Australasian Journal of Philosophy

ISSN: 0004-8402 (Print) 1471-6828 (Online) Journal homepage: www.tandfonline.com/journals/rajp20

Prudence and the Reasons of Rational Persons

Duncan MacIntosh

To cite this article: Duncan MacIntosh (2001) Prudence and the Reasons of Rational Persons, Australasian Journal of Philosophy, 79:3, 346-365, DOI: 10.1093/ajp/79.3.346

To link to this article: https://doi.org/10.1093/ajp/79.3.346

Australasian Journal of Philosophy
Vol. 79, No. 3, pp. 346–365; September 2001

PRUDENCE AND THE REASONS OF RATIONAL PERSONS

Duncan MacIntosh

I. Prudence and the Desire Theory of Reasons

Once I know my situation, the acts open to me, the possible outcomes of doing them, and the odds of each outcome's obtaining should I do them, what fixes which actions are rational for me? Hume says desires, things person-relative in being able to vary over persons, time-relative in being able to vary in a person over time. My simply having a desire makes it rational to do what I think would cause its end.[1] This is the desire theory of the reasons of rational agents. But Thomas Nagel says the determinants of rational acts are impersonal and timeless reasons.[2] These constitute a standard of goodness in states of affairs, one neither person- nor time-relative, though it may justify a person in an act at a time if her doing it could cause a good state. An act is rational only if it aims to cause the good. One's having a desire can affect which acts are rational only indirectly, by affecting which states would be good. Nagel thinks all action is induced by desire. Only it does not suffice for an act to be rational that one desires to do it. The desire must be 'motivated', a desire to do what one is justified in doing.[3] So the dispute is this: what makes a desire motivated? Hume says the mere having of a desire. For Nagel, timeless reasons fix which desires are motivated. I here extend decision theory and the rational kinematics of values to argue that Hume's view does not have certain absurd consequences Nagel alleges, and that much in Nagel must yield to Hume. In particular, I argue that reasons can't be necessarily timeless.

1. Nagel's Objections to the Desire Theory

Nagel says that in the desire theory one's desires are one's reasons for action; expecting a future reason (foreseeing having a desire) can't provide by itself any reason for present action; any present desire can be a reason for action towards its end.[4] He thinks absurdities then follow: my present desires may give me reason to prepare to do something though I foresee my future desires will give me reason not to do it; I may now have no reason to prepare to do something I know I will want to do later; I may now have reason to try to do something I think my future desires will give me reason to undo.[5] So the theory can justify me in acting imprudently, in not accommodating my foreseen desires. Nagel thinks imprudence irrational, that the theory absurdly entails the rationality of imprudence, that

[1] David Hume, *A Treatise of Human Nature*, second edition, ed. P.H. Nidditch (Oxford: Clarendon, 1978), pp. 413–18, 455–9.
[2] *The Possibility of Altruism* (Oxford: Clarendon, 1970).
[3] *Ibid.,* pp. 29–30.
[4] *Ibid.,* p. 39.
[5] *Ibid.,* pp. 39–40.

in doing this it condones the false belief that my future self is less real than my present, abetting disrespect of my future self and its reasons; and that only the timeless reasons theory can explain how a categorical duty to be prudent—to respect one's future self's reasons—is possible. Nagel doesn't know the timeless reasons' substantive content. But it must be immune to the absurdities; and this may leave but one content, a universalised morality: the reasons must be whatever ones could be those of all agents always could they be those of one agent at one time.[6] I would then be sure to respect the reasons of my future self in acting on my current reasons, for our reasons would be the same.

2. Can the Desire Theory Face a Prudence Problem Reductio?

I say the desire theory is not susceptible to Nagel's *reductio.* The theory makes prudence rational where it *is* rational; and to oblige more prudence would precisely be to disrespect the reasons of persons, those of their present stages in the present, and of their future stages in the future. Moreover, the timeless reasons theory substantially collapses into the desire theory.

A theory of reasons faces *reductio* in a 'prudence problem' just if one's current and foreseen reasons can conflict, the theory would call neither irrational, rationality requires one now to make an accommodation to future reasons, and either the theory absurdly says one needn't do this, or that one should, but can't say which one. Can the theory that reasons are desires have this problem? No; for it embeds limits on the possibility of the tension between a rational agent's current and future values needed to make the problem. I show this by elucidating the structure of values able to be served by instrumentally rational choices, and by deducing from it a rational kinematics of value—a theory of when an agent with such values rationally may, must or must not acquire new values. It emerges that a rational agent's early and later values can't be in certain sorts of conflict; and any remaining conflicts cannot rationalise absurd choices.

3. The Structure of Desires as Determinants of Rational Choices

Decision Theory tells us the structure desires must have to serve as the basis of rational choice—to enable one's acts to be evaluated for instrumental rationality: depending on the conditions under which one is to choose actions, one's desires must have some or all of the structure of well-ordered preferences, ones complete, transitive, acyclic, monotonic and continuous.[7]

If my choices are to be evaluable for whether they serve my values, then if I can make it certain which outcome obtains by which act I do, my values must be a ranking of all the

[6] On this aspect of Nagel, see Nicholas Sturgeon, 'Altruism, Solipsism, and the Objectivity of Reasons', *The Philosophical Review* LXXXIII (1974), pp. 374–402.

[7] See David Gauthier, *Morals By Agreement* (Oxford: Clarendon, 1986), Ch. II. These requirements are contested. E.g., see John Broome, 'Can a Humean Be Moderate', in R.G. Frey and C. W. Morris, eds., *Value, Welfare, and Morality* (Cambridge: Cambridge University Press, 1993), pp. 57–73. To avoid controversy, I specify them not as conditions on rational values, but on values being things in whose service acts are assessable for instrumental rationality. Even should these conditions be wrong, the right ones may yet limit possible conflicts between a rational agent's current and future values. Also, as a referee notes, maybe no one meets such strict conditions. So, lest we deny that agents ever make rational choices, we might say an agent's choices can be assessed for instrumental rationality not *iff* her values meet the conditions, but *in so far as* her values meet them.

states of affairs I think would result of my possible choices from most to least preferred. This is *completeness*: if I leave some states unranked, my values give me no basis of choice among them; for my preferences about them are 'incomplete'. And if each of my possible acts can yield different states, if I haven't ranked all the states, I've no basis for choice among any acts. (I've no basis for choosing any given act with its outcome over the other acts with theirs unless, for each other outcome, I prefer the given outcome to it.) Further, to give me a basis for choice among three acts each with its own outcome, outcome *A*, *B*, and *C*, respectively, my values must be *transitive*—if I prefer *A* to *B* and *B* to *C*, I must also prefer *A* to *C*; and to be able to choose among two acts with outcomes *A* and *B*, my values must be *acyclic*—I may not prefer *A* to *B* to *A*. Otherwise, I both will and won't have reason to do one act over another. E.g., if my values are cyclic, I'll have reason to do the act yielding *A*, because I prefer *A* to *B*, and won't have reason to do it, because I also prefer *B* to *A*. So no choice would be made unambiguously rational by my values.

If I'm to choose where my acts can make outcomes only likely, not certain, my values must also be monotonic and continuous. The *monotonicity* condition says that, given a choice among acts equal in how likely they make their outcomes, if, due to my valuings of outcomes, I prefer doing act *x* to act *y*, I should prefer doing *y* instead should it become more likely to yield those outcomes. I.e., my attitudes to risks must be fixed by just one sort of thing: my valuings of the possible outcomes of taking risks. For if I also separately ranked risk, preferring, say, more to less, I would have conflicting measures of the rationality of acts. If I liked risk, and liked outcome *O*, my liking of risk might recommend an act, *a*, on grounds of *a*'s being unlikely to produce *O*, while my desire for *O* might recommend against *a* due to its lower chance of causing *O* compared to another act. So my all-in basis for choice, being conflicted, couldn't recommend an act.

Continuity says that to be able to make a choice under risk with *A*, *B* and *C* as possible outcomes, if I rank *A* over *B* over *C* there must be odds of winning a lottery with *A* as the prize, *C* the penalty for losing, which would make me indifferent between playing the lottery or just getting *B*. Preferring *A* most, *C*, least, my preferences' strengths are then implied in the utility I'd get for each outcome. My utility for *A* is fixed at 1, *C*, 0, and *B*, the odds I demand (as a fraction of 1) of winning *A* to be indifferent between playing and *B*. My ranking now says how strongly I prefer each state to the one ranked under it, fixing how much risk I should accept in trading the latter for a shot at the former; my strength of preference for each state is the odds I rationally would demand to be indifferent between it and a chance at each other state. So my preference for each outcome is 'continuous' with a preference for some odds of getting each other outcome.

My values are now a function from my beliefs about which states obtain onto utilities. If I think the world is as I most prefer it to be, my utility is 1; if least, 0; etc. My utility function plus my beliefs about the odds of the obtaining of each possible outcome of doing each act open to me, fix an expected utility for each act, the sum of the products of the utilities of each possible outcome of the act and the odds of each outcome's obtaining given my doing of the act. My rational duty is to do the act with the highest expected utility. Absent this structure my values can't fix the rational acceptability of risks—I have no basis for choice among risky acts. But with it, my choices are so evaluable by whether they maximise my expected utility.

Humeanism doesn't hold it rationally obligatory to have some preferences rather than none, nor to have only well-ordered ones. But depending on how your values fail of order,

there can be situations where they afford no way to evaluate your acts for whether they serve your values. So your values must be well-ordered if your acts are always to be instrumentally rationally evaluable.

4. The Rational Kinematics of Values

If one now has well-ordered preferences, these rationally constrain on instrumental grounds, and due to what it is to have preferences, the preferences one may or must form later. E.g., if I prefer *A* to all available *non-A* and learn doing *x* would get me *A*, I should prefer to *x*; while if it would prevent my getting *A*, I should not. For to have a preference is to have a reason for doing what one thinks would advance its target; and the relation between a preference for outcomes and one for acts is that the former, in making it that an act is rational—one you have reason to do—makes the act rationally preferable for you. So if you learn an act would cause your preference's target, you have reason to do it, and so to prefer to do it.

Further—as I argue elsewhere[8]—instrumental rationality can evaluate one's ends once one has ends. It requires one to do what one thinks would cause what one prefers. But in paradoxical situations, to do this for one's preferences, *P*, one must replace *P* with preferences action on which wouldn't advance *P*. Think of a variant Prisoners' Dilemma: suppose others will help reduce my jail time—the aim of my preference, *P*—only if I stop caring only about that and come more to prefer to keep agreements to co-operate, though my breaking them would better advance *P*. Here I'm obliged by my current ends to adopt different ends; my values are 'self-effacing'—retaining them is wrong by their own measure.[9] Or say I want to be attractive to women: then I should drop that want, for males who have it look unattractive; and as dropping it will affect the acts and attitudes of others (will make women attracted to me, ask me out) in ways helpful to me satisfying my current value, I should revise that value.[10]

In general, one rationally must criticise one's values by asking whether one's holding them maximises on them. In paradoxical situations, revising them maximises, because it will induce other agents to have attitudes or to do acts which would advance the targets of one's current values. But where the only effect of one's having one's values is to cause one's own acts, keeping the values maximises, for they would induce one to make their targets more likely, dropping them, less; so here, one shouldn't revise.

[8] See my 'Categorically Rational Preferences and the Structure of Morality', in Peter Danielson, ed., *Modeling Rationality, Morality and Evolution; Vancouver Studies in Cognitive Science, Volume 7* (New York & Oxford: Oxford University Press, 1998), pp. 282–301; my 'Persons and the Satisfaction of Preferences: Problems in the Rational Kinematics of Values', *The Journal of Philosophy* 90 (1993), pp. 163–180; and my 'Preference-Revision and the Paradoxes of Instrumental Rationality', *Canadian Journal of Philosophy* 22 (1992), pp. 503–30. See also David Schmidtz, 'Choosing Ends', *Ethics* 104 (1994), pp. 226–51.

[9] See Derek Parfit, *Reasons and Persons* (Oxford: Oxford University Press, 1984). My views on these situations are more influenced by Gauthier (*Morals*, Ch. 3). He used Decision Theory to study when it is rational to alter one's dispositions to make choices given one's preferences; I use it to study when it is rational to alter one's preferences. This approach lets us be precise on which conditions make it obligatory given one's values to revise the values, and on which revisions are required.

[10] The example is from David Schmidtz, *Rational Choice and Moral Agency* (Princeton, NJ: Princeton University Press, 1995). I see it as illustrating a different point than Schmidtz thinks.

Yet David Schmidtz says sometimes it can advantage us to alter our values so as to manipulate *our own* behaviour.[11] I have the bare desire to live, but get bored and careless single-mindedly pursuing one goal, making me liable to death by predation. To move myself to keep at my goal, I form a new desire—to raise kids. Forming it will likely move me to meet my first goal since, to raise kids, normally I must try to stay alive; but this desire is truly new since it might happen that to save my kids I must give up *my* life, *contra* my first desire. Still, if it maximised for me to get the new desire, even given the risk that it would make me frustrate the old, it was rational for me to get the new.

But it is helpful to revise my values here only because this compensates for my being irrationally bored and careless, things which retard my goal, and which are odd given that I'm pursuing what I care about intensely—surely me having my goal gives me all the motive I should need *not* to be bored and careless.[12] Boredom is appropriate only when I'm not in a position to pursue my goal (that's *why* I'm bored); or when I dislike the means to it—for its repetitiveness, say—though not enough to make me drop the goal.[13] And carelessness is appropriate only for goals unimportant to me. Further, if I'm fully rational, surely if I most want x, I now have all the motivation I need to act to cause x. Why do I need another goal for this? And how, rationally, can I be any more moved to seek x than I am already by most wanting x? So the kids case shows not rationality in value acquisition for a fully rational being, but what it is rational to do to combat one's non- or ir-rational tendencies, e.g., akrasia.[14]

On to another issue. If I have well-ordered preferences, P, may I make P ill-ordered? Often this would be anti-maximising on P, since it would mean choice would be impossible for me, and so choice advancing P; so it would make it less likely P will be advanced, something irrational by P—my rational duty to P is to maximise on P, advance P's satisfaction. But say that, unless I form ill-ordered values, a demon will defeat P, which otherwise have a chance of being satisfied with no further act of mine; then I should form ill-ordered values to stop his defeating P. Or say P are well-ordered desires to form ill-ordered ones; then I should form such values, for this would maximise on P.[15]

We can now see how having preferences gives one rational duties. To have a preference, P, is to have reason to do the acts most likely to cause P's target. This is just to maximise. So to have P is to have a rational duty to maximise on P. And sometimes this obliges one to retain P, other times, to revise it.

Do I assume one can get values simply by choosing to have them when one's prior values make this maximising? No. Only that values respond to reason: if rationality requires my now forming value V, then if I am rational I will form V. But how does this happen? Are values formed by choice, habituation, reward-conditioning? Are they formed

[11] This point, and the example which follows, are from Schmidtz, *Rational Choice*.

[12] I assume boredom and carelessness are evaluable for rationality, and are not, as tori McGeer suggested to me, *non*-rational states, like exhaustion. I don't mean one irrationally *chooses* to be bored or careless (thanks to a referee for the issue); only that it's rationally unintelligible (as in an inappropriate emotion) to be like this in an activity one most ardently prefers to be doing. What decides which states are rationally evaluable in these ways is too big an issue to deal with here.

[13] Here, perhaps I really had two desires, a strong one for the original end, and a weaker desire not to take the means to it.

[14] See note 22, below. Let me acknowledge that Schmidtz is nicely portraying rationality for people as they really are, with irrational and non-rational tendencies.

[15] I owe these cases to a referee.

voluntarily, or non-voluntarily as a direct response to argument? Big topics—for another occasion.[16]

A last issue before applying all this to the prudence problem: surely on my instrumental theory of practical reason, I shouldn't call any values rational. My values, *V*, can make it rational for me to seek to have other ones, *V**, as in paradoxical situations. But that it's rational relative to *V* for me to *aim* to have *V** doesn't mean it's rational for me to *have V**. For once I have *V**, *V*—and its reason-giving power—is gone. Even if *V* is non-self-effacing and so just persists, my having *V* today is neither rational nor irrational even if yesterday it was rational relative to *V* for me to aim still to have *V* today.[17] A value has justificatory power only while held. I'm justified in an act by *V* only while I have *V*, justified in valuing the means to the end I value only while I value that end. If my *replacing* my current value, or my *having* a new one in place of the old, is the means to my current value's end, I should aim to acquire/have the new value; and I can be rational in so aiming because it occurs while I have the old value justifying the aiming. But once I have the new one, the old no longer exists to justify my having the new.

Reply: say the means to the end, *E*, of my current value, *V*, is my keeping *V* (for *V* will make me act to cause *E*, making *E* more likely—my keeping *V* maximises on *V*). Then since I have *V* now, I have what I need to be justified in now having *V*—my having *V* maximises on a value I have, *V*; for it might even now be making me act to cause *E*. Ditto for each moment I have *V*. So even if my holding *V* at *t* can't justify, at later time *t**, my *holding V* at *t** (only my *seeking*, pre-*t**, to have *V* at *t**), my holding *V* at *t** can justify my holding *V* at *t**. While if my now having *V* justifies me in seeking to have *V**, upon me acquiring *V**, *V** justifies my having *V** unless and until events mean *V** justifies my seeking to replace *V**; and so on.[18]

We now have a Humean theory of the rationality not just of acts, but of values. If one lacks preferences, one rationally may acquire any ones. Should these be well-ordered, if keeping them would maximise on them, rationally one must keep them; if revising them would maximise on them, one must revise them (ditto the revisants); and (we may add) if neither course would affect the odds of their ends' obtaining, one may do either. So after one forms preferences, one's preferences are rationally permissible iff derived from the first ones by non-anti-maximising revisions. Acts are rational iff maximising on permitted current preferences. There is no primitive rational duty to have any values *ab initio*. But if one gets well-ordered ones, having them creates duties to keep some, eschew others, and to replace some with others, all as part of one's rational duty to maximise on one's values. So if one has well-ordered values, that makes false that just any are motivated—able to make acts rational—given those had. Any initial well-ordered values are motivated; but they delimit what others can be so. This is a Humean theory, however. For the basis of rational choice is one's values. But one's having them delimits what other ones one rationally may acquire.

[16] See my 'Persons'; and my 'Preference-Revision'; also Schmidtz, *Rational Choice*; and Eric M. Cave, *Preferring Justice: Rationality, Self-Transformation and the Sense of Justice* (Boulder, CO: Westview, 1998), Ch. 9.

[17] Thanks to a referee for this issue.

[18] Thanks to tori McGeer for discussion.

5. The Rational Constraints on Changes in One's Values, and the Immunity of the Desire Theory to Prudence Problems

We may now expect consistency between a rational agent's current and future values, and so compatibility between the choices rational given her current values and those rational given her future ones. For a person rational in acquiring new values given her current ones, there will be tension between her current and future values only when it becomes irrational for her to hold the former. The permissibility of one's values is time-indexed: it can be permissible to have one value now,[19] obligatory to form a new one later. But since the values permissible to have change, never would rationality require that one respect—choose from or take into account in choices—both values at once. If it is obligatory at time *t* to acquire new values, it ceases post-*t* to be rational to respect the old. They don't have a timeless claim to inform acts. So the desire theory can't have a prudence problem.

Depending on circumstances, the following relations between old and new values can meet the demand that the new be rational successors to the old:

a) the values are the same (I'd rather collect peaches than rocks now, ditto later). Since one's values now and later are the same, there's no conflict between them, and so no prudence problem.

b) one's new values are for things valued because they help advance old values (I now prefer collecting peaches; I later learn where the peach and rock markets are and prefer to go to the former since I can then indulge my peach preference). Since the things newly valued are things whose obtaining advances the old values, any tension between the values is due only to poor information about the causal relations between their targets, not to conflict between the values as bases of rational choice. But what if the thing later valued wouldn't advance the thing earlier valued until later, so that it would be wrong, given current values, to value it earlier? Won't there be conflict between current and later values until when it will be right by the measure of the earlier to form the later? No. For the later have no independent grip on rationally correct behaviour. The only thing making it right to form and act on the later values is that their targets will then advance the earlier values. So the later ones don't contend for being a basis of choice until it's time to form them. So their proper role as choice basis doesn't conflict with that of one's current values. On any account it should be irrational to advance them prior to when forming them is mandated. I want on Friday to be in Montreal Monday; to get there I must board a jet Monday; so on Monday I'll want to board. But that desire isn't one I must have or act upon on Friday—not till Monday.

c) one's new values are for things implicated by logic or identity in what one earlier valued (I prefer peaches; I learn these are a fruit; 'OK', I think, 'so I prefer a kind of fruit'). There are no conflicting demands from successive values; for the new values are just for things implicated by logic or identity in the old. This may not even be a value change since if one values *x*, and if *y* is identical to or entailed by *x*'s obtaining, maybe one values *y* too; but it may take time and information to see this. In any event, just as in *b*), where, once one learns *x* would cause satisfaction of a current value, one will value *x*, so here, when one learns it's logically part of one valued thing, *x*, that thing *y* obtain, too

[19] Permissible just in this sense: forming the current value was not anti-maximising on any immediately prior values one had.

(or learns the obtaining of *x is* the obtaining of *y*), rationally one will come to value *y*; so there is no tension between the old and new values. For the basis of (the motivation for having and acting on) the new values is the old values.

d) one's new values are rationally permissible given one's old ones, *V*, for revising *V* wasn't anti-maximising on *V*. (Once I preferred collecting peaches. But they became extinct; those I collected rotted away. So my preferring peaches ceased to make it more likely that I'd have peaches. So I was free to form new values without irrationality relative to the old. Perhaps I now want to be a philosopher; and in the circumstances, replacing my old want with this one isn't irrational.) Since the values are such that holding one now and another later are both rationally permissible—because replacing the old with the new wasn't anti-maximising on the old—they don't conflict as choice bases. There is nothing now the old values would have had me do that conflicts with what the new values dictate. For here there is no longer any act which would advance the old values. So, no prudence problem.

But two possible relations of rational succession among values may pose problems, *e*) and *f*):[20]

e) one's later value is one whose possession causally advances the thing valued in one's old value; so possessing the new is rationally mandated by the old. So-altering one's value is rationally required only where satisfying an old value is made more likely by acquiring the new. This holds only if rational action on the old and new could differ; for values differ just if, in some possible circumstances, different acts maximise on them. But then action on the new could defeat the end of the old, even if its *acquisition* would not. Think of a variant Prisoners' Dilemma[21] where it advances one's old value to replace it with one whose possession will likely make other agents advance one's old value's end, and yet which may motivate one's self to act against that end. Adopting a conditionally co-operative preference may induce other agents to co-operate, reducing one's jail time, the end of one's old value; but the new one would have one co-operate in turn, when it would have been better by one's old one to defect. So it seems there can be conflict between old and new values as bases of rational choices, and so a prudence problem.

But there is a problem only if at some time, rationally both the old and the new values should be advanced. And this is false if it was rational to replace the old with the new. Before events mandated the change, it was rationally obligatory to advance the old values; after, the new. To insist one now has a rational duty to both is to fail to recognise the duty to have altered one's values, and to fail to see that this changes what should serve as the basis for choice. It disrespects the old values because it insists that one not do what they justify—form a different choice basis; and it disrespects the new because it says that though one should have formed them, one shouldn't choose by them.

Lastly, *f)*: say my biological maturation will *inevitably* replace a value I now hold, *V*, with a value, *V**, whose satisfaction precludes *V*'s. *Prima facie*, *V** isn't rationally

[20] There is another possibility, *g)*: one's earlier and later values were for things objectively valuable; but what had value changed; so rationally one's values had to change. But this is no part of Humeanism, where nothing has objective value. On *g)*, see my 'Moral Paradox and the Mutability of the Good' (unpublished MS, Dalhousie University, June, 2000); and my 'The Prudence Problem' (unpublished MS, Dalhousie University, October, 2000).

[21] See my papers, above, for more on value-change here.

succeeding *V*; for it would be anti-maximising on *V* to replace *V*, as that would make it less likely I'll do acts maximising on *V*; so it would be irrational for me to trade *V* for *V** had I a choice.[22] So on the desire theory, *V** has no claim to base the rationality of my current acts, no claim to accommodation. Nor should my current self refrain from advancing *V*; for that too would be anti-maximising on *V*. Yet if ought implies can, it can't be that I oughtn't to form a value I can't but form; so forming *V** can't count as irrational, only non-rational. And once I have *V**, it bases my rational choices. Thus the desire theory calls both values rationally permissible: *V* because (let's say) forming it wasn't anti-maximising on my prior values, and because keeping it is obligatory once it is formed, since that would induce me to maximise on it; *V** because it is fated and so can't be condemned. But this doesn't make a prudence problem. Rationality doesn't require me now to advance *V** just because I will have it; for that would be anti-maximising on my current value, *V*. Nor, once *V** replaces *V*, does rationality require me to serve *V* just because once I had it; for that would be anti-maximising on *V**. Instead, I should advance *V* earlier, *V** later; never am I bound by both.[23]

Still, that I shall inevitably non-rationally form a value *can* sometimes affect which choices are rational given my current values; but only in ways provided for in the desire theory:

i) I might now have a 'prudent' preference, one to satisfy a future value (I might now prefer that my foreseen craving for a drug be met). If so, I have reason in my present values to arrange to satisfy my future one (reason now to get the drug for then). And if I now wanted that life go well for my future self—a general prudent preference—I would have reason now to arrange to meet those fated future desires meeting which equals his life going well. Here, satisfying a future value is logically essential to satisfying a present one. For satisfying a prudent preference entails satisfying the future value over which it ranges.

ii) Satisfying a future value may be causally essential to satisfying my present values, in which case I have reason in them to arrange to satisfy it. Maybe I now want to write later; but to do that, I'll need to have satisfied the desire I'll later feel to eat; so though I'm not now hungry, I now have reason to get food so I'll have it when I later desire to eat before work.

iii) Say I now prefer most a state incompatible with one my future self will prefer. If the fact that he will have his value makes it more likely his state will obtain, because his

[22] Suppose it isn't inevitable that I'll loose *V* to *V**, but I will unless I reinforce my holding *V*: then, for the reasons just noted for its not being rational to trade *V* for *V**, I should reinforce. (See Schmidtz, *Rational Choice*.) But on my view, I'm irrational should I need reinforcing. If it maximises on *V* for me to keep *V*, rationally I should automatically keep *V*. (See note 14, above, and the associated main text.)

[23] Or maybe one should still act on *V* if one can since *V* was the last value one rationally held; and one should maximise only on rational values. Acting on *V** isn't acting rationally *tout court*, only relative to irrational values. In the text, I portray rational action as maximising on present values, just as the desire or present-aim theory is standardly thought to require. But since I think the theory can criticise desires, maybe it wouldn't always say to maximise on present values; one should act on *past* ones if altering them wasn't rational. (One is still never bound now by both current desires and ones held earlier or later.) See my 'The Prudence Problem'. Yet what of desires by which one is normally beset as one matures? Is acting on them rational, since they are one's new present desires, or irrational, since forming them didn't maximise on one's old values? A big issue.

value will make him try to cause it, that lowers the chance that an act of mine now could cause my favourite state; so the expected utility of some other act of mine now, one with a higher chance at causing a state I now prefer less, may be higher; so I should do *it*. And my retreating to an act likely to yield a state ranked lower in my current values has the effect of preparing—or at least not obstructing—the way for my future self meeting his want. Here, the fact that I will later have a certain value, *V**, alters the odds that acts of mine done now would advance my current values, *V*; and this is a strategic factor in fixing which current choices are rational given *V*. *V** is relevant to my current choices not as a reason for them, but as a force to be considered in choosing given my current reasons as such. My current values are my reasons *qua* determinants of which things I want for their own sakes, or as means to them: my now having the values gives me reason now to act to attain their ends. But I do not now want the end of my future want. Still, the want will make me act in certain ways in the future; it's like a force of nature I must begin working with, or around, to get what I now want. I now most want a career needing a high commitment, second-most, one needing a medium commitment, third, a series of temp jobs. But I foresee being gripped by a desire to have kids. Since that desire will likely ruin a demanding career, I'd more likely succeed in a medium-commitment one; and better that than having to take temp jobs mid-life. So I should train for the less demanding career instead of the one I now most prefer.[24]

Summary: in the Humean view a rational agent's old and new values might conflict. But she will have new ones only when and because it was rationally required or permitted, given her old ones, for her to change values. So it will then be rational for her to have dropped her old ones. So they then have no rational claim on her; she should choose by her new ones. But for as long as she's required to have her old ones, her future ones aren't yet rational for her to form; so they have no claim on her current conduct. So the conflict between values can't make a prudence problem; for it is right at each time to choose only from values then held,[25] false that one rationally must choose now accommodating also foreseen reasons. An agent might be fated to form a value, *V**, non-rationally. But as her current values forbid forming *V**, *V** has no rational claim now to supplement them as a choice basis, as if she had both now and had to choose given that they conflict. Her current wants are things she should consider *qua* reasons; her propensity to acquire *V** is more like a psychological frailty she must deal with to get what she now wants.

II. Timeless Reasons–Decision-Theory and the Very Idea

6. Limits On the Role of Timeless Reasons

Nagel thinks timeless reasons (the goodness of states acts might yield) fix which acts are rational (those causing good states), and that acting to cause a state entails desiring so to act, or desiring that state; so the reasons fix which desires are motivated, i.e., rational to have (ones for good states), what it is rational to desire (good states).[26] But if Part I was

[24] Thanks to Schmidtz for the example, and for help explicating what it illustrates.

[25] Provided forming them was permissible given prior values, at least depending on how we resolve the issue of note 23, above.

[26] Or at least a desire isn't decisively motivated—*qua* justified—if it is inconsistent with what timeless reasons require.

right, my rationally now having a desire affects what I rationally can desire later. Say the reasons ask me now to do an act, and so to have certain desires. If I form them, they constrain what desires I rationally can have later. But then the reasons can't later demand desiring anything whose desiring violates those constraints.

Maybe they won't. Say state *S* would be good at *t*. Then, *per* Nagel, any time pre-*t*, there is timeless reason for me to desire to bring it about that *S* holds at *t*. Call this desire '*d*'. Surely *d* would rationalise my forming only desires for states identical to, part of, or means to *S*. And such desires don't *conflict* with timeless reasons to cause *S*; for *S*'s goodness would also justify my desiring things identical to, part of, or means to *S*.

But recall *e*): sometimes one's desiring a state justifies one's forming a desire for its opposite. Call these Self-Effacement Situations—*SESs*. E.g., consider Gregory Kavka's Special Deterrence Scenario (*SDS*):[27] you lead the U.S., you desire harm-minimisation—a motivated desire, dictated by timeless reasons, since the fewer harms in a state of affairs, the more good it is. But likely the S.U. will launch a nuclear attack, killing all your people, unless you form desires which would justify your retaliating; then likely S.U. won't attack. (To deter you must alter your desires; for S.U. can detect failures of genuine resolve.) So to have a chance of saving half the world, you must form a desire that would make you kill the other half were you attacked, a desire to meet attacks with harms—useless ones, ones not means to preventing greater harms; everyone will be dead after you retaliate, so *retaliating* isn't deterrent, only *desiring* to retaliate. Here, your desire, *d*, for the good, *S*, rationalises you in forming a desire, *d**, for the bad, *-S*; and if your gambit fails, *d** will justify you in causing *-S*. But *d**'s target isn't identical to, part of, or a means to *S*. Yet there is reason to form *d**: *forming d** is a means to the good state, *S*, targeted in the desire, *d*, the reasons first dictated. And surely it is rational to take the means to the end of one's desire for a good state.

Here, the desires one is justified in forming by those first dictated by the reasons aren't ones the reasons find rational—motivated: since *S* remains good, the reasons would still find desire *d* for *S* motivated, *d** for *-S*, not. But *d* justified one's replacing *d* with *d**. So, *is d** rational, motivated? Nagel must say no, else reasons aren't timeless, but vary with the desires rational to have over time. He must say desires are motivated only if directly dictated by timeless reasons. So one's motivated desires aren't a function of one's previous desires, but of good states. Thus Nagel would reject my model of how motivated desires rationally evolve. I say a motivated desire can motivate later desires. If the reasons ask one to have desire *d* for state *S*, thus making *d* motivated, *d* can motivate desires for states identical to, part of, or means to, *S*, *or desires for states the desiring of which* is a means to *S*. But Nagel would dispute me on when a desire is motivated: he would say reasons stay timeless because desires can't, independently of timeless reasons, justify acts (make them rational), nor motivate later desires. Some of the desires I said *d* motivates, *are* motivated—but by *S*'s goodness, not by *d*: one should desire states identical to *S*, parts of *S*, or means to *S*, because, and only because, they are identical to a good state,[28] or are

[27] Gregory Kavka, 'Some Paradoxes of Deterrence', *The Journal of Philosophy* LXXV (1978), pp. 285–302.

[28] So if one learns description *P* fits the same state as a description one now prefers to be true, one should prefer *P*'s truth, too.

parts of one,[29] or means to one.[30] But a desire whose formation is justified by advancing a desire for a good state isn't motivated if it isn't a desire *for* a good state; for it isn't directly motivated by goodness in its target. So the mere fact that there is practical reason to have *d** because this would advance a motivated desire doesn't make *d** motivated.

For Nagel, then, desires are rational iff they target good states. Where such desires rationally oblige forming opposite desires, this just means one should become irrational, form irrational desires. True, our standard for rational acts (the timeless reasons ranking act outcomes) can't then always control acts; for once one internalises it in desires for good states, it here asks that one drop it, whence it stops ruling one's acts: if I rationally had to do the act of revising my desires, I now have desires that could make me cause bad states, *contra* the standard. But the standard—an act is rational iff it advances desires for good states—still determines whether my acts are rational.[31]

But this can't be right. For if the reasons rationally obliged me to revise my desire, and also timelessly measured desire rationality, calling the revisant irrational, there could be an act—forming a desire—on whose rationality the reasons would be ambiguous; so they couldn't always fix rational acts. They would violate monotonicity (defined above), giving competing tests of rationality: one says an act of revising desires is rational if dictated by desires dictated by timeless reasons, the other, only if the revisants target states good by timeless standards.

Nagel would be holding that, though the desires first dictated by timeless reasons rationally obliged one to replace them with ones for bad states, one rationally must do the act needed to cause states good by the reasons' timeless measure. Yet on his own theory, one can act only if one has a concordant desire; so if one has a rational duty so to act, one also has a duty to have (kept) the desires the reasons first required. So: one should form the desires first required, and these demand self-replacement; but one retains a duty to have the originals. But now the theory is conflicted on which desires one should have, and so on which acts (of forming desires) to do, and so on which acts there is reason to do—the theory of reasons is non-monotonic.

Monotonicity constrains the rationality of desires to act given desires for act outcomes: the rationality of a desire to act is to be decided solely by one's ranking of outcomes, not also by an independent ranking of acts, say an affinity for risk. For had one two measures of act rationality, i.e., two preference-rankings, one for outcomes, the other for act features, one's basis for rational acts would be conflicted and so couldn't determine their rationality. I am extending this concept, taking as its core that the rational preferability of acts must be fixed by one ranking only. For Nagel, one's motivated desires are to be a ranking of the goodness of outcomes; rational acts are ones that cause the best outcomes. So if the act of revising one's desires for good outcomes does this, it's rational so to act. But if one also has an independent ranking of desire rationality in which desires are rational iff they target good states, one's measure of act rationality is non-monotonic,

[29] Unless the other parts needed to make that good state won't obtain and it would be bad if only part of it obtained.

[30] Assuming it isn't worse for the state to come about by such means than not to come about, or to come about by less bad means.

[31] A referee put the idea this way: surely a theory of reasons (e.g., reasons are desires for good states) telling us to accept another theory (reasons are the desires whose adoption would advance the former desires) can yet be the truth.

giving a conflicted choice basis that can't always determine act rationality. For in *SESs*, while, by the first measure, it is rational to do the act of forming a certain desire (as this will cause a good state previously desired), by the second measure, it is not, because the desire is for a bad state, and so the second measure forbids its formation.

The problem arises because Nagel's theory rates the rationality both of desires—they should target good states—and acts—acts should advance rational desires. And where the act advancing such desires is revising them, the theory of rational acts doubles as a theory of rational desires conflicting with the explicit theory. A desire can't be made rational to have *solely* by its target state's being good, *and* by the desire's adoption advancing a desire for a good state; for in *SES*s it advances a desire for a good state to form a desire for a bad. Since this is mandated by a motivated desire, by one of Nagel's standards you must do the act needed to form the new desire (see a hypnotist, say); but since it is a desire for a bad state, by his other standard, you must not. So his theory of act rationality is non-monotonic. Obversely, his theory of rational desires implies a theory of rational acts: the act of forming a desire is rational iff the desire targets a good state. Yet by his explicit theory, an act is rational iff it advances a desire for a good state; and in our SES, this asks me to do the act of forming a desire for a bad state, since that advances a desire for a good one.

Yet maybe a desire's rationality is separate from whether forming it is rationally obligatory; so Nagel could say: there is one standard for desire rationality—only desires for good states[32] are rational; and one for act rationality—only acts advancing such desires are rational; but where it advances such desires to form irrational ones, forming them is obligatory. This might make Nagel's theory monotonic: what makes an act rational is solely that it would *cause* good states—so the mere fact that a state was good wouldn't make obligatory the act of forming a desire for it. And what makes a desire rational is solely that its target state is good—so the mere fact that forming a desire for a bad state would cause a good state wouldn't make the desire for the bad rational. So there are no competing standards making an act both rational and irrational, nor a desire either.

But Nagel thinks it rationally obligatory to act to cause good states, and that acting entails desiring. But then it's obligatory to have the desires needed to induce rational acts—desires for good states; so one should form them. Yet Nagel also thinks it obligatory to do the act advancing a rational desire. So if, as in *SESs*, that act is forming irrational desires, it is there obligatory to form *them* instead. So Nagel's theory of rational acts is non-monotonic on which acts of forming desires one should do;[33] one should do the acts of forming the desires for good states needed to induce the rationally obliged acts of causing good states; yet in *SESs*, the obliged act is forming desires for *bad* states, since this would advance desires for good ones. So calling the desire for a bad state inherently irrational won't yield monotonicity on which acts of forming desires one should do. Nor will it help to say: 'there are two ways a desire can be justified, a) by its targeting a good state; b) by its formation's advancing such a desire. a) makes a desire inherently rational; b), acquiring one, pragmatically rational. Acquiring one on b)'s pretext is rationally

[32] Or for states identical to, part of, or means to them.

[33] Even if forming a desire isn't an act, there are acts that would *result* in one's forming desires (taking a pill, seeing a hypnotist, hanging with a different crowd); and Nagel's theory violates monotonicity on the rationality of *these* acts.

motivated irrational desire acquisition.' This won't help because in *SESs* one is pragmatically justified in forming a desire whose rejection is inherently justified; so, for different reasons, one should and shouldn't form it—the theory of reasons remains non-monotonic.

This sinks any theory saying rational acts are ones justified and induced by attitudes timelessly rationally required by timeless goodness in states: for say state *G* is good. Then supposedly *G*'s goodness asks me to have a pro-attitude to *G*, one suited to inducing and justifying my acting to cause *G*. It might be a conative attitude like desiring *G* (Nagel); or a cognitive one representing *G*'s virtue, a belief, perception or intuition that *G* is good or that *G* should be procured (as in Kant or moral cognitivism: a rational agent is ruled by reason, by her beliefs about which acts are apt, which states, good). But for any attitude, *A*, which *G*'s goodness may ask, there can be a *SES* where, to cause *G*, I must revise *A*. The theory says it's rational to do what would cause *G*. So I should revise *A*. If the theory approves the revisant, *A**, it is time-relative which attitudes are rational; but if the theory asks me to keep *A* (as it does if it calls *A** irrational) *and* to cause *G*, it is non-monotonic on rational choices. For the duty to cause *G* conflicts with the duty to have attitude *A* to *G* in *SESs* where, to cause *G*, one must drop *A*.

Maybe Nagel shouldn't say that, to be rational, one's acts must be justified and induced by one's attitudes.[34] For his main idea is that the ultimate reasons for rational acts are the goodnesses of states. So maybe acts are rational iff they cause good states; and timeless reasons—state goodness—can always monotonically decide act rationality if they rate acts directly by whether they cause good states, not indirectly by whether they are induced by pro-attitudes to good states.

But this is problematic given what acts are. Action Theory says my acts are behaviours of mine induced by my *reasons*. Two theories of what it is to have a reason are germane, internalism and externalism. On internalism, to have reason to do *x* or to bring about *y* is to have a pro-attitude to doing *x* or to causing *y*,[35] one tending to induce act *x* or action to cause *y*, e.g., a belief that *x* should be done, *y*, caused, or a desire to do *x*, cause *y*. If the reason is a belief, it entails a motivating desire to do *x*, cause *y*, or it itself motivates. So acts are behaviours induced by pro-attitudes (conative or cognitive) to the behaviours or to their yields. Thus any theory rating acts for categorical rationality also rates attitudes: if it's categorically rationally obligatory to act to cause good state *G*, I should form a pro-attitude to *G*, one to make me cause *G*.[36] And if I shouldn't act to cause bad state *B*, I shouldn't form a pro-attitude to *B*. But say *G* is good, so I should act to cause *G*, and so should form a pro-attitude to *G*. Say I do; but now say I face a *SES* where, to cause *G*, I must form a negative (con-) attitude to *G* (or one pro-*B*). (Forming it causes good states independently of effects of acts the attitude may cause.) The theory of required acts, since it implies a theory of required attitudes, is non-monotonic. For it says I must act to cause

[34] Even Nagel may be reluctant to reify desires as causes of actions, and so to think rational acts are ones justified by motivated desires. True, he thinks all action induced by desire. But he sometimes talks as if to say a behaviour was desire-induced is just a way of saying it was an act—*trivially*, all action is induced by desire (and so all rational action, by a motivated desire). On this reading, see Jonathan Dancy, *Moral Reasons* (Oxford: Blackwell, 1993), pp. 7–9.

[35] Or it is to have a pro-attitude practical reasoning from which justifies pro-attitudes to doing *x*, to causing *y*.

[36] In Humean instrumental rationality, by contrast, it is rationally obligatory to cause *G* only if one desires *G* as an end or means. Still, if one's behaviour, *B*, causes *G*, *B* is the act of causing *G* iff justified and induced by a desire one has for *G*.

G, and so must form a pro-attitude to *G*; but where the act needed to cause *G* is forming a con-attitude to *G*, it also says I must instead form a con-attitude to *G*. Since forming it would cause a good state I should form it; but since it is a con-attitude to a good state (or a pro-attitude to a bad), I should not.

On externalism, I can think I have reason to do *x* or to cause *y*, yet have no motivation to do *x*, cause *y*. But my having reasons only externally can't yield acts. However, when I have an external reason, maybe it's also true that I should internalise it in a motivating attitude (I have external reason to internalise the former reason): if my reason rationally obliges me to do act *a*, it also obliges me to have an attitude which would, by inducing a certain behaviour, yield *a*. But then even on externalism, evaluating my attitudes remains implicated in evaluating my acts; so *SESs* will still make trouble.

But maybe what makes a behaviour of mine, *B*, a rational act isn't that *B* is induced by my rational attitudes, but by a brute disposition to *B* iff *B*-ing would yield good states. Yet what if I am unaware of my *B*-ing, or think it irrational, think it will cause a bad state? Surely my *B*-ing is an *act* only if I'm aware of it, represent it as a thing I've reason to do (e.g., as causing good states) and do it for that reason; I must be *B*-ing *because* I think it apt to do so, this marking off things I do for reasons from *mere* behaviours. But to represent *B* as apt is to have *B*, or *B*'s yield, as the satisfaction condition of a representation, *R*. And in *SESs*, to cause *B*, I must alter *R*—our old problem.

Even if my behaviour is a rational act if caused by a disposition of mine justified by tending to yield good states, to say I have external reason to cause states due to their goodness would be to say I should be disposed to cause them. But then a *SES* can arise where, to cause them, I must revise my disposition into one to cause bad states. So the theory of which acts are rational will imply a theory of which dispositions are rational; and the theory's two aspects will conflict for *SESs*.

Nagel's theory is non-monotonic because it posits a *timeless* standard for rational desires and acts: acts are timelessly made rational to do by their yielding timelessly good states, desires, timelessly made rational to do the act of forming by being desires for such states. But in *SES*s, the act which would yield good states would be forming desires for bad states, dropping desires for good. So one shouldn't and should form desires for bad states—shouldn't because the states are bad, should because forming the desires would yield good states. How can we recover monotonicity in a theory of reasons?

Well, maybe when a desire you were rationally obliged to have by the goodness of the state it targeted is best advanced by you forming a desire for a so-called bad state, *B*, *B* becomes good for you—goodness is relative. Since *B* is now good, you meet the duty to form desires for good states by forming the desire for *B*. So the desires state goodness demands will be the same as those advised for their formation's advancing desires for good states.

This saves a link between the rational duty to have a desire and the goodness of the state it targets, and so between the rational duty to do an act and the goodness of states the act might cause; and state goodness would oblige one to do the same acts as those required by the desires one was rationally obliged to form by the desires state goodness first asked one to form. But if the new desires oppose those state goodness first dictated, the acts the new desires demand can be the same as those state goodness, by then, demands, only if state goodness *changed* with rationally required changes in desires; so if a rational agent's reasons would be the goodness of the states her acts might cause, her

reasons would be time-relative. The goodness of states acts might cause could still be what makes acts rational; but since the states one is rationally obliged to desire are different after rationally required change in desires, the standard of state goodness must have changed concordantly. So it isn't timeless, but time-relative; but then so (*contra* Nagel) are reasons.

I think goodness does mutate like this in *SESs*;[37] but most philosophers would disagree. So say it can't mutate. Then, *contra* Nagel, the (timeless) goodness of states can't necessarily always properly be our reason for action. For it is the very timelessness of Nagel's standard of rationality that yields non-monotonicity. So we must say this: if an act is rational just if induced by a motivated—rational—desire, the reasons for acts won't necessarily be state goodness, Nagel's 'timeless reasons'. Instead, they must be the desires appropriate to form given desires first dictated by 'timeless' reasons. But then if one now has desires recommended by the reasons, and so now has rational desires, this constrains what desires one rationally can have later (recall Part I). And if one's motivated desires, i.e., those rational for one to have, are the only reasons able to make one's acts rational (as they are if acting entails desiring, rational acting, acting from rational desire), then one's *effective* reasons—those able rationally to be one's reasons for, and able to result in one's doing, rational acts—are time-relative, because the desires rational to have are time-relative. So the timeless reasons theory collapses into the desire theory, except that one is first required to have certain desires by the 'timeless' reasons. So even if there are desire-independent reasons, they pass their role of being one's reasons to the current desires they require; and *these* pass it to the desires they recommend later.

Summary: Nagel says an act is rational iff induced by a motivated desire, a desire one is justified in having by its target being one timeless standards of state goodness, timeless reasons, say should obtain. But say one's current desires are motivated: I argued that, if one's reasons are to be monotonic, it must be that, given what it is to have desires and rationally to prosecute them, one's current ones inform the rationality of one's later ones; and if the act needed to cause a current desire's target is revising the desire, the revisant is rationally obligatory to have, and so is motivated, even if it is for a state the first desire opposes.[38] So if timeless reasons ask that one form certain desires, the reasons pass to

[37] See note 44, below.

[38] I say if it advances rational desires to replace them with ones for opposite states, the new ones are rational to have, motivated. But most philosophers would think these aren't rational—forming them is rationally motivated irrationality. And one might object that my view leaves no room for such irrationality. But it does. Say I'll *certainly* frustrate your current rational desires (ones for good states), unless you take a pill that *may* make you act against them. Taking it maximises on your desires, for it saves a chance that they will be satisfied. Say you take it and unluckily it makes you act against your desires: since, on my view, rational acting should advance one's desires, this is irrational acting; but it is rationally motivated, for it was rational for you (since maximising on your desires) to take the pill with this as a possible result. And say the pill may make you act to form desires whose formation would retard those you will have just after taking it. If it does, then since, on my view, a desire is rational iff forming it advances the desires one had when forming it, this is irrational desire formation; but it is rationally motivated because it was rational for you (since maximising on your desires) to take the pill with this as a possible result. So rationally motivated irrationality is possible, but in a non-traditional form: rationally motivated irrational action is not action later against desires you have now, but against ones you rationally will have then. And rationally motivated irrational desire formation is not forming a desire for a state incompatible with states currently desired; it is forming a desire whose formation retards desires one holds when forming it.

them the rating of later acts and desires for rationality; and these desires may rationally oblige their own revision. Thus my reasons may change; I'm justified in having them only time-relatively.

So say one has a desire recommended by its target's goodness. And say one faces an *SES* where, to advance the desire, one must replace it with one for an opposite state. Then *it* is the desire now rational to have; and *SESs* show that the rationally obliged reasons behind rationally obliged acts can decouple from state goodness (unless it mutates, and is relative). A rational agent wouldn't always be moved by the goodness of the states her acts might cause, any more than an epistemically rational agent would always be moved to beliefs by their truth. What makes a desire rational isn't necessarily its targeting a good state; so what makes an act rational isn't necessarily its advancing such a desire. Now, I said the timeless reasons theory collapses into the desire theory except for first desires. But maybe not even *these* should target good states if good states would be caused by one's desiring bad ones instead. Maybe state goodness determines initial desire rationality only indirectly: the desires rational to have are those *the having of which would advance* good states, not necessarily desires *for* good states. This has been our theme—the conflict between desiring a state because one sees its goodness (or being rationally obliged to desire it because it is good), and desiring a state (or being obliged to desire it) because desiring it will cause goodness.[39]

I argued that no monotonic theory of which acts are rational can be timeless. For it will imply which attitudes or dispositions are rational; and to be monotonic, it must call this time-relative; ditto for what are one's proper reasons. Nor can a monotonic theory of which attitudes are rational be timeless. For attitudes rationally oblige acts advancing them (if they are conative attitudes) or the states they favourably represent (if the attitudes are cognitive); and for any attitudes, there are *SESs* where the acts the attitudes oblige consist in revising those attitudes. But since no true theory of reasons can be non-monotonic—for it wouldn't always give unambiguous, coherent advice on which acts are rational—reasons can't be necessarily timeless.

Nagel's view exemplifies a paradox at the heart of our conception of rationality, a paradox analogous to one in our conception of morality: as Kavka taught us (though he didn't say it decision-theoretically), our moral scheme violates monotonicity: we think an act is morally required if it would yield a good outcome; but also that an intention to act is morally permitted only if it is an intention to make a good outcome. But then where the act likely to yield a good outcome is forming an intention to make a bad one (an intention it is unlikely one shall have to act on), our morality is conflicted: we should do the act of forming the intention, for this will likely have good effects; yet we should not, since the act it is an intention to do would have bad effects if done. And it now seems that this conflict between consequentialist and deontological standards for the *morality* of an act is reproduced for its *rationality*—consequentialist assessments of forming intentions whose formation would have good effects recommend forming the intentions; but deontological assessments of forming such intentions recommend against forming them, because they are intentions to do bad (or irrational) actions, actions likely to have bad (or dispreferred) effects.

[39] It would be interesting to connect this with Peter Railton, 'Alienation, Consequentialism, and the Demands of Morality', *Philosophy and Public Affairs* 13 (1984), pp. 134–71.

I see only one resolution: whatever the standard for the rationality (or morality) of acts, it is right to revise it in *SESs*; the revisant is then the correct standard. Thus I offer a unified theory: all desire-justification is pragmatic: *desires* are innocent; it is only *having* or *forming* them in the circumstances that is rational or irrational. (Derivatively, a person is rational or irrational for forming them then.) Could some desires be rational to have in all possible circumstances, others, in none? I suspect no desire is such that there couldn't be a situation in which it would be best advanced by being revised, nor such that there couldn't be a situation making it rational to adopt that desire given some other desire; so no desire is necessarily always rational (to have), nor always irrational. In any case, what is justified or not, then, is forming a desire, not a desire itself; and what justifies forming one is always that this advances the desires one has when assessing desires: depending on the situation, a desire can be advanced by keeping it, or by adding a desire for a state identical to, part of, or means to, the state the original desire targets; but sometimes, instead, by replacing the desire with one *whose formation* makes more likely the state the former desire targets. This theory of reasons is monotonic because its only standard for desire rationality is the desires held while assessing prospective desires; and the standard is updated whenever the desires comprising it oblige their own revision. So there is no standard, constant in content, to conflict with the rationally self-evolving standard, and so no violation of monotonicity at any time of evaluation.

I claimed Nagel's theory of reasons is non-monotonic: it can't guarantee unambiguous verdicts on which acts are rational. I then generalised: there can't be a necessarily timeless standard of act rationality, nor reasons that, necessarily, all rational agents must always have. So no true theory can posit as reasons, things necessarily timeless; for then reasons would be non-monotonic, and so couldn't really be our reasons.

What does this mean for the prudence problem? Nagel found it absurd that our desires should be our reasons, for then we might have reason in our current desires to act in ways frustrative of our later ones, and *vice versa*. Instead, our desires can be related to reasons only as things that might inform the timeless goodness of states. E.g., some future desires might be such that, because we will have them then, the states they target are good; and we have timeless reason to cause the good; so we might now have reason to cause states we won't desire until later, reason now to act against our current desires, advancing future ones. And since to rationally act is to desire so to act, we may have reason now to desire to act in ways retarding our current, pre-critical desires, ways that will advance our future desires.

But if I'm right, the rationality of later desires is time-indexed; and I can't be required to desire now to accommodate my future desires, except as rationally mandated by my current ones and ones rationally evolved from them *via* the desire theory's strictures. Not even if my current desires are ones timeless reasons ask me to have now. For then those desires are, in effect, my current reasons; and having them, rationally I can't act except as they say, nor alter them till such time as they favour this. So I can't be bound by future desires until forming them is favoured by current ones. And thereafter I can't rationally be bound by the past ones. Since this comes of what was first dictated by reasons supposedly timeless, even on the timeless reasons theory my reasons (motivated desires) must be time-relative.

But maybe the timeless reasons theory can have rational current acts be both determined by desires timelessly dictated, and informed by future desires: the reasons

would take account of my current and future pre-critical desires as things which inform which states are good, and then would recommend only post-critical desires giving due weight to both as the desires which should always guide my choice (as motivated—justified—desires). So there are post-critical desires I should always have and follow, thereby accommodating my future pre-critical ones.[40]

Yet even for post-critical desires, there is a possible *SES* where, to advance them, one must revise them into desires for states opposite those first desired, whence the desires rationally oblige their own revision, as in *SDSs*. The timeless reasons theory says it is rational to have and advance the post-critical desires timelessly dictated; so should a person with such desires face a *SES* for them, since they then oblige their own revision, the 'timeless' reasons are, again, dictating some desires only relatively to the time of the required revision, forbidding others. Again, one's reasons (motivated desires) are time-relative, and must be rationally successive as per the desire theory.

True, if one never faces *SESs* one should always have the same desires.[41] But this would be mere luck; that one *could* face *SESs* for one's desires makes the theory that all agents' reasons *must always be timeless* self-refuting. For if such reasons dictate certain desires to everyone always, they do so to me at *t*. But to have a desire is to have a thing which, if motivated, not only fixes rational acts, but constrains which later desires are motivated—rational to have. Thus even if, at *t*, I have a desire dictated by 'timeless' reasons, post-*t*, they can't monotonically inform my rational duty to desire and act, except as they are just the time-relative reasons I would be mandated in having by the desire theory (the desires I would be motivated in having time-relatively—as in *SESs*—by desires first motivated by 'timeless' reasons). So apart from initially dictated desires, the timeless reasons theory collapses into the desire theory.[42]

Or does it? Maybe one's current desires, even if dictated by timeless reasons, partly constrain the rationality of one's later desires. But surely other desires could later be added without violating these strictures. So timeless reasons could then dictate with which other desires I should supplement my first ones.

But if the reasons may not later dictate desires whose adoption would be irrational given the first-dictated ones or ones rationally evolved from them, at best they can later dictate desires only for states on which the reasons were silent in dictating first desires. For if they dictate later desires for states on which I already have desires, they are demanding a change in my desires about those states, one inappropriate given the first desires. But if the reasons were silent on some states, they didn't dictate preferences decision-theoretically complete:[43] some states weren't ranked. But then the preferences couldn't fix the rationality of acts; so the reasons weren't then determinative of what I should do. So if ever the reasons dictate desires fixing the rationality of acts, they must ever after be silent. If prior dictates gave me complete desires, there can be none the timeless reasons may require me to add.

[40] Thanks to a referee for this suggestion.

[41] Or at least any new ones should be for states part of or means to the ends one should always desire.

[42] This also argues for the relativity of reasons—motivated desires—to persons and situations. Say timeless reasons dictate certain desires to all persons; but then say one person faces an *SES* for the desires, another, not. Then the first person will be justified in forming different desires; so the two persons should have different motivated desires, and so different reasons.

[43] I defined complete preferences in Part I, Section 3, above.

An exception: recall from *d)* in Part I, Section 5, above, that if my desires become such that replacing them won't affect whether their targets obtain, because these became such that no act could make them more likely, it isn't anti-maximising on the desires to replace them. So timeless reasons could recommend new desires; but only once it becomes true that it does not retard the old to drop them. So the recommendability of the new is conditional on a time-relative fact. So they are right (are motivated so as to be my reasons) only time-relatively.

How general are these results? The structure Decision Theory says desires need to determine rational acts is needed of anything with this power. Say timeless reasons are truths ranking states by goodness; to be rational is to cause the best state one can. But the ranking must be complete, transitive, etc., else there will be states the rationality of causing which the truths won't fix, and on whose rank they conflict. And the truths must be internalised to make one's behaviours be acts. Maybe one needs beliefs about the ranking and an inclination to cause states in rank order; or maybe just apprehending the truths motivates. But my having an inclination now may forbid my having other inclinations, and may oblige me to revise it later; and my having an act-inducing apprehension now may require me to have another one later. And monotonicity requires the new standards of rationality I had to adopt to count as correct. So the standard for rational attitudes and acts is relative. No matter their psychological or metaphysical kind, then, if ever timeless reasons are to cause acts and rate their rationality, they can't thereafter do so. Reasons can't be necessarily timeless.[44,45]

Dalhousie University

Received: June 2000
Revised: November 2000

[44] For more on this result's generality, see my 'The Prudence Problem'. I said no monotonic theory of rational acts or attitudes can hold that necessarily one rationally must always have the same attitudes. Instead, one must have ones whose formation best serves those one was previously obliged to have; and for any attitudes once rationally obligatory due to their targets' goodness, there can be *SESs* where one rationally should *drop* the attitudes, even though they target states good pre-*SES*—attitude rationality is relative to time, agent, situation. So no state can be necessarily timelessly rationally obligatory to value. For argument that, on plausible meta-ethical assumptions, this means the goodness of *states* is relative, see my 'Moral Paradox'.

[45] My thanks to audiences at the 1999 Alabama Philosophical Society meeting (for which this was the plenary address), and at Dalhousie University, the University of British Columbia and the 1991 Canadian Philosophical Association meeting. For discussion, I'm grateful to Nathan Brett, Bob Bright, Richmond Campbell, Sue Campbell, trish Leadbeater, Carolyn McLeod, Bob Martin, tori McGeer, Peter Schotch, Rob Shaver, Susan Sherwin, Heidi Tiedke, Tom Vinci, Mark Vorobej and Russ Weninger. For written comments, my thanks to David Schmidtz, the editors of *AJP*, and referees of various versions. My work was aided by a Grant from the Social Sciences and Humanities Research Council of Canada.