

Global Rationality

Richard Chappell

Introduction

We typically conceive of rationality in atomistic terms, as a matter of doing what seems best from the local perspective of the moment, or maximizing expected utility. I want to explore an alternative – more holistic – way of thinking about rationality, which applies primarily at the global level of the whole, temporally extended person, in contrast to the local level of each momentary person-stage. This distinction may be motivated by noting that global optimality can – for various reasons – require us to do other than what seems optimal within the confines of a moment. Holistic rationality, as I envisage it, tells us to adopt a broader view, transcending the boundaries of the present and identifying with a timeless perspective instead. Of course, even atomists will want to go beyond the present by including knowledge about the future in their decision procedure. But I wish to make the stronger claim that the procedure *itself* should be, in an important sense, time-invariant. We might think of this as a kind of universalization constraint: to assess a decision procedure or form of practical reasoning, you should consider the consequences, not just of employing it at *this* momentary stage, but at all your relevantly similar stages.

Julian Nida-Rumelin has suggested: “It is perfectly rational to refrain from point-wise optimization because you do not wish to live the life which would result.”¹ Yet refraining from optimization might seem to violate the principle that one rationally ought to do what seems best. To reap the rewards of global rationality, we must be willing to treat the dictates of the broader perspective as rationally authoritative, no matter how disadvantageous it may seem from the particular perspective of our local moment.² This amounts to an intrapersonal analogue of the

‘social contract’: each of our momentary stages abdicates some degree of rational autonomy, in order to enhance the rationality and autonomy of our person as a whole.

The challenge here is to make sense of the normative status of the advice issuing from these conflicting perspectives. What, for instance, should consequentialists say about those particular instances in which following the globally optimal strategy would actually have worse consequences? I will draw on – and further develop – the theoretical apparatus of contemporary philosophy of normativity in order to diagnose, clarify, and assess the various ways such conflicts may arise, and to rebuff charges of inconsistency or paradox. Overall, this essay seeks to establish two things. First, the distinction between local and global levels of rationality can be fruitfully applied, as an extension of the standard framework, to help us understand the puzzles discussed by Parfit and others. Secondly, we would do well to favour the global level in cases where the two conflict.

Reasons and Rationality

First, let us distinguish the objective and subjective modes of normativity, as reflected in contemporary philosophical use of the terms ‘reasons’ and ‘rationality’.³ *Reasons* are provided by facts that count in favour of an action. For example, if a large rock is about to hit the back of your head, then this is a reason for you to duck, even if you are unaware of it. As this example suggests, the objective notion I have in mind is largely independent of our beliefs: *p* can be a reason for you, in this sense, even if you do not believe that *p*.⁴ As inquiring agents, we try to discover what reasons for action we have in virtue of our actual situation, and hence what we should do. Such inquiry would be redundant according to subjective accounts, which restrict reasons to things that an agent already believes. So I use the term exclusively in its objective sense. Consequentialism implies that we have reason to bring about good states of affairs, and to prevent bad ones from obtaining.⁵ I take it as analytic that we have *most reason* to do what

would be *best*. I will follow Parfit's terminology in saying that this is what one *ought*, in the reason-implying sense, to do.⁶

There is another sense of 'ought', tied to the subjective or evidence-based notion of *rationality* rather than the objective or fact-based notion of 'reasons'. As Kolodny puts it, rationality is "about the relation between your attitudes, viewed in abstraction from the reasons for them."⁷ Sometimes the evidence can be misleading, so that what seems best is not really so. In such cases, we may say that one *rationally ought* to do what seems best, given the available evidence. But due to their ignorance of the facts, they would not be doing what they actually have most reason to do. Though they couldn't know it, some alternative action would have been better.

This raises the question of what to say when reasons and rationality diverge. Suppose that someone ought, in the reason-implying sense, to X, but that they rationally ought to Y. If innocent people have been convincingly framed, for instance, then the (unknown) facts provide reasons for acquittal, though it may be rational for a jury to condemn them based on the available evidence. Which takes precedence? What is it that the jury *really* ought to do? There is some risk of turning this into a merely terminological dispute. But we can make some substantive observations here. In particular, I think that the reason-involving sense of 'ought' is arguably the more fundamental normative concept. This is because it indicates the deeper goal, or what agents are ultimately seeking.

The purpose of deliberation is to identify the best choice, or reach the correct conclusion. In practice, we do this by settling on what seems to us to be best. But we do not think that the appearances have any independent force, over and above the objective facts. We seek to perform the best action, not merely the best-seeming one. Of course, from our first-personal perspective we cannot tell the two apart. That which seems best to us is what we take to truly *be* best. Belief is, in this sense, "transparent to truth". Because our beliefs all seem true to us, the rational choice

will always *seem* to us to also be the best one.⁸ We can thus take ourselves to be complying with the demands of both rationality and fact-based reasons. Nevertheless, it is the latter that we really care about. One way to bring this out is to consider the advice that would be given by a helpful third party: recognizing your false beliefs, they would surely advise you to do what was truly best, rather than what merely seemed to you to be so.⁹

This is especially clear in epistemology. We seek true beliefs, not justified ones. Sure, we would usually take ourselves to be going wrong if our beliefs conflicted with the available evidence. Such conflict would indicate that our beliefs were likely false. But note that it is the *falsity*, and not the mere *indication* thereof, that we are ultimately concerned with. More generally, for any given goal, we will be interested in evidence that suggests to us how to attain the goal. We will tend to be guided by such evidence. But this does not make following the evidence *itself* our ultimate goal. Ends and evidence are intimately connected, but non-identical. Normative significance accrues in the first instance to our ends, whereas evidence is merely a means: we follow it for the sake of the end, which we know not how else to achieve. Applied to the particular case of reasons and rationality, then, it becomes clear that reasons are closer to the real goal, whereas rationality is merely the guiding process by which we hope to achieve it. Since arriving at the intended destination is ultimately more important than faithfully following the guide, we may conclude that the reason-implying sense of ‘ought’ takes normative precedence.¹⁰

I will use this as my default sense of ‘ought’ in what follows.

Indirect Utilitarianism and Blameless Wrongdoing

Act Utilitarianism claims that our actions ought to maximize the good. Paradoxically, if people tried to act like utilitarians, this would plausibly have very bad consequences. For example, authorities would engage in torture or frame innocent persons whenever they believed that doing so would cause more good than harm. Such beliefs might often be mistaken, however, and with

disastrous consequences. Let us suppose that attempts to directly maximize utility will generally backfire. Utilitarianism then seems to imply that it is wrong to be a utilitarian. But the conclusion that utilitarianism is self-defeating only follows if we fail to distinguish between *criteria of rightness* and *decision procedures*.¹¹—

We typically conceive of ethics as a practically oriented field: a good moral theory should be action-guiding, or tell us how to act. So when utilitarianism claims that the right action is that which maximizes utility, it is natural for us to read this as saying that we should try to maximize utility. But utilitarianism as defined above does not claim that we ought to *try* to maximize utility. Rather, it claims that we should *achieve* this end. If one were to try and fail, then their action would be wrong, according to the act-utilitarian criterion. This seems to be in tension with the general principle that we rationally ought to aim at the good. The utilitarian criterion instead tells us to have whatever aims would be most successful at *attaining* the good. This difference will be clarified in the section on ‘object- and state-based modes of assessment’, later in this essay. For now, simply note that the best utilitarian consequences might result, say, from a steadfast commitment to respect human rights no matter how expedient their violation might appear. In this case, the utilitarian criterion tells us that we should inculcate such anti-utilitarian practical commitments.

This indicates a distinction between two levels of normative moral thought: the intuitive and the critical.¹²— Our intuitive-level morality consists of those principles and commitments that guide us in our everyday moral thinking and engage our moral emotions. This provides our moral *decision procedure*. It is often enough to note that an action would violate our commitment to honesty, for instance, to settle the question of whether we should perform it. This is not the place for cold calculation of expected utilities. They instead belong on the critical level, when it comes time to determine which of our intuitive principles and commitments are well-justified ones. For the indirect utilitarian, honesty is good because being honest will do a better job of improving the

world than would being a scheming, opportunistic, direct utilitarian. The general picture on offer is this: we use utility as a higher-order criterion for picking out the best practical morality, and then we live according to the latter. Maximizing utility is the ultimate goal, but we do well to adopt a more reliable indirect strategy – and even other first-order “goals” – in order to achieve it.¹³—

What shall we say of those situations where the goal and the strategy conflict? Consider a rare case where, say, framing an innocent person *really would* have the best consequences. Such an act would then be *right* – or what we have most reason to do – according to the utilitarian criterion. Yet our practical morality advises most strongly against it, and *ex hypothesi* we ought to live according to those principles. Does this imply a contradiction: the right action ought not to be done? Only if we assume a further principle of normative transmission:

(T) If you ought to accept a strategy S, and S tells you to X, then you ought to X.¹⁴—

This is plausible for the rational sense of ‘ought’, but not the reason-involving sense that I am using here. We might have most reason to adopt a strategy – because it will more likely see us right than any available alternative – without thereby implying that the strategy is perfect, i.e. that everything it prescribes *really is* the objectively best option. S might on occasion be misleading, and then we could have more reason to *not* do X, though we remain unaware of this fact. So we should reject (T), and accept the previously described scenario as consistent. To follow practical morality in such a case, and refrain from expedient injustice, would constitute what Parfit calls “blameless wrongdoing”.¹⁵— The agent fails to do what they have most moral reason to do, so the act is wrong (objectively suboptimal). But the agent herself has the best possible motives and dispositions, and could say, “Since this is so, when I do act wrongly in this way, I need not regard *myself* as morally bad.”¹⁶—

Parfit's solution may be clarified by appealing to my earlier distinction between the local and global levels. Our 'local' assessment looks at the particular act, and condemns it for sub-optimality. The 'global' perspective considers the agent as a whole, with particular concern for the long-term outcomes obtained by consistent application of any given decision-procedure. From this perspective, the agent is (*ex hypothesi*) entirely praiseworthy: their psychological makeup will in fact lead to better consequences *overall* than any available alternative.¹⁷ The apparently conflicting judgments are consistent because they are made in relation to different standards or modes of assessment. I have illustrated this with the example of indirect utilitarianism, but the general principle will apply whenever some end is best achieved by indirect means. More generally than "blameless wrongdoing", we will have various forms of (globally) optimal (local) sub-optimality.

Meta-coherence and Essential By-products

The above discussion focused on the objective, reason-involving sense of 'ought'. Let us now consider the problem in terms of evidence and what one *rationally ought* to do. Rationality demands that we aim at the good, or *do what seems best*, i.e. maximize expected utility. But the whole idea of the indirect strategy is to be guided by reliable rules rather than direct utility calculations. One effectively commits to occasionally acting "irrationally" (in the local sense), though it is rational – subjectively optimal – to make this commitment. Parfit thus calls it "rational irrationality".¹⁸ But we may question whether expected utility could really diverge from the reliable rules after all.

Sometimes we may be in a position to realize that our initial judgments should be revised. I may initially be taken in by a visual illusion, and falsely believe that the two lines I see are of different lengths. Learning how the illusion worked would undercut the evidence of my senses. I would come to see that the *prima facie* evidence was misleading, and the belief I formed on its

basis likely false. Principles of meta-coherence suggest that it would be irrational to continue accepting the appearances after learning them to be deceptive, or more generally to hold a belief concurrently with the meta-belief that the former is unjustified or otherwise likely false.¹⁹ This principle has important application to our current discussion, potentially unifying the two levels of local and global rationality.

We adopt the indirect strategy precisely because we recognize that our direct first-order calculations are unreliable. The utilitarian sheriff might think that framing an innocent subject would have high expected utility. But if he recalls his own unreliability on such matters, he should lower the expected utility accordingly. As a good indirect utilitarian, he believes that in situations subjectively indiscernible from his own, to follow a strict policy of never framing innocent persons will generally yield the best results. Taking this higher-order information into account, he should revise his earlier judgment and instead reach the all-things-considered conclusion that seeing justice done maximizes expected utility even for this particular act.²⁰ This seems to collapse the distinction between local and global rationality. When all things are considered, the former will come to conform to the latter.²¹

This will not always be the case, however. A crucial feature of the present example is that one can consciously recognize the ultimate goal at the back of their mind, even as they employ an indirect strategy in its pursuit. But what if the pursuit of some good requires that we make ourselves more thoroughly insensitive to it? Jon Elster calls such goods “essential byproducts”, and examples might include spontaneity, sleep, acting unselfconsciously, and other such mental absences.²² Such goods are not always susceptible to momentary rational pursuit, even with indirect strategies. The problem is no longer mere ignorance of how best to achieve the good. Rather, to achieve these goods we must relinquish any conscious intention of doing so. As we relax and drift off to sleep, we cannot concurrently conceive of our mental inactivity *as* a means to this end. One cannot achieve a mental absence by having it “in mind” in the way required for

the means-ends reasoning I take to be constitutive of rationality. In the event of succeeding, one could no longer be locally rational in their pursuit of the essential byproduct, for they would not at that moment be intentionally pursuing it at all.

Nevertheless, there remains an important sense in which a person remains perfectly rational in having their momentary selves abdicate deliberate pursuit of these ends. If we attribute the goal of nightly sleep to the *whole temporally extended person*, then this abdication is precisely what sensible pursuit of the goal entails. In this sense, we can understand the whole person as acting deliberately even when their momentary self does not. So the distinction is upheld: global rationality recommends that we simply give up on trying to remain locally rational when we want to get some rest.

Object- and State-based modes of assessment

Odd as it may seem, even local rationality recommends surrendering itself in such circumstances. From the local perspective of the moment, pursuit of the goal is best advanced by ensuring that one's future self refrains from such deliberate pursuit. The puzzle arises because mental states are subject to two very different modes of assessment: one focusing on the *object* of the mental state, and the other focusing on the *state* itself.²³ Suppose an eccentric billionaire offers you a million dollars to believe that the world is flat. The object of belief, i.e. the proposition that the world is flat, does not merit belief. But this state of belief would, in such a case, be a worthwhile one to have. In this sense we might think there are reasons for (having the state of) *believing*, which are not reasons for (the truth of) the *thing believed*.²⁴ It seems plausible that desire aims at value in much the same way that belief aims at truth. Hence, indications of value could provide object-based reasons for intention or desire – much as indications of truth provide object-based reasons for belief – whereas the utility of having the desire in question

could provide state-based reasons for it.²⁵ This is the difference between an *object's* being worthy of desire, and a desire for the object being a *state* worth having.

There are various theories about what reasons we have for acting, and hence what objects merit our pursuit. For example, we may call “*Instrumentalism*” the claim that we have reason to fulfill our own present desires, whatever they may be. *Egoism* claims that we have most reason to advance our own self-interest. And *Impartialism* says that we have equal reason to advance each person’s interests. I propose that for any such account of reasons, we can pair it with a corresponding account of object-based local rationality, based on the following general schema:

(G) Rationality is a matter of pursuing the good, i.e. being moved by the appearance of those facts ____ that provide us with reasons for action.

Let us say that S has a *quasi-reason* to X, on the basis of some non-normative proposition p, when the following two conditions are satisfied: (i) S believes that p; and (ii) if p were true then this would provide a reason for S to X. Note that on my above account, S need not *recognize* that the truth of p would provide reason to X. We may then understand (G) as the claim that one rationally ought to do what one has most *quasi-reason* to do. This clarifies the proposed relation between reasons and rationality, as distinct from an alternative view that holds simply that one is rationally required to do whatever they *believe they have most reason to do*.

The different theories mentioned earlier posit different reasons, so different quasi-reasons, and hence different specifications of rationality in this sense. For example, according to Egoism, agents are rational insofar as they seek to advance their own interests. There is a sense in which the theory thereby claims this to be the supremely rational aim.²⁶ But let us suppose that having self interest as one’s dominant local aim would foreseeably cause one’s life to go worse, as per “the paradox of hedonism”.²⁷ Egoism then implies that we would be *irrational* to knowingly

harm ourselves by having this aim. This conclusion seems to contradict the original claim that this aim is “supremely rational”. On this basis, Dancy claims that such theories are flatly inconsistent: they “say... that there is a single rational aim which it is not rational... to aim at.”²⁸—

The distinction between object- and state-based assessments may help resolve this problem. We might say that an aim *embodies* rationality in virtue of its object, in that it constitutes supreme sensitivity to one’s quasi-reasons. Or the aim might be *recommended* by rationality, in the sense that one’s quasi-reasons tell one to have this aim, in virtue of the mental state itself. As before, the apparent incoherence can be traced to the conflation of two distinct modes of assessment. The aforementioned theories should be interpreted as claiming that their associated aims supremely embody rationality, even though it might not be rationally recommended to embody it so. This reflects the coherent possibility that something might be desirable – worthy of desire – even if the desire itself would, for extrinsic reasons, be a bad state to have.

It is worth noting that this distinction appears to hold independently of the local/global distinction. We might imagine a good – perhaps sainthood²⁹ – that would be denied to anyone who ever entertained it as a goal. If one sought it via the standard “globally rational” method of preventing one’s future momentary selves from deliberate “locally rational” pursuit, it would already be too late: this initial plotting would suffice to disqualify one. There is no rational way at all, on any level, to pursue the good. Still, being of value, the good might *merit* pursuit. It might even provide reasons of sorts, even though one could never recognize them as such.³⁰— So although the object/state distinction may recommend a shift from local to global rationality, it further establishes that even the latter may, in special circumstances, be transparently disadvantageous. Reasons and rationality may come apart, even when no ignorance is involved, because it may be best to achieve a good without *ever* recognizing it as such. This would provide reasons that elude our rational grasp, being such that we ought to act unwittingly rather than by grasping the reason that underlies this very ‘ought’-fact.

Rational Holism

We have seen how various distinctions, including that between the local and global levels of rationality, can help us to make sense of the indirect pursuit of goods. If we know our first-order judgments to be unreliable, then meta-coherence will lead us to be skeptical of those judgments. Indirect utilitarianism stems from recognizing that expected utility is better served by instead following a more reliable – globally optimal – strategy, even if this at times conflicts with our first-order judgments of expedience. Global rationality paves the way for utilitarian respect for rights, and meta-coherence carries it over to the local level. Essential by-products re-establish the distinction, as we may understand a goal as being rationally pursued at the level of the temporally extended person, if not at the level of every momentary stage or temporal part. Although the object/state distinction implies that even global rationality may be imperfect, the preceding cases suggest that we would do well at least to prize the global perspective over the local one. I now want to support this conclusion by considering a further class of problems that could be fruitfully analyzed as pitting the unified agent against their momentary selves.

Consider Newcomb's Problem:³¹ a highly reliable predictor presents you with two boxes, one containing \$1000, and the other with contents unknown. You are offered the choice of either taking both boxes, or else just the unknown one. You are told that the predictor will have put \$1,000,000 in the opaque box if she earlier predicted you would pick only that; otherwise she will have left it empty. Either way, the contents are now fixed. Should you take one box or both? From the local, momentary perspective, the answer seems clear: the contents are fixed, it's too late to change them now, so you might as well take both. Granted, one would do better to be the sort of person who would pick only one box. That is the rationally recommended dispositional state. But taking both is the choice that embodies rationality, according to the local view. This atomistic reasoning predictably leads to a mere \$1000 prize.

Suppose that one instead adopted a more holistic perspective, giving weight to the kinds of reasoning that, judging from the timeless perspective, one would want one's momentary stages to employ. This globally rational agent is willing to commit to being a one-boxer, and so will make that choice even when it seems locally suboptimal. This predictably leads to the \$1,000,000 prize, which was unattainable to the atomist.

Similar remarks apply to Kavka's toxin puzzle.³² Suppose that you would be immediately rewarded upon forming the intention to later drink a mild toxin that would cause you some discomfort. Since you will already have received your reward by then, there would seem no narrowly local reason for you to carry out such an intention. Recognizing this, an atomist about rationality cannot even form the intention to begin with. (You cannot intend to do something that you know you will not do.) Again we find that atomistic reasoning disqualifies one from attaining something of value. The rational holist, by contrast, is willing to follow through on earlier commitments even in the absence of temporally local reasons.³³ She wishes to be the kind of person who can reap such rewards, so global rationality leads her to behave accordingly.

In both these cases, the benefits of global rationality require that one be disposed to follow through on past commitments. One must *tend* to recognize one's past reasons as also providing reasons for one's present self. This allows one to overcome problems, such as the above, which are based on a localized object/state distinction.³⁴ But occasional violation of this disposition might allow one to receive the benefits without the associated cost. (One might receive the reward for forming the sincere intention to drink the toxin, only to later surprise oneself by refusing to drink it after all.) So let us now consider an even stronger sort of case that goes beyond the object/state distinction and hence demands more than the mere disposition of global rationality. Instead, the benefits will accrue only to those who follow through on their earlier resolutions.³⁵

Pollock's *Ever Better Wine* improves with age, without limit.³⁶ Suppose you possess a bottle, and are immortal. When should you drink the wine? Atomists could never drink it, for at any given time they would do better to postpone it another day. But to never drink it at all is the worst possible result! Or consider Quinn's *Self-Torturer*, who receives \$10,000 each time he increases his pain level by an indiscernible increment.³⁷ It sounds like a deal worth taking. But suppose that the combined effect of a thousand increments would leave him in such agony that no amount of money could compensate. Because each individual increment is – from the narrowly local perspective of the moment – worth taking, rational atomism will again lead one to the worst possible result. A good result is only possible for agents who are willing to let their global perspective override local calculations. The agent must make in advance a *rational resolution* to stop at some stage n , even though from the local perspective of stage n he would do better to continue on to stage $n+1$.

It seems clear that the global perspective is rationally superior. The agent can foresee the outcomes of her possible choices. If she endorses the local mode of reasoning then she will never have grounds to stop, and so will end up stuck with the worst possible outcome. It cannot be rational to accept this when other options are open to her. If she is instead *resolute* and holds firm to the choice – made from a global or timeless perspective – to stop at stage n , then she will do much better. Yet one might object that this merely pushes the problem back a step: how could one rationally resolve to choose n rather than $n+1$ in the first place?

The problem of Buridan's ass, caught between two equally tempting piles of hay, shows that rational agents must be capable of making arbitrary decisions.³⁸ It cannot be rational for the indecisive ass to starve to death in its search for the perfect decision. Indeed, once cognitive costs are taken into account, it becomes clear that all-things-considered expected utility is better served by first-order satisficing than attempted optimizing.³⁹ (“The perfect is the enemy of the good,” as

the saying goes.) Applying this to the above cases, we should settle on some n , any n , that is good *enough*. Once we have made such a resolution, we can reject the challenge, “why not $n+1$?” by noting that if we were to grant that increment then we would have no basis to reject the next ones, *ad infinitum*, and that would lead us to the worst outcome.

Though our reasoning activity takes place at particular moments, the perspective we adopt at that time need not be so limited. We can, in the present, reason *as if* from a global perspective – and the above cases suggest that this is precisely what we rationally ought to do. Hence I propose the following negative principle of rational holism:

(H-) *In any given situation: one rationally ought not to employ any decision procedure that could not be endorsed from a global perspective (i.e. abstracting away from one’s present temporal location or momentary stage).*

The above discussion shows that point-wise optimization is not temporally universalizable in some cases. If one endorses the arbitrary shift from n to $n+1$, parity of reasoning would lead one’s future selves to never stop. The consequences of universalizing that method of reasoning would be clearly undesirable, so (H-) requires us to reject it. The atomist’s “local rationality” is in fact not rational at all. Instead, the way we should reason at a moment derives from the methods we would endorse from a timeless perspective. Of course, in unproblematic cases, the locally optimal option may happen to coincide with what is truly (globally) rational. Our idealized timeless selves might endorse our present use of the decision procedure. But as the diverging cases show, it is the global perspective that takes precedence. Hence the positive holistic principle:

(H+) *In any given situation: one rationally ought to act and reason as one would recommend from a timeless perspective.*

It is not enough simply to determine that $n+1$ is better than n , and hence take that step before determining what to do next. In cases where local and global optimality diverge, one must instead look at the “big picture”, resolve where one wants to end up, and act accordingly. Of course, in reality our reasoning always takes place within a specific temporal context. But we may nevertheless adopt a timeless *perspective* by abstracting away such details, and thus refusing to employ methods of reasoning that we recognize as non-universalizable. Thus localized decisions may be governed by global norms.

Conclusion

The standard picture of rationality is thoroughly atomistic. It views agents as momentary entities, purely forward-looking from their localized temporal perspective. In this essay, I have presented an alternative, more holistic view. I propose that we ascribe agency primarily to the whole temporally extended person, rather than to each stage in isolation. This view allows us to make sense of the rational pursuit of essential by-products, as we may ascribe deliberate purpose to a whole person even if it is absent from the minds of some individual stages. Moreover, global rationality sheds light on the insights of indirect utilitarianism, though meta-coherence allows that these conclusions may also be accessible from a temporally localized perspective. I have shown how to develop Parfit’s framework to better accommodate these puzzling cases, in response to charges of outright inconsistency. Finally, I have argued that there are cases where reasoning in the atomistic manner of point-wise optimization leads to disaster. Such disaster can be avoided if the agents embrace my holistic conception of rational agency, acting only in ways that they could endorse from a global or timeless perspective. Persons are more than the sum of their isolated temporal parts; if we start acting like it then we may do better than traditional decision theorists would think rationally possible.

References

- Dancy, J. (1997) 'Parfit and Indirectly Self-defeating Theories' in J. Dancy (ed.) *Reading Parfit*. Oxford : Blackwell.
- Elster, J. (1983) *Sour Grapes*. Cambridge : Cambridge University Press.
- Hare, R.M. (1981) *Moral Thinking*. Oxford : Clarendon Press.
- Harsanyi, J. (1980) 'Rule Utilitarianism, Rights, Obligations and the Theory of Rational Behavior' *Theory and Decision* **12**.
- Kavka, G. (1983) 'The toxin puzzle' *Analysis*, **43**:1.
- Kolodny, N. (2005) 'Why Be Rational?' *Mind*, **114**:455.
- McClennen, E. (2000) 'The Rationality of Rules' in J. Nida-Rumelin and W. Spohn (eds.) *Rationality, Rules, and Structure*. Boston : Kluwer.
- Musgrave, A. (2004) 'How Popper [Might Have] Solved the Problem of Induction' *Philosophy*, **79**.
- Nida-Rumelin, J. (2000) 'Rationality: Coherence and Structure' in J. Nida-Rumelin and W. Spohn (eds.) *Rationality, Rules, and Structure*. Boston : Kluwer.
- Nozick, R. (1993) *The Nature of Rationality*. Princeton, N.J. : Princeton University Press.
- Parfit, D. (ms.) *Climbing the Mountain* [Version 7 June 06].
- Parfit, D. (1987) *Reasons and Persons* (2nd ed.). Oxford : Clarendon Press.
- Quinn, W. (1990) 'The puzzle of the self-torturer' *Philosophical Studies*, **59**:1.
- Railton, P. (2003) 'Alienation, Consequentialism, and the Demands of Morality' *Facts, Values and Norms*. New York : Cambridge University Press.
- Sorensen, R. (2004) 'Paradoxes of Rationality' in Mele, A. and Rawling, P. (eds.) *The Oxford Handbook of Rationality*. New York : Oxford University Press.
- Weirich, P. (2004) 'Economic Rationality' in Mele, A. and Rawling, P. (eds.) *The Oxford Handbook of Rationality*. New York : Oxford University Press.

[1](#) Nida-Rumelin, p.13.

[2](#) Harsanyi, p.122, seems to be getting at a similar idea in his discussion of the "normal mode" of playing a game.

[3](#) See, e.g., Kolodny, pp. 509-510, and Parfit (ms.), p.21.

[4](#) Of course we can imagine special circumstances whereby one's holding of a belief would itself be the reason-giving 'fact' in question. If I am sworn to honesty, then the fact that *I believe that P* may itself provide a reason for me to assert that P.

[5](#) I leave open the question whether such value is impersonal or agent-relative.

[6](#) Parfit (ms), p.21. I also follow Parfit's use of the term "rationally ought", below.

[7](#) Kolodny, p.509, italics omitted.

[8](#) Cf. Kolodny's "transparency account" of rationality's apparent normativity (p.513).

[9](#) I owe this idea to Clayton Littlejohn.

[10](#) Of course, the agent may not know of any more reliable option than to follow the guide. As noted above, we must move beyond the first person perspective before this distinction will seem significant.

[11](#) This is a familiar distinction, see e.g. the Stanford Encyclopedia entry on 'Rule Consequentialism', <http://plato.stanford.edu/entries/consequentialism-rule/#4> [accessed 21/6/06].

[12](#) The following is strongly influenced by R.M. Hare.

[13](#) Hare, p.38. This is distinct from "rule utilitarianism", which I take to be the claim that whatever the rules advise is *ipso facto* what we ought to do. The example below should make the difference clearer.

[14](#) This is a generalization of the sorts of transmission principles that Parfit (1987) rejects.

[15](#) Parfit (1987), pp.31-35. Cf. my section on 'meta-coherence' below, where I argue that this is not really "wrongdoing" at all from the perspective of subjective rationality. However, the current section is concerned with the objective, fact-based sense of 'ought', independent of what the agent happens to know.

[16](#) *Ibid.*, p.32. (N.B. Here I quote the words that Parfit attributes to his fictional agent 'Clare'.)

[17](#) This arguably means that the so-called "locally optimal" action isn't optimal at all: the closest possible world where the agent so acts is a worse world overall, because in that world she acts worse in *other* situations. But this raises complicated issues about the appropriate comparison class for the relevant counterfactuals, which goes beyond the scope of this essay. See Dancy, pp.9-10.

[18](#) Parfit (1987), p.13. Though the more radical (non-epistemic) cases he discusses are better covered by my treatment of "essential byproducts" below.

[19](#) I owe the idea of "meta-coherence" to Michael Huemer.

[20](#) Of course, an injustice might *in fact* yield the best results. But this section is discussing the *evidential* question, i.e. *expected*, rather than *actual*, utility.

[21](#) Two further points bear mentioning: (1) We might construct a new distinction in this vicinity, between *prima facie* and *all things considered* judgments, where the former allows only first-order evidence, and the latter includes meta-beliefs about reliability and such. This bears some relation to 'local' vs. 'global' considerations, and again I think the latter deserves to be more widely recognized. Nevertheless, I take it to be distinct from the "momentary act vs. temporally-extended agent" form of the local/global distinction, which this essay is concerned with. (2) Even though a meta-coherent local calculation should ultimately reinforce the indirect strategy, that's not to say that one should actually carry out such a decision procedure. The idea of indirect utilitarianism is instead that one acts on the dispositions recommended by our practical morality, rather than having one "always undertake a distinctively consequentialist deliberation" [Railton, p.166]. So my local/global distinction could do some work here after all, in the manner described below.

[22](#) Elster, pp.43-52.

[23](#) Cf. Parfit (ms), p.30.

[24](#) Musgrave, p.21.

[25](#) This evidence-based notion is another common use of the term “reasons”. But in light of my earlier remarks, I should instead say that objective reasons are provided by the ultimate facts, not mere “indications” thereof. The more subjective or evidence-based “reasons” might instead be conceptually tied to rationality, as per my “quasi-reasons” below.

[26](#) Indeed, Parfit (1987) takes this as the “central claim” of the self-interest theory.

[27](#) Railton, p.156.

[28](#) Dancy, p.18.

[29](#) Thanks to Jeremy Shearmur for suggesting this example to me.

[30](#) For example, one would plausibly have reason not to entertain the good as a goal. But one could not *recognize* this reason without thereby violating it, for it would only move an agent who sought the very goal it warns against.

[31](#) Nozick, p.41.

[32](#) Kavka, pp.33-36.

[33](#) Of course, the atomist will allow that facts about other times can provide reasons in the present. But in the example discussed, there seems an important sense in which the reason *itself* is not present – except, perhaps, through a kind of imaginative projection on the part of the globally rational agent.

[34](#) From a local perspective, the state of intending is worth having, even though the object (e.g. the act of drinking Kavka’s toxin) does not in itself merit intending. The problem arises because we regulate our mental states on the basis of object-based reasons alone (as seen by the impossibility of inducing belief at will). The global perspective enables us to overcome this by treating past reasons for intending as present reasons for acting, and hence transforming state-based reasons into object-based ones.

[35](#) For more on the significance of rational resolutions, see McClennen, pp.24-25.

[36](#) Sorensen, p.261.

[37](#) Quinn, pp.79-90.

[38](#) Sorensen, p.270.

[39](#) Weirich, p.391.