
Article:
Lenman, J.W. (2017) Reasons without humans. Analysis, 77 (3). pp. 586-595. ISSN: 0003-2638

https://doi.org/10.1093/analys/anx071

Reasons without Humans

JAMES LENMAN

1.

Brian Hedden, in this impressively learned and ingenious, if somewhat maddening book[1], defends a view he calls *Time-slice Rationality*, a view comprising two central claims. They are these.

> *Synchronicity*: All requirements of rationality are synchronic. (8)
>
> *Impartiality*: In determining how you rationally ought to be at a time, your beliefs about what attitudes you have at other times play the same role as your beliefs about what attitudes other people have. (9)

Hedden begins by objecting to principles of rationality he deems dubiously diachronic. The first is *conditionalization*. This tells me that if I believe the conditional probability of H given E to be x, then, should I discover E to be true, I should believe H to have probability x. The objection to this that it involves an unmotivated conservativism: it gives weight after you learn new information to what you believed beforehand. Hedden also complains that it is unhelpful about what to do in cases where we forget stuff or otherwise lose evidence. It therefore at least 'cannot be the whole story when it comes to rational belief change.' (p. 43) He then considers and rejects an analogous principle for preferences:

> It is a condition of rationality that ultimate preferences – preferences over maximally specific possibilities – do not change over time. (47)

He goes on to urge rejection reflection principles that tell us to defer to our future selves, both with respect to preferences and with respect to beliefs.

Those who believe in synchronic rationality fear that without it we may be led to perform what Hedden calls *tragic* sequences of actions where a tragic sequence is a sequence such that at all times we would rather be performing some other sequence. Tragic attitudes are attitudes that open you to the risk of performing a tragic sequence. Thus someone departing from the epistemic conservativism implicit in conditionalization opens herself to exploitation through a Dutch book (a set of bets that together guarantee a loss). And instability in our preferences opens us to courses of action that tragically defeat our own purposes. Thus we can imagine a Russian nobleman (in a variation of Parfit's classic example) who in his liberal youth donates generously to liberal causes and in his conservative old age again donates generously to conservative causes where the donations cancel each other out in such a way he might as well have kept his money. Hedden thinks to avoid this worry for Synchronicity by arguing that the rational *ought* does not apply to *sequences* of actions. Having stable intentions, he suggests, is rather like having a good memory. It's a desideratum, nice way to be, but we are not *irrational* when we fail to be this way. As far as the theory of rationality is concerned this is a desideratum we can simply, as we might say, *outsource*. Failing to satisfy it is a cognitive and practical misfortune perhaps but not a form of irrationality.

---

[1] Brian Hedden: *Reasons Without Persons: Rationality, identity and Time* (Oxford: OUP, 2015). References are to this book unless otherwise specified.

He then tells us what he thinks are some correct principles of rationality. They are:

> *Uniqueness*: Given a body of total evidence, there is a unique doxastic state that it is rational to be in. (130)

And:

> *Synchronic conditionalization*: Let P be the uniquely rational prior probability function. If at time t you have total evidence E your credence at t in each proposition H should equal P(H/E). (138)

Where preferences are concerned, Hedden urges that we accept *preference uniqueness*:

> *Preference uniqueness*: Given a body of total evidence, there is a unique set of (ultimate) preferences that it is rational to have. (149)

Where ultimate preferences are preferences over maximally specific possibilities, complete possible worlds. In fact he prefers a strong version of this claim that drops the relativisation to bodies of evidence. He also proposes a *Principle of expert deference*: where an expert is a person who is perfectly rational and has strictly more evidence than you do, you should agree with any expert about anything.

Hedden acknowledges that his time-slice view of rationality is only very promising for propositional justification (a matter merely of whether your evidence supports your beliefs) and will not so easily handle doxastic justification (whether the basis on which you hold your beliefs is a sound one). The latter may for example make reference to whether one possesses a general disposition to belief what one's evidence supports and whether one has deliberated in accordance with such a disposition in arriving at a given belief. Here he considers two options, the first being to outsource doxastic rationality and limit the scope of his theory to propositional justification; in the second, taking his inspiration from Williamson, Hedden suggests that time-slice centric norms of propositional justification are primary, other, less plausibly time-centric norms that may govern doxastic justification are derivative from these. Epistemology is 'time slice first' (182).

Hedden's theory is addressed to the snapshot concern, *Do we get it right?* He has more or less nothing to say about the question *how* we get there, or fail to. That is a matter of *reasoning*, the stuff we need to do to be at all successful. Well, yes, Hedden argues, but while we need to do it, imaginary ideal creatures might not. Imaginary ideal creatures just get it right have no need of the reasoning mechanisms on which we, as a matter of our contingent limitations, depend. And it is appropriate, Hedden urges, that the theory of rationality should focus on what rationality requires as a matter of necessity on rational creatures of whatever kind. The contingencies of how we go about satisfying these constraints are something else he is happy to outsource.

2

A central aim of Hedden's in this book is to improve the extensional adequacy of our understanding of rationality to cover various more or less fanciful thought experiments involving, inter alia, teletransportation, people who split in two, or have their beliefs and memories tampered with in strange and fanciful ways. My core worry, adumbrated in my title, is that in his eagerness to do

justice to the fanciful, Hedden loses sight of the everyday. This latter failing is evident, for example, when it comes to the principle of expert deference. The principle is interesting but of no practical interest whatever. For there are of course *no* experts as Hedden defines experts and there are never likely to be. So the principle of expert deference, which tells me to adjust my beliefs so as to align them with the beliefs of experts in effect tells me nothing. If, as Hedden proposes at the outset (10-12) the point of the concept of rationality is to help us evaluate, predict and guide actions, the principle of expert deference looks, at least for real human beings, decidedly pointless. The unreality of the principle comes to the fore when Hedden address the worry, What if experts disagree? More precisely what if two experts have different credences for the same proposition. In those circumstances doesn't the principle offer inconsistent advice?

Well, sure, experts can disagree is by that we mean regular experts of the sort you and I might meet in the university cafeteria. But the principle of expert deference is not about *them*, it's about experts in Hedden's idealised sense, people who are perfectly rational with strictly more evidence than you have. And such people, apprised as they necessarily are qua rational with the uniquely rational prior probability function, cannot knowingly disagree where each knows the other's credences: this follows, Hedden tells us, from a result proved by Robert Aumann. And without relying on Aumann, Hedden also offers his own reasoning to the consistency of the principle assuming *inter alia* that not only the experts but you yourself are perfectly rational.

OK, OK, the reader wants to know. But what about real people in real life? Where none of us is rational as Hedden understands it and there are no experts in his highly ideal, technical sense. Well, we are told:

> As for real life experts, I suspect that there will be no exceptionless principle for how to take their opinions into account, but in most cases the way in which you ought to defer to the opinions of real-life experts will more or less approximate that given in our formal principle.(167)

I'm not at all clear, given how extreme the idealisation, how strong the assumptions Hedden needs to make his principle consistent, what that can even mean. Hedden is clearly most comfortable with these rarefied levels of idealisation as they 'offer the benefit of making things more precision and formally tractable' (167) but this level of remoteness, indeed divorce, from the realities of human experience seems rather a high price to pay for precision and formal tractability.

More unreality confronts us where the discussion of practical rationality takes it point of departure with expected utility theory. While certainly precise and formally tractable, taken as a theory of rationality, this is not immensely informative. Expected utility theory tells you what to do if you are an agent whose preferences satisfy some rather exigent completeness, transitivity, continuity and independence axioms. Actually it doesn't even do that, as Hedden observes (94), unless it is taken to embrace an imperative to maximize expected utility. And even that imperative is doubtful: wide-scopers about rationality favouring a safer but still less helpful imperative in effect to either maximize expected utility or reconsider. Of course outwith perhaps certain very artificial toy circumstances, it is a safe bet that no human being has ever arrived at this attractive condition of perfect comprehensiveness and coherence in his preferences. Should any ever do so there would be a good way to describe her deliberation. Completed, done, finished. Expected utility theory as such is silent about the deliberative process that is supposed to get you there.

Divorce from reality only deepens when we think of the preferences of rational agents, as Hedden urges we do, as *ultimate* preferences, imposing order on maximally specific possibilities, in effect fully specified possible worlds. (See in particular, 46 and section 8.2.3.) Once again, human beings don't have these. Fully specified possible worlds are just too large and complex to be, at least singly, objects of thought at all, excepting only perhaps the actual world – demonstratively – and a few extremely spare and simple toy worlds of interest only to metaphysicians.

Likewise with the epistemic case, conditionalization principles offers instruction to those who are already extremely credally opinionated what adjustments to make in the light of new information. But of what credal opinions we should adopt in the first place it does not speak. Again as an account of epistemic rationality this might reasonably be thought to leave rather a lot out. But if conditionalization is not very helpful, synchronic conditonalization is less so. It basically tells us to have the credences we rationally ought to have given our evidence. This is good advice but, as a theory of epistemic rationality, a little empty.

3.

But while standard expected utility theory and standard conditionalization are where Hedden starts out, as the book procedes we leave them behind for the more purely synchronic principles he urges we should prefer. In effect Hedden urges we abandon a picture of rationality that says, *Stick to your guns!* for a picture of rationality that says, *Get it right!* Or a little more particularly, *Get it right, never mind how!* Fussing about the how is after all only an issue for creatures like ourselves with our contingent limitations. Imaginary ideal creatures, 'creatures intellectually superior to ourselves' with 'no need for reasoning' (185), don't need to bother. They just get it right. The value of reasoning is instrumental and gets outsourced. Ideally rational creatures just get it right.

This might be true for example of *Professor Instinct*. Professor Instinct, let's suppose, doesn't go in for any reflective reasoning of any kind ever. He just gets pushed around by his instincts like a brute beast. But his instincts are good instincts. Whoever designed him designed him extremely well. His instincts guide him so reliably that he reliably decides what the uniquely correct utility function says he should decide and believes what the uniquely rational prior probability function tells him to believe on the basis of his evidence. I guess he himself doesn't need to know what the uniquely rational prior probability function is so long as the system that generates his instinctive beliefs is implementing it correctly. Indeed if what matters is getting it right he doesn't need to know what his 'evidence' is. If we can outsource everything else, we can surely outsource that too and count as evidence any kind of causal input to the system that furnishes it (it, not him) with information about the world.

Do we want to call Professor Instinct rational? It seems a little odd to. It is a very natural commonplace pointedly to contrast reason and instinct. But Professor Instinct satisfies what Hedden takes to be the strict requirements that any being must satisfy to count as rational. So I guess Hedden has to say, by his own lights, that Professor Instinct is rational. And maybe fair enough. But what about *Professor Lucky*? Professor Lucky is a what we might call a *Randomizer*, someone who arrives at her beliefs and decisions by some entirely random procedure. Of the many possible Randomizers, most do very badly, getting almost everything wrong almost all the time. But there are a small minority who get lucky and do pretty well. A *fantastically* lucky *very* tiny minority, one in a few squillion perhaps, do just perfectly and get from one end of life to the other believing and

deciding exactly as the uniquely correct utility function and the uniquely rational prior probability function would tell them to. Professor Lucky, the lucky so-and-so, happens to belong to that very tiny minority. If what matters is getting it right, Professor Lucky gets full marks. But is she rational? Well, surely not. She doesn't form her beliefs and decisions by any kind of reliable mechanism or process. But a reliable mechanism or process again is surely a merely *instrumental* good, one that almost all of us need, but that Lucky Randomizers like Professor Lucky do not. To be sure she couldn't have known in advance how lucky she was going to be but that only has her needing a reliable mechanism in the rather attenuated sense of need in which someone who, as it turns out, will never be afflicted by a fire, can be said to need fire insurance.

Professor Lucky is surely not rational in any meaningful sense. But she gets it right. If that's the end and we can outsource the means, it looks wrorryingly like we can outsource basically everything including everything we ordinarily take to distinguish the rational from the irrational. At least in the *epistemic* case we can. In the *practical* case, it won't so obviously give us all we want to have decisions that conform to what the uniquely correct utility function would demand without it mattering where and how they originate. It won't give us all we want if, for example, we are Aristotle and think the normatively authoritative human *ergon* is *ψυχῆς ἐνέργεια κατὰ λόγον*[2], 'rational activity of the soul'. More generally, it might credibly be supposed that autonomous practical reasoning is more than an instrumental good but a central part of what we value in our lives such that we would be properly unwilling heteronomously to outsource our practical reasoning to some external agency however perfectly reliable in making correct decisions.[3]

This worry expands naturally to Hedden's suppostion that we see rationality as time-slice first, the primary norms if perhaps not the derivative being time-slice centric. Here again, whatever we think of the epistemic case the practical case looks problematic. Again if you are Aristotle the end of practical reason is *eudaimonia* and that isn't something you have or fail to have at a moment but an enduring, settled state. If you are Hume, perhaps, what you are after is a life that, viewed as a whole, is able to bear your survey.[4] Of course on a one time-slice-friendly reading of expected utility what matters is the maximal satisfaction of your preferences now. But Hedden can't very credibly accept that as it amounts to the present-aim theory of rationality that is the orginal target of Parfit's Future Tuesday supposed reductio that so impresses him.[5] And of course Hedddon ultimately only accepts expected utility theory insofar as ones preferences match up with the uniquely correct utility function. And what this tells me to do is aim for the whole world to be as good as possible. As the world is extended in time that again is not really a time-centric basic norm. Still, it might be countered, my basic concern is to have, at any given time a utility function that matches up with the ultimately correct one. But surely not. For one thing if what matters is that the world be as objectively good as possible, surely what I ought to do is have the utility function I have in whatever attainable-by-me world is best. And that might *not* be the *correct* utility function. The correct utility function Uc might surely be *self-effacing*: the best way to attain the best outcome might be to instantiate some other, quite different utility function Uo. For another, even prescinding from such discomfiting possibilities, the desirability of my having at any given time a utility function that

[2] *Nicomachean Ethics* 1098a.
[3] Griffin 1986, p. 9. Crisp 1997, pp. 61-2.
[4] *Treatise* 3.6.6.
[5] Parfit 1986, pp. 123-4.

matches up with the correct one is surely not itself normatively fundamental. If it's better that a peace deal is struck than the war continue, then it's ordinarily no doubt better if people with hands on the relevant causal levers so prefer. But we want them to prefer that way because of what we want to happen, not vice versa. With epistemic rationality our ultimate aim is simply truth. Direction of fit is word-world and we seek simply to understand and not to change the world. So time-slice centeredness makes some sense. But with *practical* rationality, where the direction of fit is reversed the final aim is not – at least not for the sort of consequentialist perspective that appeals to Hedden, to get our preferences and decisions right but that the world be as good as we can make it not just now but in the very long term.

4.

Hedden's discussion of preference uniqueness will raise many eyebrows. It comes in two parts. The first is a discussion of the approximately Humean claim that there are no substantive (as opposed to formal) rational constraints on preferences. Here he first considers only to reject Broome's well-known argument against so-called 'moderate Humeanism' and then considers and endorses PArfit's argument from the irrationality of 'future Tuesday indifference'. He then argues (simplifying a little) that preference uniqueness is not threatened by Lewis's famous desire as belief result providing we restrict the desires/preferences we consider to those ultimate desires/preferences that apply to maximally specific possibilities. He then argues:

> Plausibly, the fundamental normative facts, such as which moral theory is correct, are a priori. And arguably, ideally rational agents are certain of all a priori facts. After all an agents evidence always a priori entails these facts and it is natural to think that an ideally rational agent will be certain of everything that is a priori entailed by her evidence. Putting these two things together, we get the result that whenever one world is better than another, an ideally rational agent will be certain that the one world is better than another. (160)

That is so breathtakingly quick and dirty a passage of argument for a claim at once so extremely strong and so central to Hedden's understanding of the rational that it is hard to know where to begin. For one thing we can note again how little bearing it can credibly have on the evaluative, predictive or deliberative activities of human beings given the wildness of the idealization involved. For another thing it is not just highly arguable that the fundamental normative facts can be known a priori but that there are any, or any very interesting, fundamental normative facts if by that is understood normative facts that float free of any dependence on particularities of human nature and human society.[6] For another thing, I'm not sure how to take 'entail'. Hedden takes himself to be opposing the approximately Humean view that there are no substantive (as opposed to formal) rational constraint on preferences. In fact you might not believe you need to take that anti-Humean line to defend something like uniqueness. R. M. Hare famously believed you could get your preferences into correct alignment with what they rationally ought to be given only the minimal start-up kit of 'logic and the facts';[7] but I think almost nobody now believes that form of moral logicism. That may mean something a bit meatier than logical entailment will be needed and it would be nice to know what. The transition from 'is' to 'ought' is a notoriously tricky one and it would be nice to know how ideally rational agents do it. Perhaps they just have compelling intuitions

[6] See Rawls 1972, pp. 159-160, Lenman 2000, pp. 361-362.
[7] Hare 1981, pp. 6, 101ff

which reliably track independent normative truths. After all, Hedden doesn't think rational beings need to actually *reason* their way to getting things right so long as they get things right. (Even if in this case that might involve getting very lucky indeed.[8]) But we mere humans, with our conflicting and uncertain intuitions do need to do some reasoning to determine what to do. And Hedden really has nothing to tell us about how that might work.

The phrase 'such as which moral theory is correct' in the passage just quoted is a telling one. Right at the start of the book, Hedden notes that morality might be understood in ways unfriendly to time-slice rationality. Ideas like promissory obligation or the Rawlsian doctrine of the separateness of persons seems to depend on taking the relation of personal identity over time very seriously. He proposes two ways out of this worry. The first is another outsourcing stratagem that distinguishes morality from rationality. The second elects to understand morality along some rigorously utilitarian lines that concerns itself only with how much welfare there is without caring how it gets distributed over persons or times. By the time we get to the passage just discussed he seems to have come down pretty clearly in favour of the latter escape, including moral facts among the normative facts the ideally rational agent is supposed, in virtue of being an ideally rational agent, in a position to know a priori. But there is no argument here or anywhere else in the book to support this extremely strong claim about how morality should be understood or to address the many serious objections to it.

5

Hedden's critique of Humeanism is also a little puzzling to me. What has me puzzled is the central role he gives to Parfit's example of future Tuesday indifference. We all remember how it goes:

> A certain hedonist cares greatly about the quality of his future experiences. With one exception, he cares equally about all the parts of his future. The exception is that he has *Future-Tuesday-Indifference*. Throughout every Tuesday he cares in the normal way about what is happening to him. But he never cares about possible pains or pleasures on a *future* Tuesday.[9]

It is a nice example of very plausibly irrational preferences. But notice something about the irrationality that is involved. It is *diachronic*. It is a failure of stability over time in FTI-person's pro-attitudes to Tuesday pleasures and pains. He doesn't care about future Tuesday pleasures and pains when the Tuesday in question is future but when Tuesday arrives he cares about them 'in the normal way'. And from reading chapter 7 I thought I understood what Hedden had to say about this kind of instability. What I thought he wanted to say was, It is very nice not to be like that because if you are like that all manner of suboptimal stuff will happen to you. But it is like having a good memory: it's nothing whatever to do with *rationality*. Which makes the centrality bestowed on the example puzzling. What got outsourced in chapter 7 should surely stay outsourced in chapter 8. We might perhaps try to eliminate the element of diachronic instability by trying to imagine a person who *even on Tuesday* is indifferent to what happens to him on that very Tuesday. This character, the consistently Tuesday indifferent person would be a very strange creature indeed. So strange that, as Sharon Street persuasively urges in her highly instructive discussion of the argument, we really have

[8] See Street 2006.
[9] Parfit 1984, p. 123-4.

no grounds confidently to write him off as irrational once we seriously try to imagine what such a person would be like.[1011]

There are three crucial things Hedden thinks we want the notion of rationality to do for us. (10-12) We want to use it to evaluate thoughts and actions of thinkers and agents, ourselves and others, we want to use it to predict and explain them and we want to use it to think and deliberate what actions to perform and what beliefs to adopt. He goes on to offer a defense of his two key claims, synchronicity and impartiality, which relentlessly pursues two thematic strategies: idealise like crazy and outsource almost everything. My fear for him is that he pursues them to an extent that the resulting picture of rationality is of very limited interest to us very unideal human beings in trying to understand our own our very unideal efforts to do these very crucial things.

Department of Philosophy, University of Sheffield,
45 Victoria Street, Sheffield S3 7QB
j.lenman@sheffield.ac.uk

References

Aristotle. *Nicomachean Ethics*. Many editions and translations.
Crisp, Roger. 1997. *Mill on Utilitarianism*. London: Routledge.
Griffin, James. 1986. *Well-Being: Its Meaning, Measurement and Moral Importance*. Oxford: OUP.
Hare, R. M. 1981. *Moral Thinking: Its Levels, Method and Point*. Oxford: OUP.
Hume, David. *A Treatise of Human Nature.* Many editions.
Lenman, James. 2000. Consequentialism and Cluelessness. *Philosophy and Public Affairs* 29, 342-370.
Parfit, Derek. 1984. *Reasons and Persons*. Oxford: OUP.
Rawls, John. 1972. *A Theory of Justice*. Oxford: OUP.
Street, Sharon. 2006. A Darwinian Dilemma for Realist Theories of Value. *Philosophical Studies* 127, 109-166.
Street, Sharon. 2009. In Defense of Future Tuesday Indifference: Ideally Coherent Eccentrics and the Contingency of What Matters. *Philosophical Issues* 19, 2009, 273-298.

[10] Street 2009.

[11] Hedden might perhaps just accept this, saying he doesn't strongly insist on the strong version of preference uniqueness and a weaker, more relativistic version would be adequate to his purposes. Indeed he claims in a footnote (p. 158), though he does not argue, that preference uniqueness is perfectly consistent both with expressivist and subjectivist understandings of metaethics. In that case he may have no quarrel with someone like Street. But then I start to lose my grip on why he takes himself, as he pretty clearly does to have a quarrel with the Humean to which the appeal to the future Tuesday example is relevant.