Rational Temptation

CLAIRE FINKELSTEIN

1

"To characterize a person as rational is not to relate him to any order, or system, or framework, which would constrain his activities," writes David Gauthier.¹ Reason is free, he thinks, to develop its own path to any ends it takes as given. The suggestion appears in the earliest of Gauthier's essays in which he recognizes the freedom of reason from the strict constraints of maximization. And while Gauthier did not intend his words in this sense, he might have put his point about maximization in this form as well: To conceive of practical reason as subservient to a system of ends is to conceive its independence from any determinate path in the pursuit of those ends. The form of reasoning, as well as its content, must be instrumentally supplied. Gauthier's famous claim that instrumental reason sometimes demands the constraint of its own maximizing activity is thus dependent on the recognition of reason's freedom from constraint.

But to say that reason is free from constraint in this sense is to posit reason's constraint in another. Like many committed to understanding practical reason instrumentally, Gauthier accepts the economist's picture of the ends it is reason's purpose to pursue as established by the preferences of the agent whose ends they are. Reason's freedom, Gauthier thinks, is the freedom to serve the preferences of rational agents. Gauthier's recognition of reason's freedom from maximization thus depends on his willingness to accept the idea of reason's constraint, constraint by the preferences of the rational agents it serves.

As we shall see, a tension emerges in the attempt to treat reason both as constrained by preferences and as free to conduct itself in any manner necessary to serve the ends those preferences establish. Constrained maximization requires a rational agent to act at times against the weight of her current preferences in pursuit of reason's ends. But the economist's understanding of practical reason as preferential is imperialistic: Once it is accepted anywhere in an account of rational agency, it becomes difficult to reject that same restriction on a rational agent's choice of each and every action. It thus becomes difficult to reject the idea of preferential constraint as a pervasive feature of rational agency.

In this essay, I shall argue that making sense of preferential constraint as a standard feature of practical reasoning commits one to regarding counter-preferential agency as irrational across the board. I shall argue the point in reference to the problem of changing preferences over time. As we shall see, Gauthier and others who share the central features of his account lean heavily on a supposed asymmetry between the way a rational agent forms plans and the way she executes them: Plan-adoptions are preferentially constrained, but the performance of actions in the execution of a plan is potentially counter-preferential. I shall argue that this asymmetry is not defensible and that proponents of this approach must choose between preferential constraint and freedom from preferences generally. I shall then make the further suggestion that the tension inherent in the sort of account Gauthier's theory exemplifies is best resolved by rejecting the idea that preferences ever function as an absolute constraint on practical deliberation for a rational agent. While I cannot argue for this rather bold claim in any detail, I can present some intuitive reasons for thinking it correct; reasons I hope will be of special interest to the rational choice theorist. My thought is that those who accept the broad picture of rational agency as self-interest ought to hesitate before articulating that view in terms of preference maximization.

2

I shall focus on a kind of change-of-preference case we might call rational temptation. A rational temptation is a temptation an agent faces due to a temporary shift in her preference ordering. By calling the temptation rational, I mean to distinguish it from another sort of familiar case. Sometimes agents are drawn to do things on the basis of appetites or urges, despite the fact that their overall preferences do not support doing them. A smoker trying desperately to quit might retain a preference that she not smoke, even while she finds herself drawn to take another cigarette. On all but the narrowest possible interpretation of "preference," we might describe this by saying the smoker is tempted to take another cigarette against the weight of her preferences.² In this way, we can distinguish an agent who yields to a rational temptation
from an agent who yields to an appetite or urge. The former faces a problem in rational choice theory. What she requires is an argument to convince her of the rationality of resisting temptation. The latter faces a problem in action theory or psychology. She is convinced of the rationality of resisting the temptation; she simply cannot conform her behavior to what she knows it is rational to do.

As our case of rational temptation, let us take an example of Michael Bratman's. A pianist performs nightly at a club, prior to which he eats dinner with a friend who fancies good wines. Each night the pianist is tempted to drink with dinner. He knows, however, that drinking wine will impair his evening performance, and thus prior to dinner he prefers that he refuse the wine. Nevertheless the pianist invariably finds himself tempted to sacrifice his performance for the sake of a good Cabernet.

The pianist's case is one of shifting preferences. Prior to and after dinner he prefers the option drink no wine with dinner and play well tonight. During dinner, however, the pianist has a reversal of preference. He temporarily prefers the sequence drink wine with dinner and play less well tonight. If he were to determine his behavior in accordance with deliberations conducted during dinner, he would maximize his preferences at that time by opting for the wine. Were he to do this, however, he would later regret his decision, since after dinner he would prefer to play well that evening and would wish he had forgone the wine. He is, moreover, aware that he will regret it if he drinks, for he knows that after dinner he will return to his prior preference ordering, which favors protecting his evening performance above all.

If the pianist is myopic with regard to his future preferences, he will go to dinner thinking it best to refuse the wine, but he will end up choosing to drink after all, since his preferences will have shifted by the time he must decide what to do. The myopic deliberator is hostage to his preferences at all times. The economist suggests a solution: The pianist should find a way of precommitting to the decision to resist wine with dinner before the change in preference occurs. The pianist might accomplish this, for example, by deciding not to go to dinner at all. This is the sophisticated solution to rational temptation cases. There is a sense in which it is an unfortunate solution, since it requires the pianist to deprive himself of dinner in order to avoid drinking. He might even wish he could go to dinner and not drink rather than having to avoid dinner entirely, but he knows he will drink if given the chance. He should therefore settle for eliminating the source of temptation altogether, by precommitting to a course of action that precludes drinking. The sophisticated solution is accordingly second-best, but it is a clear improvement over the myopic solution.

Drawing on reflections about interpersonal coordination in bargaining contexts, however, an alternative to the economist's solution has recently gained adherents, Gauthier among them. Thinking of the pianist ex ante (prior to dinner) as a different self from the pianist ex post (during dinner) suggests a way of conceiving the situation that makes going to dinner but not drinking an available solution. Call this the resolute solution. The resolute solution is available, the argument runs, because it is a better solution for both ex ante self and the ex post self than is the other available solution, namely, the sophisticated solution. It is better than the sophisticated solution because avoiding dinner altogether would give the ex ante pianist his second-best solution and the ex post self his worst solution. The resolute solution, on the other hand, would give the ex ante pianist his best solution and the ex post pianist his second-best solution. And although the ex post pianist would prefer the myopic solution, that solution is unavailable, since the ex ante self has every incentive to avoid it by being sophisticated and skipping dinner altogether. Going to dinner and not drinking is thus optimal. Speaking of the pianist as divided into different time-slice selves is of course intended metaphorically. The resolute solution does not depend on any particular account of personal identity. The point can be put just as well in terms of a unified self: The pianist should regard himself as faring best under the plan that involves choosing to go to dinner and not drinking, since he does better under that plan at each moment in his history than he would do under any alternative plan he might adopt.

It is not clear, however, that the foregoing argument can allay concerns about the resolute solution's availability. At dinner, after all, the pianist will have nothing further to gain from refraining from drinking. The only reason he accepted the plan in the first place was to ensure he would be able to go to dinner. But here he is at dinner, so what is now to stop him from drinking? Unfortunately the pianist will realize prior to dinner that he might come to reason in this way, and this will make him think that the plan of going to dinner and not drinking is infeasible. Against the resolute solution, it might be argued generally that plans calling for actions contrary to an agent's preferences at the moment of execution are not available for a rational agent to adopt. Indeed, we might say that the resolute plan is either infeasible or unnecessary: It is feasible only where the actions it calls for are preferential, in which case it is unnecessary, since a rational agent would have acted in accordance with it even in the absence of a prior intention or plan.

Gauthier and Bratman, who both endorse the resolute solution in this context, account for its availability by specifying conditions under which it is rational for an agent to refrain from reconsidering a previously formed intention or plan. Bratman argues that it is rational to refrain from reconsidering a prior intention in cases that satisfy what he calls the "no-regret condition." That condition asks whether the agent believes that at some time in the future she will be glad she stuck with her intention, and whether she...
believes she will wish she had stuck with it if she does not. In this case the pianist knows, even at dinner, that he would regret abandoning his plan to refrain from drinking and that he would be glad he had stuck with the plan if he does. Assuming that no new information emerges that would cast doubt on the original advisability of the prior intention, the pianist should think it rational to refrain from reconsidering his intention not to drink with dinner.

For Gauthier, it is rational to constrain one’s deliberations in accordance with a prior plan when one can expect to do better under the plan than one would have done if one had never adopted the plan at all.10 When this condition is met at a given time, Gauthier says the plan is “confirmed” for that agent at that time. “Full confirmation” is when a course of action is “confirmed in each possible situation in which it would require some particular action or decision.”11 Full confirmation constitutes a filter on eligible plans: “only courses of action that the agent expects would be fully confirmed are eligible for adoption.”12 Gauthier presumably would regard the pianist’s plan to go to dinner and refrain from drinking as fully confirmed, since the pianist can think of himself as better off under it than he would have been without the plan both in the ex ante and in the ex post contexts, and thus he can regard himself as better off under the plan at each point at which he must make a decision – first when he must decide whether to go to dinner, and later when he must decide whether to drink with dinner.

One way to defend the resolute solution to cases of rational temptation is thus in terms of a two-tier story involving constraint by preferences in the selection of intentions or plans, on the one hand, and a relaxation of that constraint in the context of plan-execution, on the other. This is the defense of resolution Bratman and Gauthier favor.13 Our question will be whether they can resist allowing counter-preferential deliberation outside of plans once they have accepted it in the execution of an existing plan. If, as I shall argue, they cannot resist the extension, they will have reason to rethink their use of a prior intention or plan in cases of rational temptation. If counter-preferential behavior cannot be ruled out by the theory of rationality outside the context of plan-execution, then the fact that the pianist does better in terms of his current preferences would not be sufficient to make his adoption of the plan required by the resolute solution rational. I shall suggest accordingly that the formation of the prior intention the resolute choice theorist advocates is an explanatory fifth wheel in rational temptation cases, since the benefit resolution is supposed to provide turns out to be available without it.

Rationality sometimes seems to require agents to reason counter-preferentially outside the context of a previously formed plan. A slightly modified version of our pianist case will illustrate the point. Suppose the temptation
to drink wine with dinner is sprung on the pianist, such that he must decide whether to drink without the advantage of advance deliberation. This means he must decide whether to drink at the same moment that his preferences most strongly support drinking. What could possibly entitle the resolute choice theorist to say that the pianist should abstain from drinking in the modified case? Since the pianist has no opportunity to form an intention not to drink, a counter-preferential reason to refrain from drinking cannot be derived from a preferentially formed intention to do so, as it might be in the original case. The answer that is consistent with the solution Bratman and Gauthier favor, then, is to say that the pianist should drink with dinner in the modified case. It is not that he lacks the ability to resist wine with dinner, or at any rate, not just that he lacks the ability. It is rather that it would not be rational for him to refrain from drinking with dinner, since there is no reason in that case not to maximize the preferences he has at the moment of choice. In Gauthier’s terms, the plan to refrain from drinking with dinner is not “fully confirmed” in the modified case. The rational choice theorist is thus apparently committed to saying that resisting drinking with dinner would be irrational for the pianist in the modified case.

Now I think this answer is likely to sit uncomfortably with anyone who is not already in the grip of a theory that endorses it. One is inclined to say that whether it is rational for the pianist to refrain from drinking with dinner should not be a function of whether he had advance warning he would be offered wine with dinner. That, at any rate, would be a curious principle for a theory of rational choice to endorse. Furthermore, the rationality of refusing wine with dinner in the modified case is suggested by the fact that the pianist knows he will regret it later if he drinks now. Should we really think it rational for a person to do something he knows he will regret? Bratman in particular might be concerned about a theory with such an outcome. Why, after all, should regret have significance for a rational agent inside a plan but none outside it? Let us then consider whether the resolute choice theorist can account for the rationality of not drinking with dinner in the modified case.

A recent suggestion of Gauthier’s holds out some hope. In quite a different context, Gauthier distinguishes between an agent’s “proximate” preferences, meaning the preferences he has at a given moment of choice, and his “vanishing-point” preferences, the preferences he acknowledges when choice is not imminent. Gauthier suggests that “at any given time, although a person may want to act on his now proximate preferences, he does not want to act at other times on what would be his then proximate preferences, where these are in conflict with the vanishing-point preferences that he now holds.”14 A rational agent will thus realize that he must not deliberate in a way that leads him always to act on his proximate preferences. He must instead adopt a policy of deciding in accordance with his vanishing-point preferences.

The example Gauthier gives is a person faced with a choice of chocolate
or fruit on each of five occasions. Suppose that when the choice is imminent, the agent prefers chocolate, but at any other time he prefers fruit. Gauthier suggests that the agent can regard his choice between a chocolate and a fruit as a choice between two policies or modes of decision-making he might adopt: choosing on the basis of proximate preferences, on the one hand, or choosing on the basis of vanishing-point preferences, on the other. Although the agent prefers chocolate to fruit on each occasion on which he must choose, he does not prefer the mode of decision-making that involves choosing on the basis of proximate preferences, for then he would choose chocolate on all five occasions, and this is something he does not want to do. Indeed, he prefers a mode of decision-making that requires him to choose fruit on all occasions to a mode that requires him to choose chocolate on all occasions, even though he currently prefers a chocolate to a fruit. He should thus choose the mode of decision-making that requires him always to choose in accordance with his vanishing-point preferences, and this will result in his choosing a fruit on this occasion, despite his current preference for chocolate.

Might the resolute choice theorist use the idea of choice over modes of decision-making to solve the modified pianist case? The pianist does not want to choose on future occasions in accordance with his proximate preferences, since he does not want to drink every night prior to performing at the club. He prefers to drink wine with dinner tonight, but he does not now prefer that he drink wine with dinner on each successive evening on which he must face the same choice. He should therefore choose in accordance with his vanishing-point preferences now, and this will enable him to refrain from drinking on this and on all future occasions. Bratman suggests he might endorse this sort of solution as well. In "Planning and Temptation," he says that cases of temptation may be solved by the adoption of what George Ainslie calls "personal rules." A personal rule applied to a series of choices means that an agent can think of himself as choosing "a whole series of rewards at once." Indeed, he claims it a virtue of a planning theory of agency, by contrast with standard utility theory’s outcome-oriented assessment, that it is able to make sense of personal rules of this sort.

While the appeal to personal policies is a natural one for the resolute choice theorist to make, it ultimately proves unsuccessful. The reason is that the pianist in the modified case will have no reason to select the series that requires him to refrain from drinking on all occasions. That would be irrational, given his current preferences. True, the pianist prefers that he not drink on all future occasions, but his preferences with respect to the present occasion support his drinking now. The rational series for the pianist to adopt would thus be drink now and refrain from drinking on all future occasions. The same is true of the choice between fruit and chocolate. I may have a strong preference for choosing fruit on all future occasions. But what is to supply the preference for choosing fruit on this occasion? After all, the sequence chocolate now and fruit on future occasions would satisfy my preferences better than any other available sequence. And if I worry that I would reason that way on each future occasion, and so choose chocolate each time, my fears should be allayed by the thought that on those occasions I will merely be implementing a plan that is already in place. I should therefore be able to resist the chocolate on those occasions. Since my present choice is a choice of which plan to adopt, by the resolute choice theorist’s own lights it seems I cannot choose counter-preferentially, and thus I am constrained to choose the plan that involves chocolate now and fruit later.

What Gauthier seems to have in mind is that a choice over a series of choices is structurally different from a choice on a single occasion. Insofar as the former involves a choice of a mode of decision-making, he thinks there will be constraints on an agent’s choice that there would not otherwise be. Thus Gauthier presumably would want to say that there is no mode of decision-making that corresponds to a choice of chocolate now and fruit on all future occasions. Such a choice would involve choosing on the basis of proximate preferences for this occasion and choosing on the basis of vanishing-point preferences for all future occasions. This does not represent a generalized pattern of decision-making. In this sense, Gauthier’s theory is asymmetrical as between modes of decision-making and plans, since he apparently sees no comparable structural constraints on what can count as a feasible plan.

Two obvious questions present themselves. First, why is choice over a series of choices structurally different from choice on a single occasion? Why, in particular, does it involve a choice of a mode of decision-making? Second, what is Gauthier using to individuate modes of decision-making, and thus what is he using to rule out some potential ‘modes’ from consideration? If the test of the feasibility of a plan is its benefit to the agent who adopts it, as measured by his current preferences, why would we not use the same test for modes of decision-making? Presumably the feasibility of a choice of a mode of decision-making should be the product of a theory of rational choice (as it is in the case of plans) rather than a constraint on it. And since it would be preference-maximizing for an agent to adopt whatever mode allowed him to choose a chocolate now and a fruit on all future occasions, that mode should be feasible for a rational agent to adopt.

If the foregoing is correct, appeal to a series of choices can do no work in the modified pianist case, for it cannot enable the pianist to resist temptation on the present occasion. The only reason the pianist may be able to resist future temptation is that he knows about the previous occasions in advance. That is, with respect to each future member of the series, he is like the pianist in the original case prior to being tempted with the wine. If he knows in advance
that he will be subject to temptation, he can form the intention to resist and then rely on some reasonable approach to reconsideration to avoid abandoning the intention once formed. But if he does not know about the temptation prior to being subject to it, he has no basis for resisting, and so he must choose to eat the chocolate or drink the wine, as the case may be.

Now I want to claim that its inability to deal with cases of rational temptation absent advance notice of the temptation is a major defect of the resolute approach. The rational response to temptation ought not depend on how an agent is situated temporally with respect to his exposure to it. The best that can be said in support of the asymmetric positions of the pianist with advance warning and the pianist without it is that the asymmetry may capture something of our experience in dealing with temptation: We may find it easier to resist temptation if exposed to it after having had an opportunity to prepare ourselves for its onslaught. But even if phenomenologically plausible, this surely does not suggest that an account of the rational response to temptation should enshrine this asymmetry as a virtue. The problem, then, with the resolute solution is that it must treat a rational agent as excluded from resisting some temptations to which she knows she will regret succumbing, temptations it would be rational for her to resist given advance notice. This seems a highly dubious feature of a theory of rationality.

The appeal to modes of deliberation, had it succeeded, would have given the resolute choice theorist a way of justifying decisions that currently are counter-preferential, consistent with the requirement of preferential constraint on plans. It would thus have made preferential constraint on plan-adoptions more plausible requirement, for that requirement would not then have stood in the way of the intuitive answer to the modified pianist case and its ilk. Absent some other solution to cases of this sort, the resolute choice theorist should feel himself under pressure to abandon preferential constraint on plan-adoption. And if he abandons that requirement, the prior intention becomes an unnecessary element of a solution to the original pianist case as well.

That is the intuitive case against the resolute solution to cases of rational temptation. But what if the resolute choice theorist is willing to bite the bullet in the modified case and accept that the pianist has no reason to refrain from drinking with dinner? I shall argue in the next section that he faces difficulties even if he is willing to make this move. The reason is that the resolute choice theorist's account of plan-execution commits him to the possibility that plan-adoption might fail to be preferential. This supplies a second reason to think he cannot have his solution to the original pianist case, since, if correct, the formation of the prior intention would once again turn out to be otiose. I shall further suggest that there are at least two reasons for the resolute choice theorist to favor his account of plan-execution over his account of plan-adoption, and to abandon the requirement of preferential constraint: First, the

modified case is most sensibly dealt with by saying that the pianist should form an immediate, counter-preferential plan not to drink with dinner. Second, preferential constraint on plans can be defended only if one is prepared to accept preferential constraint on both practical judgment and decision. I shall suggest, however, that these are implausible requirements.

It seems natural to think the pianist would reason as follows when confronted with the choice whether to drink in the modified case: "If I drink now, I will regret it, and so I will refrain from drinking with dinner." That, at any rate, would be a Bratman-type thought to have. Alternatively, he might think, "I will fare better overall if I refrain from drinking with dinner," to frame matters as Gauthier suggests. The question is why Bratman and Gauthier think the pianist cannot reason in this way, and why, in particular, he is constrained to reason on the basis of his current preferences.

The explanation must lie in some principle or thesis they accept that rules out counter-preferential behavior outside the context of a previously adopted plan. What is this principle or thesis? First, they might subscribe to a thesis having to do with the nature of judgment. Suppose a rational agent lacked the capacity to judge that she had reason to do a thing if doing it would conflict with her preferences at the time of judgment. Then no rational agent could adopt a plan that was not supported by her current preferences at the time of plan-adoption, since presumably she would have to judge it favorably in order to incorporate it into a plan. Let us call this the thesis of the dependence of judgment on preferences (TDJ).

Second, they might think a rational agent lacked the capacity to decide to do what she judged she had most reason to do if her decision would conflict with her immediate preferences at the moment of decision. Although an agent could judge that she had reason to forgo a certain temptation, she might be unable to allow that judgment to determine her actions, because she would be unable to decide to act on the basis of a judgment that derogated from her preferences at the moment of action. Let us call this the thesis of the dependence of decision on preferences (TDD).

Third, a rational agent might simply be unable to perform actions that would derogate from her preferences at the time of action. That is, although she might be able to judge that she had reason to adopt a course of action that would be counter-preferential at the moment of execution, and although she might be able to decide in favor of such a course of action, she might be unable to translate judgment and decision into action in defiance of preference. Let us call this the thesis of the dependence of action on preferences (TDA). The question we shall ask is whether Bratman and Gauthier might
subscribe to any of these principles to preclude counter-preferential plan- 
adoption. If any one of these principles is correct, they would have a basis 
for maintaining preferential constraint at the level of plan-adopt. I take it, 
furthermore, that there are no other plausible theses by which the rational 
choice theorist might seek to justify preferential constraint. So he must sub-
scribe to one of these three if he is to defend his requirement that plan-
adopt be preferential. The question will then be whether the relevant thesis 
can be adopted consistent with retaining counter-preferential plan-execution.

TDA can be summarily dismissed for present purposes, for clearly neither 
Bratman nor Gauthier can subscribe to it. This follows directly from a feature 
that is common to their solutions: They think an agent can act against the 
weight of his current preferences, as long as he has formed the intention to 
do so in advance. The resolute choice theorist thus assumes the falsity of 
TDA, since he is attempting to offer an account of the very phenomenon it 
declares impossible. Indeed, even the economist believes that rational agents 
can act against the weight of their preferences. He simply arranges matters so 
that the counter-preferential action does not depend on counter-preferential 
rational processes at the moment of execution. It is thus the resolute choice 
theorist’s commitment to deliberative counter-preferential behavior that pro-
vides the thin edge of the wedge: Once he accounts for counter-preferential 
behavior by positing counter-preferential deliberative processes, he is hard-
pressed to account for the necessity of preferential deliberation elsewhere.

Let us now consider whether the resolute choice theorist can explain his 
commitment to preferentially constrained plan-adopt in terms of either 
TDJ or TDD.

We might formulate TDJ as follows:

judge at a time \( t \) that it would be rational for her to perform an action 
\( A \) rather than another action \( A' \) if \( A' \) is better supported by the balance of 
her preferences at \( t \).

It does not matter for TDJ what the agent believes her preferences at the time 
of action will be: TDJ requires only that the content of the agent’s current 
preferences, whether preferences over present or future actions, determines 
the content of the judgment she forms about what she ought to do. Should the 
resolute choice theorist think of judgment as constrained by preferences 
in this way?

The thesis that judgment is constrained by preferences seems implausible 
on its face. After all, judgment is a cognitive faculty, “a faculty of thinking,” as Kant says.\(^\text{19}\) As such, it need make no essential reference to the preferences 
of the judging agent. Moreover, it would be patently absurd to think of 
judgment about theoretical matters as constrained by preferences. Why should 
practical judgment, namely judgment about what an agent ought to do, be 
any more constrained? While I think the idea that judgment is constrained by 
preferences as implausible in practical as in theoretical matters, we need not 
make the case against this view of judgment directly at present. As I shall 
argue, the resolute choice theorist’s view of reconsideration within plans 
commits him to rejecting TDJ, and thus it is not open to him to adopt it.

Both Bratman’s no-regret principle and Gauthier’s account of plan-
confirmation suggest that an agent can judge counter-preferentially in the 
context of the execution of an existing plan, since on both accounts an agent 
must judge that she should not reconsider in the face of current preferences 
that support reconsideration. The agent is to reach this determination by 
judging in accordance with the relevant principle that suggests when it is 
reasonable to reconsider. On one version, the agent judges that she should not 
reconsider on the grounds that she would regret not performing the action to 
which her prior intention committed her. On the other version, she judges 
that she should not reconsider on the grounds that she can see herself as 
better off having performed and executed her prior intention than she would be 
had she never formed the intention in the first place.

It is somewhat unclear whether the resolute choice theorist thinks of the 
relevant judgment as judgment in favor of the action called for by the prior 
intention or whether the judgment pertains solely to reconsideration, that is, 
whether it is a judgment by the agent that she ought not reconsider. But the 
difference between these formulations is irrelevant for our purposes, for either 
way the judgment must be counter-preferential. It is counter-preferential by 
ypothesis if it is a judgment in favor of the action called for by the plan. 
And it is presumably also counter-preferential if it is a judgment in favor of 
non-reconsideration, since the refusal to reconsider is what stands in the way 
of the agent’s maximizing her preferences at the time of action. So both Brat-
man and Gauthier recognize counter-preferential judgment in at least one 
context. The burden is on them to explain why they would disallow it in others.

The resolute choice theorist may wish to argue that the relevant judgment 
is not counter-preferential, since it is only a judgment in favor of non-
reconsideration, and the agent in fact has no preference in favor of reconsider-
ation itself. What she has is a preference for the action she would take were 
she to reconsider. Since she does not know what she would do until she 
reconsiders, he may argue, she has no preference in favor of reconsideration 
*per se*. But this interpretation of the resolute choice theorist’s account strikes 
me as implausible. For each principle suggests that the agent’s decision about 
whether to reconsider is transparent with respect to the action she would take 
were she to reconsider. Transparency is presupposed both by the no-regret 
condition and by the confirmation condition, since whether the agent would 
regret reconsidering or would regard herself as better off under the prior plan
is at least partly a function of the action she would perform after reconsidering, as compared with the action required by the plan. So I think it fair to regard the judgment in favor of non-reconsideration on the resolute choice theorist’s view as counter-preferentially formed.

If this is correct, the resolute choice theorist appears to endorse counter-preferential judgment in the context of a plan at the same time that he denies its possibility in the context of plan-adoption. If he cannot account for the difference between these contexts, he must abandon one claim or the other. If he abandons preferential constraint on judgment in the context of plan-formation, that does not yet commit him to abandoning the idea that plan-formation must be preferential. For he might still accept preferential constraint on decision, and that would provide an independent ground for the restriction on plan-formation. So let us turn to TDD:

*Thesis of the Dependence of Decision on Preferences* (TDD): A rational agent cannot deliberatively decide at a time *t* to perform an action *A* if another course of action *A'*, is better supported by her preferences at *t*.

Commitment to TDD is essential if preferential constraint on plan-adoption is to be maintained, now that we have rejected TDI, since otherwise an agent could judge a counter-preferential plan best and could decide in favor of it. Moreover, it should be clear at this stage that TDD is essential for the claim that plan-formation is necessary for dealing with cases of rational temptation, even apart from the stance one takes on judgment. For it is only because an agent has no resources for deciding against the weight of her current preferences that she must resort to plan-formation, according to the resolute choice theorist. So the resolute choice theorist needs TDD. The question is whether he can have it.²⁰

There are two reasons for thinking the resolute choice theorist cannot accept TDD. The first is that it is not compatible with rejecting TDI. The second is that he is committed to accepting counter-preferential decision in the context of reconsideration within an existing plan, and once again he offers no basis for distinguishing the context of plan-adoptions from the context of plan-execution.

First, suppose the pianist in the modified case formed a judgment of the sort I *ought not to drink with dinner*. If judgment is not constrained by preferences at the moment of judgment, then he could form such a judgment, despite the fact that he preferred to drink with dinner. Suppose we accept TDD. The pianist would then be incapable of deciding to adopt a plan that required him to refrain from drinking, a plan he himself judged best. So a commitment to TDD, without the analogous commitment to TDI, would imply that a rational pianist could not adopt a course of action he thought it best for him to adopt. This sort of split between judgment and decision is generally thought to be the mark of akrasia, an irrational condition. An akratic agent intentionally decides to do something that her better judgment speaks against doing, or intentionally decides not to do a thing that her better judgment speaks in favor of doing.²¹ TDD thus appears to endorse irrational behavior, unless the analogous restriction on judgment is assumed as well. But, as we have seen, there are independent reasons for thinking the restriction on judgment cannot be accepted.

Second, by both Bratman’s and Gauthier’s lights, a resolute agent will need to make a counter-preferential decision in the course of executing a plan that calls for counter-preferential action: He will need to decide in favor of carrying out the prior intention, and that decision arguably will be counter-preferential. Again, we have an ambiguity in this account between the decision to carry out the action called for by the plan and the decision to refrain from reconsideration. But again it looks as though the decision implemented in the context of a plan is counter-preferential either way. If the decision is a decision in favor of the course of action, it is counter-preferential by hypothesis. And if the decision is a decision to refrain from reconsideration, it is counter-preferential too, for we can suppose the agent wants to reconsider, so that he can decide to do what will maximize his preferences. Once again, we can treat reconsideration as transparent with respect to the action performed pursuant to reconsideration, since the agent knows that by reconsidering he would decide in accordance with his current preferences, and this entitles us to conclude that the agent would prefer to reconsider. The resolute choice theorist thus appears to affirm the possibility of counter-preferential decision in the context of a plan at the same time that he denies possibility outside the context of plan-execution. And this creates pressure for him to choose between his account of decision at the level of plan-adoptions and the account he offers at the level of plan-execution.

Now the resolute choice theorist can choose to side with his account of plan-adoptions only if he is prepared to accept both TDD and TDI, since he cannot have one without the other, as I have argued. The economist would certainly favor this solution. For him, judgment and decision cannot come apart from preference, since their connection is assumed in his articulation of the notion of revealed preference.²² Unlike the resolute choice theorist, the economist is thus free to accept both TDI and TDD, and in fact, their acceptance is fundamental to his account. This is why he rejects the resolute choice theorist’s account of plan-execution, and also why he believes in the need for precommitment. My claim, then, is that the resolute choice theorist’s account of plan-adoptions requires him to accept the economist’s position on preferential constraint, since it requires him to accept both TDI and TDD. But, as I have also suggested, the view of judgment and decision that TDI and TDD reflect is implausible. The resolute choice theorist should therefore
distance himself from the economist and abandon the idea of preferential constraint.

The resolute choice theorist might consider two possible solutions to this difficulty, both of which are compatible with the idea of preferential constraint. First, there is a version of resolution that does not accept the asymmetry between plan-adoptation and plan-execution. Ned McClenen, for example, has suggested that the adoption of a plan requiring counter-preferential action produces an endogenous preference change, such that the action called for by the plan will not be counter-preferential at the moment of execution. The suggestion, in short, is that the preferences of a rational agent will support the actions she determines it is in her interest to take. It looks, however, as though endogenous preference change is itself a fifth wheel in an account of rational agency. For the change in preference is supposed to follow from the fact that the action is a rational one to perform. The preferences are not themselves the grounds for regarding the action as rational. The fact that the action is preferential can do no work in recommending it to a rational agent on McClenen's account, and the notion of preferential constraint has not itself been vindicated.

Second, there is a strategy that would allow the resolute choice theorist to defend the asymmetry between plan-adoptation and plan-execution. This would be to adopt a non-deliberative account of plan-execution, such as an account based on habit. Bratman himself seems once to have held such a view. If the decision to reconsider were a matter of habit, instead of deliberation, then the resolute choice theorist would not need to assert the existence of counter-preferential judgment and decision, and thus the difficulty I have pointed to for the resolute choice theorist would not emerge. Notice that the appeal to habit is not far from the economist's sophisticated solution, which is also a non-deliberative account of plan-execution. And as we noted in our discussion of TDA, the device of precommitment allows for the possibility of counter-preferential behavior, without requiring counter-preferential deliberation.

Without fully exploring the defects of such non-deliberative accounts, it may suffice to point out a few of the reasons that weigh against them. First, precommitment, whether of the habitual or the economist's external variety, involves costs. That was the basis for thinking the resolute solution preferable to the sophisticated solution in the first place. It applies no less to those non-deliberative accounts that appeal to the notion of habit, for it may cost something to form a habit, and there may also be costs associated with the manifestation of a habit once formed.

Second, rationality seems to require that agents hold their intentions and plans open to reconsideration, and non-deliberative accounts of plan-execution have the effect of foreclosing reconsideration. The advocate of the habitual account might point out that the relevant habits pertain to patterns of reconsideration itself, and thus the habitual account leaves the possibility of reconsideration open. But the question is whether the account can make sense of an agent's performing a deliberatively based counter-preferential action pursuant to reconsideration. If not, then reconsideration will always lead to the abandonment of the prior intention anyway, and so accounting for it in a given case will not be compatible with accounting for counter-preferential behavior in the execution of a plan.

Finally, and I think most significantly, an account of rational plan-execution seems to require that the action performed in furtherance of a plan be deliberatively grounded. If not, it is hard to distinguish rational plan-execution from what some would think of as rationally induced irrationality. Performing an action called for by a plan through habits of non-reconsideration would be like getting oneself to perform an action by taking a pill. The fact that the action is beneficial to the agent does not suffice to make its performance an instance of rational behavior. For these and other reasons, I do not think non-deliberative accounts of plan-execution will prove successful. But further exploration of these points would be necessary before definitively ruling out such accounts.

I have been arguing that the project of finding a middle way between the economist's full-blown preferential constraint and the conception of practical deliberation as entirely free from preferences is not tenable. The departure from preferential constraint in the context of plan-execution the resolute choice theorist proposes has a tendency to spread to other aspects of rational agency and cannot be contained in the way Gauthier and others have attempted to contain it. This is because freedom from preferences in the choice of actions within a plan establishes the possibility of counter-preferential practical judgment generally, which in turn carries counter-preferential decision in its train. Together, counter-preferential judgment and counter-preferential decision seem sufficient to establish the possibility of counter-preferential deliberation in any context.

Allowing for counter-preferential judgment, along with counter-preferential decision in accordance with that judgment, seems to capture our ordinary thinking about practical reasoning better than accepting thoroughly going preferential constraint in deliberation. It also appears to capture the phenomenology of practical deliberation better than a theory premised on preferential constraint. It is common for a person to feel that she goes to the dentist, for example, because she thinks she should go, not because she has preferences...
in favor of going. It would be odd to think that a rational agent would cancel her trip if she were to reflect on whether to go in the middle of executing a plan to go, or that she would cancel her trip were it not for the fact that she was preventing herself from thinking about what she really wanted to do at the time. She plans to go because she knows she should go, and that is also why she carries out her plan. In such cases, we are most inclined to think not only that judgment is not constrained by preferences but also that an agent’s judging that she ought to do a thing can be sufficient to explain her doing it. These considerations tend naturally in the direction of a cognitive account of motivation.

The rejection of preferential constraint on judgment tends in the direction of a purely cognitive account of motivation, but it does not lead inexorably to one. For the question still remains whether there might be a non-cognitive source of motivation other than preferences, and thus whether judgment, and hence decision, might not be thought of as constrained by it. I am inclined to think that the problems with conceiving of judgment as constrained by preferences would likely apply to a conception of judgment as constrained by any other non-cognitive item as well. But I cannot argue the point here, and I wish to suggest only that the cognitivist solution is an attractive way to think of practical judgment.

The strategy of trying to explain away the counter-preferential character of much of human behavior has a persistent and somewhat surprising allure. The resolute choice theorist’s two-tier approach is only one among many accounts in this vein. The idea of preferences over preferences, or second-order preferences, has been another prominent variant. These accounts rest on a mistaken thought about what is of value for rational agents: The thought is that we must understand practical rationality as preferential if we also want to understand it in terms of the personal commitments of the deliberating agent. Tying practical reasoning to preferences thus seems the only way to avoid an objectivist account of the demands of rational agency. But if that is the basis for attempting to preserve preferential constraint, the motivation misses its mark. For a subjective account of the value of an action or state of affairs for an agent is compatible with a non-preference-based account of rational motivation. Instead of maximizing his preferences, an agent might reason on the basis of what we can call his interests, where these are distinguishable from any set of preferences he currently has. Granted, a person may regard it as in his interest to satisfy certain of his preferences. But others, he may think, are best ignored. The notion of an interest can remain thoroughly subjective, because we can still refuse to count as among an agent’s interests any consideration the agent himself is not prepared to recognize as contributing to his life going as well as possible. The reason it is rational for the pianist to refrain from drinking with dinner is that he recognizes it as in his interest to do so. This need not imply that the importance to the pianist’s life of refusing wine with dinner is determined by his preferences with regard to such an action, even if “importance to his life” is entirely a matter of the importance he attaches.

It counts strongly in favor of resolving the tension in the resolute choice theorist’s account, I have argued, that preferential constraint in plan-execution requires a general commitment to constraint on judgment and decision and that this seems inconsistent with any plausible account of ideal rationality. Worse, preferential constraint on both judgment and decision looks more like a symptom of deep irrationality than like a condition on rational agency. Rational agents should have the ability to act in accordance with their assessments of what they ought to do, assessments reached by considering what would be best for them overall, even in the face of their immediate preferences to the contrary. Moreover, as I have suggested, there is a perfectly plausible account of how practical reason can operate that makes no mention of preferences: A rational agent can reason practically on the basis of more general assessments he makes about what would increase his well-being across his life as a whole. He would then assess particular actions in terms of the contribution they would make to that well-being. And once agents are no longer restricted to the kind of local reflection that preferential reasoning seems to involve, the need for a prior intention drops out in many cases in which the resolute choice theorist posits its necessity, in particular in cases involving temporary shifts in what an agent is inclined to do on the basis of his more immediate concerns.

By contrast, I am prepared to allow that the prior intention or plan is not dispensable in the so-called autonomous-benefit cases, cases where an agent stands to benefit from forming the intention to do something that it will not benefit her to do. Turning to the notion of an interest from that of a preference does not seem to solve the difficulty there. The point is clearest in the toxin puzzle. Under no plausible account of what an interest is can the agent regard actually drinking the toxin as independently in his interest. What is in his interest is intending to drink, and so by hypothesis the prior intention is not dispensable in that case. Less obviously, a prior intention may be necessary in the reciprocation cases as well. Here, the move from preference to interest does not give the person who must perform second any additional reason to make good on his promise to perform. The special features of these cases, however, should not mislead us into thinking of rational agency as requiring an intention or plan every time an action contrary to preference is called for. The excessive focus on the peculiarities of rational agency in the more extraordinary cases perhaps explains why the resolute choice theorist has gotten off to the curious start he has in analyzing the more ordinary problems of rationality the change-in-preference cases raise.
There is an ambivalence in Gauthier's work about preferential constraint, an ambivalence manifested even in those parts of his account where he is at his most "preferential." In _Morals by Agreement_, for example, he specifies that rational agents maximize value only if they reason on the basis of their _considered_ preferences. Preferences are considered, Gauthier says, "if and only if there is no conflict between their behavioral and attitudinal dimensions and they are stable under experience and reflection." A considered preference, unlike the raw desiderative item the economist favors as a source of motivation, is thus a partially normatively constructed item.

In later writings, the distance between the economist's idea of preference as the source of value for rational agents and Gauthier's account of practical rationality increases. In a recent paper, for example, Gauthier acknowledges that there are cases in which a rational agent would recognize that it is better not to enunciate his current preferences in his future life by being resolute with respect to them. Some preference change is to be welcomed, and one would not want to stymie one's personal development by locking in any particular set of preferences for all future occasions. The example he gives is the boy who seeks to bind himself to a way of life in which he will be sure of avoiding the attractions of girls, since he now thinks girls are "yucky," and he is sadly aware that the day will come when he will not find them so. As Gauthier says: "Stepping back from any particular shift in preferences or change in concerns to the recognition that such shifts and changes are a normal and necessary part of human life, an agent may judge that it makes sense to deliberative about her future in a way that leaves herself open to their effects." It is to solve this sort of case that he introduces the distinction between proximate and vanishing-point preferences, along with the claim that it is more rational to reason on the basis of the latter than the former.

That Gauthier has not entirely incorporated the implications of this recent departure from preferential constraint into the body of his account is clear from the fact that cases like that of the pianist could not be solved by resolution if we accepted the foregoing point, leaving aside the problems with resolution I have signaled. For the fact that the pianist can enunciate his ex ante preference for not drinking does not yet suggest that he should on this later thought of Gauthier's. It might turn out that the ex ante preference is one the agent would be better off living without. An effort to reconcile the idea of preferential constraint on plan-adoption with the idea that some preferences are not worth enouncing in one's future life would require criteria for distinguishing those preferences it is rational for an agent to maximize from those he would do better ignoring entirely.

Now Gauthier does present something that might serve as such a criterion. The agent who chooses in favor of his vanishing-point preferences (say, for fruit over chocolate), because he will otherwise be committed to choosing in accordance with proximate preferences on all future occasions, can expect to do better in terms of his "overall concerns," Gauthier says, since he will not choose on the basis of his proximate preferences on all future occasions. This explains why it is rational for him to choose against the weight of his proximate preferences now. But a question about the relation between an agent's "overall concerns" and his preferences naturally arises at this point. If the notion of an overall concern is not limited to current preferences, Gauthier is assuming that practical deliberation can take place from a standpoint outside of preferences. It would accordingly be difficult to argue that preferences by themselves ever constrain our choices and hence that a preferentially formed intention would be required for us to act against them. On the other hand, if the notion of an overall concern is limited to current preferences, it is not clear that Gauthier can justify plans that call for current counter-preferential behavior, as he seems to want to do to accommodate cases of immature preferences. As we saw in the case of the series of choices between a chocolate and a fruit, it is not clear why an agent would ever settle on a plan that required current counter-preferential behavior, in view of the fact that there would always be some other plan he could adopt that would not call for it and so would satisfy his preferences better than one that would. The turn to the notion of an "overall concern" seems an implicit recognition on Gauthier's part that preferential constraint may _not_ function in the way his official view treats it.

More recently still, Gauthier has come quite close to embracing the idea that rational deliberation takes place on the basis of something like an agent's view of what her interests require, where the notion of an interest need not be a preferential one. In the beginning of "Assure and Threaten," for example, he says that the aim of a rational life is _one's life going as well as possible_. Although an agent must "fill in" what it means for her life to go as well as possible, choosing in a way that makes one's life go as well as possible is the maximizing activity in which an agent's rationality consists. Gauthier once again appears to miss the implications the suggestion has for his overall account. For presumably if an agent has a view of what would make her life go as well as possible, she can use _that_ as a basis for evaluating courses of action directly, for settling on plans as well as on isolated actions outside of plans. And there is no reason to suppose that when an agent settles on a course of action she regards as maximally conducive to her life going as well as possible she will be maximizing her preferences at the level of plan-adoption, any more than there was to suppose that in settling on a plan on the basis of preferences the actions that fell under that plan would be preferential. Once one allows for a more global perspective from which an agent
can ask questions about her life as a whole, apart from any particular set of preferences within that life, the local restriction that reasoning take place on the basis of preferences becomes difficult to sustain.

Gauthier himself has thus felt the tendency of counter-preferential deliberation to spread across an account of rational agency: The moves from considered preferences to vanishing-point preferences, to overall concerns, and finally to the aim of one’s life going as well as possible suggest that he has felt the pressure on the idea that reason is constrained that one would expect the recognition of reason’s freedom to provide. One is led to wonder whether the full recognition of this freedom will eventually cast doubt on the understanding of reason as instrumentally constrained itself.

Notes


2. The notion of a preference can be interpreted in a variety of ways. The common use of the term in the philosophical literature makes it a middle-sized object: It is larger than a mere appetite or urge, but smaller than everything an agent takes as valuable or worthwhile. For a general discussion of broad and narrow interpretations of the notion of a preference, see L. W. Sumner, “Welfare, Preference and Rationality,” in Value, Welfare, and Morality, ed. R. G. Frey and C. W. Morris (Cambridge: Cambridge University Press, 1993).

3. The example is from Michael Bratman, “Planning and Temptation,” in Mind and Morals, ed. L. May, M. Friedman, and A. Clark (Cambridge, MA: MIT Press, 1996). I use Bratman’s example instead of the more familiar case of Ulysses and the sirens in the hope that it will help to focus attention on the problem of changing preferences. There is a tendency to see Ulysses’ problem as one of weakness of will, and hence the case as one of ‘irrational’ rather than ‘rational’ temptation.


5. Of course, if it turns out that the ex post self prefers missing dinner altogether to going to dinner and not drinking (he goes to dinner only to drink, let us say), he will not regard himself as better off under the ex ante self’s most preferred solution. In this case he will have no reason to refrain from drinking with dinner, and so the ex ante self will have to choose his second-best option in order to avoid drinking. The optimal solution in this case is the solution that is second best for both, namely avoiding dinner altogether.

6. Economists would say that the solution is not in equilibrium, since the ex post undominated response to the ex ante decision to go to dinner is to drink. The availability of the resolve solution, they think, simply cannot be inferred from its optimality.

7. A plan calling for preferential action may still be necessary in a case in which an agent has insufficient time or resources to determine what the right decision is at the moment of action. A plan for dealing with an emergency, for example, may be necessary if the agent expects to lack the emotional and material resources necessary to determine the correct course of action when the moment for action comes, even if her preferences support acting in the way called for by the plan.

8. Bratman, however, rejects it in other contexts, in particular in the so-called autonomous benefit cases. See Michael E. Bratman, “Toxin, Temptation, and the Stability of Intention,” in Rational Commitment and Social Justice: Essays in Honor of Gregory Kavka, ed. J. L. Coleman and C. W. Morris (Cambridge University Press, 1998), pp. 59, 61–7. He thinks, for example, that it simply is not possible to win a million dollars if to do so one must form an intention to drink a toxin one will lack any reason actually to drink. And similarly he thinks it impossible to enter into a sincere agreement to assist someone later on the condition that he assist you now when his willingness to assist you depends on your intention to assist him, rather than on your actually assisting him. Bratman rejects the resolve solution for these cases, because he thinks it a basic fact about our rational agency that we reason about what to do, and not what to intend. So in a case in which the benefit to the agent arises from forming an intention, rather than from performing an action, an agent can plan in a way that will inure to his greatest benefit only if the action he must plan to take in order to equip himself with the desired intention is one that is independently to his benefit to take. It is not clear why Bratman thinks that change-of-preference cases are distinguishable from the autonomous-benefit cases on this point, for it might equally be said that the gains the ex post pianist enjoys—going to dinner rather than having to miss dinner altogether—are won simply by his intending ex ante not to drink with dinner. In this sense, the intention not to drink with dinner supplies an autonomous benefit as well, at least if one is inclined to view the necessary intention in the reciprocation cases as conferring an autonomous benefit.


12. Gauthier, “Intention and Deliberation,” p. 49. Presumably Gauthier would also accept the converse, namely that every plan that satisfies full confirmation is eligible for adoption by a rational agent.

13. McClennen does not, but there are other problems with his account. Vide infra, part 4.


15. Gauthier allows that there are cases in which the agent’s proximate preferences might overwhelm the value to her of avoiding choosing on the basis of her proximate preferences on all future occasions. These would be cases in which the agent’s proximate preferences surpassed a certain threshold of strength. Above that threshold, the agent would regard no damage to her ability to satisfy her preferences regarding future decision-making as too high a price to pay to satisfy her current preferences. In what follows, however, I shall ignore this wrinkle, since it does not affect our present concerns.

16. Gauthier has indicated to me in conversation that he would favor this sort of approach to the modified pianist case.
18. It is beyond the scope of our present discussion to consider whether Bratman is correct that standard utility theory cannot make sense of personal rules of the sort Bratman has in mind, but I suspect that he may not have given the economists their due on this point.
20. There is a nice argument that shows that Bratman implicitly accepts TDD. Consider a principle we might call the thesis of the dependence of intention on preferences (TDI).

**Thesis of the dependence of intention (TDI):** A rational agent cannot form the intention at a time t to perform an action A if another course of action A’ is better supported by her preferences at that time.

Bratman clearly accepts TDI outside the context of an existing plan: if an agent could form an intention that ran counter to her preferences at a given time, then the pianist could simply form the intention not to drink at the time of action, and this Bratman thinks he cannot do. In order to show that Bratman must accept TDI, I need only show that TDI commits him to TDD.

For Bratman, an agent’s formation of an intention to A is part of his choice of an overall scenario that has A-ing as a “by-chain” leading, in the relation of means to end, to the agent’s goal in acting. In his book, Bratman presents the following principle: “If on the basis of practical reasoning I choose scenario S in the pursuit of intended end E, then I will intend at least one by-chain in S.” Michael Bratman, *Intention, Plans, and Practical Reason*, p. 157. Suppose, then, it was possible for Bratman to be committed to TDI but not to TDD. This would mean that our modified pianist could decide, when he saw his friend in the restaurant, to have dinner with him but not to drink. That is, he could choose the following scenario: *not drink at dinner in order to play well tonight*. There is only one by-chain in this scenario According to Bratman’s principle relating intention and choice, the agent must intend that by-chain. But the agent cannot intend not to drink with dinner, according to Bratman, since Bratman accepts TDI. So Bratman cannot accept TDI and reject TDD, and since he accepts TDI, we must conclude that he is committed to TDD.


22. Strictly speaking, the economist may not be able to think of preferences as constraining judgment and decision after all. This is because he treats an agent as preferring whatever it is that she judges and hence chooses as best for herself (i.e., “revealed choice”). If an agent can be said to prefer to do everything she does, then it looks as though there can be no independent item called a preference to do any constraining. True there are conditions on preferences themselves, such as transitivity, reflexivity, and completeness, and thus an agent’s choices are constrained by conditions of rationality. But this is not itself preferential constraint.

26. This is a failing of other recent accounts of rational choice. Scott Shapiro, for example, argues for a version of resolution that eliminates reconsideration: “The Difference That Rules Make,” in *Analyzing Law: New Essays in Legal Theory*, ed. Brian Bix (Oxford: Clarendon Press, 1998). He suggests, in effect, that an agent should execute intentions automatically, namely, without allowing herself to deliberate about their adequacy at the moment of action. But I would argue that Shapiro’s account is not an account of rational agency per se, because he makes forming an intention or adopting a rule too much like swallowing a pill. It is in my view a *sine qua non* of an account of rational agency, as well as of rule-following, that the intention or the rule be executed deliberatively, namely pursuant to a rational decision at the moment of execution.
27. There might be many reasons why she requires a plan to go to the dentist, other than that she needs to overcome preferences at the time of action in a situation not going. For example, she must coordinate with other people, such as the dentist, not to mention with other aspects of her own life. Bratman’s insight that plans constitute an important part of a rational human life is not at all dependent on thinking plans necessary as a bulwark against recalcitrant preferences.
28. A common objection to an account of this sort is that it rules out any sensible account of *akrasia*. Alfred Mele makes the point thus: “Resisting temptation, Mele argues, is as he judged best. Mary’s behavior cannot be explained by noting that she judged it best to A; for if judging A best (in conjunction with such standing conditions as having the ability to A) were sufficient for A-ing, we would never act akratically.” *Springs of Action* (Oxford University Press, 1992), pp. 122–3. But Mele’s criticism is unclear to me. If an agent is ideally rational, then her judgment that she must A is sufficient to explain her A-ing. The capacity to act on one’s judgment of what it is best to do may be what is meant by calling an agent ideally rational. Akrasia is a phenomenon depending on insufficient rationality. Thus Mary’s B-ing, instead of A-ing, when she judges it best to A, only serves to suggest that she is not ideally rational. It is not clear why Mele thinks that a theory of rational agency should have to explain Mary’s behavior in this case.
29. There will be complications here. For example, what should we do about agents who are unaware that their current motivational makeup commits them to treating a certain state of affairs as desirable? Should we allow that it is in the agent’s interest to bring about that state of affairs?
32. Once again, there is a difficulty in understanding the economist’s position on this point. The economist seems committed both to the idea that preferences constrain deliberation and to the idea that there is nothing more to having a preference than an agent’s willingness to choose in a certain way under certain conditions. The notion of “revealed preference” is in this sense misleading, because the economist does not appear to think that there is anything there to be revealed, and thus the notion of preference itself seems to collapse into a kind of choice, or into a way
of understanding choice. But if preference is really to constrain choice, it must be separately identifiable from it, and this would seem to require the economist to recognize the possibility that choices and preferences might, in a given case, come apart.


34. I am again ignoring the modification of this thesis that Gauthier introduces having to do with a threshold beyond which it would be rational for one to choose in accordance with proximate preferences, since it does not affect the current discussion.