then Schauer would be correct to be skeptical. More interesting are Schauer's next three interpretations of Hart's claim, interpretations that resist the possibility of writing the defeating consideration into the rule. Under Schauer's analysis of these interpretations, a defeasible legal rule is reduced to a constant state of defeasibility in the light of particular considerations within the case. The result, says Schauer, is that rules *qua* rules disappear, being replaced by the direct application of the background justifications with respect to which the rules were thought to be instrumental.

Schauer begins this discussion with a reference to what he calls "strong non-specifiability,"[48] by which he means (in contrast to the weak form discussed above) an inability to specify, *even by broad type*, the sorts of conditions that might arise which would defeat a legal rule. Then, Schauer offers the following interpretation of defeasibility: "A rule is defeasible when its application is contingent not only upon the non-occurrence of events specifiable in advance by particular or type, but also by the non-occurrence of conditions specifiable in advance *neither by particular nor by type*."[49]

Schauer is careful to consider two possibly different sorts of situations where there might be such an unanticipated event or condition. First, there could be an event or condition that, while not precisely anticipated, clearly lies within *both* the linguistic contours of the primary rule *and* the contours of the rule's background justification. In that case, says Schauer, the rule is "simply applied, for the question of defeat does not arise when . . . language and purpose both encompass [the] case, even if it is not a case that has previously been imagined."[50] Schauer seems to think that this is not a very interesting case, although it looks to be a situation where the rule continues to operate as a rule. This appears to be because he thinks the rule is doing no real normative work here;

47. *Id.* at 231.
48. *Id.*
49. *Id.* at 232 (emphasis added).
50. *Id.*

everything is being done by a direct application of the justification or purpose that the rule serves and which, in this first sort of situation, is in agreement with what the rule requires. We shall have reason to return to this point in a moment.

The second sort of situation is one that Schauer thinks is more interesting. Here, the unanticipated event lies *within* the linguistic contours of the rule, but *outside* its background justification. Yet, says Schauer:

> [I]f any consistency between the rule (as formulated) and the result indicated by direct application of the rule's background justification is a sufficient condition for non-application of the rule, then the rule prohibits no action not prohibited by the background justification. . . . [A]ll of the normative work is being done by the justification and none by the rule.[51]

Combine this result with Schauer's understanding of the very limited role that a rule plays in the first sort of situation, where rule and background justification happened to agree, and the rule seems to disappear for all practical purposes in all possible situations. Thus Schauer concludes: "If defeasibility is purchased at the cost of the rule itself, the cost is too high, at least for the purpose of maintaining, with Hart and his followers, that ruleness and strong defeasibility can co-exist."[52]

However, this conclusion is too strong for the argument, and our earlier analysis of practical rationality, which showed it as being comprised of both reasons and normative requirements, shows why. Schauer is right to claim (as in the first sort of situation) that, if some event (although not precisely anticipated) lies within the contours of the primary rule and its background justification or (to borrow from earlier terminology) the reasons that we might have for having the rule, then the

51. *Id*. at 232–33.

52. *Id*. at 233. It should be remarked that Schauer does provide a final interpretation of Hart that saves something of the insight of the defeasibility claim, even for Schauer. Schauer argues that the force of rules might be *presumptive*, in that they are subject to defeat by the existence of particularly powerful defeating conditions that cannot and need not be specified in advance, so long as the requirement of particular power can be specified in advance and is not itself subject to defeat. This would allow the rule to carry some decisionmaking force up to some threshold point (the measure of which is itself nondefeasible), and before a direct application of the rule's background justification takes over. However, such an interpretation of defeasibility makes the choice between applying the rule and applying the background justification merely a quantitative matter. But without providing a more qualitative or categorical distinction between the application of rules and the application of background justifications, there is the possibility, and danger, that this interpretation of defeasibility will collapse the former into the latter and the normativity of rules will again disappear into the particularity of an all-embracing defeasibility. The categorical (logical) distinction between normative requirements and reasons that I suggest in this Article as a way to account for defeasible legal rules avoids this collapse.

rule is "simply applied." However, he is mistaken in thinking that we could just as easily apply the background justification directly in this sort of situation and, therefore, that the rule does no real normative work. The rule does do its own normative work here, but it does so under the aspect of a normative requirement of practical rationality, and not under the direct application of the background justification, or independent reason, for the rule.

The promising example, discussed earlier, makes this clear. Suppose we had a rule for seeing to it that our promises were performed as intended.[53] The reason or background justification for the rule might well be the welfarist one that we are better off performing our promise as intended than we would be if we made no such promise at all, and even if (we anticipate that) actually carrying out that promise makes us all worse off than we would be if we secured the benefits of making the promise and did not incur the costs of actually performing it. This presents us, of course, with a familiar problem. We have already seen that we might have a particular background justification or reason, which we can call reason *W* (for welfare), for intending to carry out a promise, or (now) having a promising rule, which can also furnish us with a reason *not* to follow the rule when the particular situation arises for carrying out the promise as previously intended. Nevertheless, as a matter of normative requirement, we should follow, or "simply apply," the rule because this is precisely the situation that was anticipated by the rule and was already accounted for in it. Moreover, we properly (rationally) follow the rule in the particular case even though there is (now) some cost in terms of reason *W* (which provides the background justification for the rule) in doing so. Thus, it is a mistake to think, as Schauer does, that, at the point where the rule is to be followed, we are simply appealing directly to the background justification. On that view, a view that (here) reduces practical rationality to acting in accordance with reason *W*, we would *not* follow the rule. But we do follow the rule, and we follow it as a matter of normative requirement. Further, we have no *independent* reason for not following the rule, having already anticipated and accounted for the countervailing consideration in terms of *W within the rule itself*.

With the practical significance of a rule now once again secured, we

53. The promising example is discussed both by HART, *supra* note 37, at 136, and by Schauer, *supra* note 42, at 224.

have only to recognize that in Schauer's second sort of situation, where some event occurs that appears to lie within the rule's linguistic contours, but which was not anticipated under the reasons that provide for the rule's background justification, there *is* an independent *pro tanto* reason for not following the rule. After all, under this more purposeful (or contentful, reason-based) understanding of the rule, the rule simply does not apply and there is no normative requirement actually to follow it. Indeed, there is only an independent reason for not following it, and for re-formulating the rule (perhaps with a further unless clause) in light of the new (unanticipated) *pro tanto* reason (or, perhaps, with a view to some new balance between new and old *pro tanto* reasons). Thus, the combined application of normative requirements and independent reasons makes sense of Hart's claim that both rules *qua* rules and the defeasibility of rules can sensibly be integrated into a full account of practical legal rationality. And it also allows us to comprehend the idea, alluded to at the beginning of this section, that particular cases can, apparently simultaneously, both determine legal rules and be determined by them.

V. CONCLUDING REMARKS

In this Article I have argued that a full account of practical rationality, as being comprised of reasons and normative requirements, can make good sense of rationally following through on commitments rationally made. Further, it can do so even if, in some circumstances, the balance of reasons is for not doing what you had reason to choose, intend, or promise to do. I have tried to suggest that there is much advantage in this for a rational agent, although it would be a mistake, a *rational* mistake, for the agent to let the securing of this advantage be the reason for doing what she does. In these circumstances, the agent should do what she has chosen, intended, or promised to do as a matter of normative requirement, *not* as a matter of reason. This is not to say, of course, that the agent should follow through on her prior commitments in some blind, mechanical way, without a view to new and unanticipated considerations. That, too, would be practically irrational; normative requirements are no more the whole of practical rationality than are reasons. Rather, what she should do is follow through on her rational commitment, even in the face of countervailing reasons, *unless* some new independent reason, not already anticipated and accounted for under the rule, prompts her to reconsider that commitment.

This suggests that a practically rational decisionmaker, at least in particular cases, will concern herself less with the substantive reason behind her prior rational commitment, at least *qua* reason, and more with the reasons that might arise for narrowing or broadening her prior commitments in light of developing circumstances. In other words, the practically rational decisionmaker will typically be working only on the margin of some more general rule adopted for application in the cases, pressing forward with it in the case unless she can be convinced that the case is unlike those she had already considered and anticipated under the rule.[54]

This, I have also tried to suggest, is precisely the sort of rationality that is manifested in common law adjudication. Of course, judges will often have to attend to a purposeful (reason-based) interpretation of the rules that they seek to apply as a matter of normative requirement. Otherwise, they would not have a full sense of what the rule really is. But their purposeful application of the rule is formal, not substantive, the stuff of normative requirements rather than independent reasons. They only address independent reasons when they are asked, typically by litigants, to either broaden or narrow the rule in light of special (novel) considerations arising in the particular case. In this way, common law judges act in a rule-based way, although they are not rule-bound. This is what Hart meant to capture in his idea of defeasible legal rules.

One final cautionary note: while I think that the practical rationality that is manifested in a system of defeasible legal rules is the sort of rationality that might prove helpful for understanding why a rational agent can rationally follow through on commitments rationally made, I do not mean to suggest that common law adjudication is solving the sort of problem that plagues the rational choice theorist confronted with the paradox of backwards induction. That paradox turns on there being a finite sequence of possible choices and the agent being able to look ahead to the end of that sequence from which the backwards induction begins. The common law has no such finite horizon, or at least not one with such a predictable ending. However, I do mean to suggest that the rational choice theorist can learn from a close study of the practical rationality that is manifested in common law adjudication. Armed with an appreciation of the role that is played by reasons *and* normative requirements in practical rationality, the rational choice theorist will be in a better position to understand and design institutions for rational behavior more generally.

54. For further discussion of this idea, and how it relates substantive reason to the formal equality of treating like cases alike, see Bruce Chapman, *Chance, Reason, and the Rule of Law*, 50 U. TORONTO L.J. 469, 477–89 (2000).