Discussion Notes

Self-Respect Regained

Jake Chandler and Adam Rieger

In a recent article, David Christensen casts aspersions on a restricted version of van Fraassen's Reflection principle, which he dubs 'Self-Respect' (SR). Rejecting two possible arguments for SR, he concludes that the principle does not constitute a requirement of rationality. In this paper we argue that, not only has Christensen failed to make a case against the aforementioned arguments, but that considerations pertaining to Moore's paradox indicate that SR, or at the very least a mild weakening thereof, is indeed a plausible normative principle.

In 'Epistemic Self-Respect' (2007), David Christensen seeks to cast doubt on the Synchronic Reflection principle, which states that one's credence in A, conditional on one's credence in A being r, should be r:

Synchronic Reflection (SR): $cr(A \mid cr(A) = r) = r$.

He suggests that one can conveniently think of SR as standing for 'Self-Respect' as well as 'Synchronic Reflection'; for to violate it is to regard one's own current credences in A as unreliable.[1]

Christensen's strategy is to consider, but reject, two arguments for SR. He concludes ultimately that SR is not a norm of rationality, and indeed that on occasion rationality demands its violation. The only rationality norms in this area are those that already follow from probabilistic coherence, and there is nothing special about the first-person perspective codified in SR; one should have no more respect

[1] van Fraassen's Reflection principle (1984) states that, as a matter of rationality, one's current credence, at time t, in a proposition A, conditional on one's credence in A at some *possibly future* time t^* being r; should already be r; thus $crt(A \mid crt^*(A) = r) = r$. As many authors have observed, there are numerous cases in which this more general principle seems false, at least in its unqualified form; see Bovens (1995) for various counterexamples. As far as we are aware, Christensen is the first author to throw doubt on the weaker, synchronic version of Reflection.

Proceedings of the Aristotelian Society, Vol. CXI, *Part* 3
doi: 10.1111/j.1467-9264.2011.00322.x

for oneself than for anyone else to whose credences one happens to have equally good access.

In this note we question Christensen's conclusions, and argue that SR (or something in its vicinity) is indeed a rationality norm.

I

The Argument from Introspection. Christensen first considers an argument that SR follows from probabilistic coherence plus an agent's perfect knowledge of their own credences. In particular, he cites the following two principles, arguing that SR follows from them:

Confidence: For any n, $cr(cr(A)=n)$ is either 0 or 1.

Accuracy: If $cr(cr(A)=n)=1$, then $cr(A)=n$, and if $cr(cr(A)=n)=0$, then $cr(A)\neq n$.

We first note a peculiarity in the proof given by Christensen. Suppose, for some x, we have $cr(A|\,cr(A)=n)=x$. Christensen observes that, by Confidence, $cr(cr(A)=n)$ must be 0 or 1. He continues: 'In the former case, the conditional probability ... would be undefined, so it must be that $cr(cr(A)=n)=1$' (Christensen 2007, p. 326), and then works his way to the conclusion that a probabilistically coherent agent satisfies SR.

The quoted sentence looks suspicious, and indeed the step taken there is fallacious. It is analogous to the following:

Problem: Solve $xy=y$ for x and y, given that y = 0 or 2.
Solution: Note that $x = y/y$. If $y=0$, this is undefined, so y must be 2; hence $x=1$ and $y=2$.

But this clearly leaves out the family of solutions with $y=0$ and $x=$anything. In an exactly similar way, Christensen's argument ignores what may happen if $cr(cr(A)=n)=0$. He states that the conditional probability is then 'undefined', but there are well-known axiomatizations of probability, for example via Popper functions (Popper 1959), which do allow this. Indeed, suppose we have $cr(A)=0.5$, $cr(cr(A)=0.5)=1$, and $cr(A|\,cr(A)=0.5)=0.5$. We could set $cr(A|\,cr(A)=r)=0.7$, say, for every $r\neq 0.5$ without any violation of probabilistic coherence, since $cr(cr(A)=r)$ must be 0 for every $r\neq 0.5$.

Thus SR in full generality does *not* follow from Accuracy and

Proceedings of the Aristotelian Society, Vol. CXI, *Part* 3
doi: 10.1111/j.1467-9264.2011.00322.x

Confidence.[2] Christensen should either formulate SR with the condition $cr(cr(A)=r) \neq 0$ explicitly stated, or equivalently in the 'multiplied out' form:

$$cr(A \wedge cr(A)=r) = r \times cr(cr(A)=r)$$

With this proviso, Synchronic Reflection *does* follow from Accuracy and Confidence. So far, however, we are no closer to settling the question of whether SR is a norm of rationality, since it is unclear whether Accuracy and Confidence are.

Christensen himself rejects Confidence, and with it the argument. But his rejection is ill-founded. Indeed, in a remarkably strong claim he says that 'an agent who satisfied Confidence would *thereby* fall short of ideal rationality' (2007, p. 328). His reason is that even if an agent satisfied Confidence, they would not be in a position to be certain that they did so. But this seems bizarre: it is surely at least *possible* (if not obligatory) that an ideal agent have this certainty. We should not demand that God, say, should refrain from satisfying Confidence on pain of irrationality. We are attempting to define the properties of an ideal agent here, not a hybrid of ideality and human frailty.[3]

II

The Dutch Book Argument. As Christensen notes, SR has received justification in the form of a so called 'synchronic Dutch book argument' (DBA). The argument proceeds from a theorem to the effect that an agent whose credences are probabilistically coherent is vulnerable to exploitation via a particular kind of betting strategy if and only if she violates SR. Since such a vulnerability is considered to be undesirable, we thereby have a reason to adhere to SR.[4]

---

[2] This is carefully noted by Sobel (1987), whom Christensen cites as the basis of his proof.

[3] We note also that Milne (1991) gives a direct Dutch book argument for (a principle equivalent to) the conjunction of Accuracy and Confidence, and that Samet (1997) shows that Confidence does not follow from SR, so that rejecting the former does not automatically require the rejection of the latter. We do not have space to discuss these matters here.

[4] The nomenclature sometimes distinguishes between Dutch book arguments and Dutch *strategy* arguments, of which the present argument is an instance. We have opted for a less discriminate terminology here.

It is worth noting that, when stating the argument, Christensen overlooks the left-to-right half of this theorem, known as a 'converse synchronic Dutch book theorem'. But, as

doi: 10.1111/j.1467-9264.2011.00322.x

The proof of the right-to-left half of the theorem can easily be adapted from the well-known proof of its analogue for full Reflection (see, for instance, Skyrms 1987, p. 14). The converse is also quick to establish.

To illustrate the kind of vulnerability that is exhibited by those who violate SR, Christensen asks us to suppose that he (i) is probabilistically coherent and (ii) has $cr(A|cr(A)=0.5)=0.1$. Christensen again assumes the ratio definition of conditional probabilities and deduces $cr(cr(A)=0.5) > 0$, equal to n, say.[5] The bets offered by the Dutch bookie are as follows:

1. A bet on $cr(A)=0.5$: Christensen's gain is £$(1-n)$ if $cr(A)=0.5$, and his loss is £n otherwise.
2. A bet against A, *conditional* on its being the case that $cr(A)=0.5$. If $cr(A)\neq 0.5$, the bet is called off. Otherwise, he loses £9 if A, and gains £1 if not-A.
3. A bet on A, which is *only offered* if $cr(A)=0.5$. In this case, he wins £5 if A, and loses £5 otherwise.

His credences guarantee that he finds 1 and 2 fair, and finds 3 fair if it is offered, because in that case $cr(A)=0.5$. It is simple to verify that if $cr(A)\neq 0.5$, he loses £n, and if $cr(A)=0.5$, he loses £$(3+n)$.

Now although Christensen admits that he is guaranteed a loss here, he denies that this shows he is irrational, for this is not a DBA of the usual sort, in which the set of bets accepted is logically guaranteed to be loss-making.

> Consider the case where I don't actually have 0.5 credence in A. Here, the bookie does not offer me Bet 3, so the relevant set of bets is just 1 and 2. But this set is not one whose payoffs logically guarantee my loss. True, they'll cost me money in the actual world. But in a world where I do have credence 0.5 in A, and where A is false, I win £$(2-n)$ on these bets. Thus the bookie can know he's going to profit in this case only because he knows that my credence for A is not 0.5. But that is a contingent fact about the world that, by hypothesis, I don't know. So it's true that the bookie can profit from me by knowing nothing more than my credences. But since the subject matter of the bets in-

is well-known, this is an essential part of the argument: the Dutch book vulnerability of those who violate SR can hardly be an incentive to comply with the principle if those who do so face the very same prospect.

[5] Once again the argument will not work if we assume a more general conception of conditional probability; in this case $n > 0$ must be taken as an independent assumption.

Proceedings of the Aristotelian Society, Vol. CXI, *Part* 3
doi: 10.1111/j.1467-9264.2011.00322.x

> cludes those very credences, the bookie is profiting by his knowledge of contingent facts beyond my ken. And that sort of guaranteed profit should not be seen as any indication of irrationality on my part. (Christensen 2007, pp. 329–30)

There seem to be two strands to Christensen's objections to the DBA here: firstly that the book is not truly Dutch, in that it does not guarantee loss at every world, and secondly that the bookie is somehow using knowledge in a way that is unfair. On the first, it is true that bets 1 and 2 jointly yield a profit in a world where $cr(A) = 0.5$ and A is false. But it is hard to see why this is relevant, since in that world *he would have been offered and accepted bet* 3. Whatever his credence in A, and whatever the truth value of A, in a world, he makes a loss at that world, and this is sufficient.

On the second point, it is true that there is an appearance of trickery here. Normally we allow that the bookie know the agent's credences (so they can offer the appropriate bets) but not that they know the truth values of the propositions betted on (else it is too easy to win: just make sure the agent takes the losing side in any bet). But here, where the agent's own credence is the subject of Bet 1, these principles seem to pull in different directions.

But we must proceed carefully here. What is clearly objectionable, with respect to knowledge of truth values, is for not-P to constitute a necessary and sufficient condition for the bookie's proposing a bet on P. But in the present case, we have no such goings-on. On the one hand, we have a bet that is offered on condition that $cr(A) = 0.5$, but whose payoff does not hinge on whether $cr(A) = 0.5$ (Bet 3). On the other hand, we have a bet on whether $cr(A) = 0.5$, but which is offered unconditionally (Bet 1). The bookie does not use their knowledge of $cr(A)$ to win any of the bets (as they could do, but which would be unfair); instead, such knowledge is only used in constructing the book, as in other DBAs. Consequently the bookie cannot be convicted of any unfairness.

III

Moorean Problems. Having dismissed the DBA, Christensen suggests replacing SR by what he calls 'moderate' self-respect, MSR: roughly, that an agent who comes close to satisfying Confidence and Accuracy should come close to satisfying SR. But, as Christensen

Proceedings of the Aristotelian Society, Vol. CXI*, Part* 3
doi: 10.1111/j.1467-9264.2011.00322.x

himself points out, this is merely a matter of obeying probabilistic coherence. Thus MSR is really no additional principle at all. Agents will tend to satisfy SR approximately because they tend to have good access to their credences, but Christensen does not think this has any normative force: we just happen generally to have better insight into our own than into others' credences. There is nothing more to respect in one's epistemic self (beyond the trivial fact that one's own credences necessarily coincide with themselves but may diverge from those of others).

Dutch book arguments are controversial. But simpler considerations suffice to show that Christensen's line here cannot be right. Since perfect self-knowledge entails Reflection, a failure to obey the latter entails a defect in self-knowledge. But one might think, with Christensen, that such a defect is a purely epistemic one, like having bad eyesight. What makes it a matter of *rationality*?

Consideration of parallel questions in the case of full belief can help here. Suppose I have a Moorean belief such as

M: It is raining, and I do not believe it is raining.

Belief in *M* seems bad—indeed, irrational. Why so, when *M* is consistent? The literature on Moore's paradox includes many attempts to answer this question. The following strikes us as plausible.[6] Although *M* is consistent, *it cannot be true if I believe it* (for then I believe the first conjunct, contradicting the second conjunct). Thus if I believe *M*, then necessarily I have a false belief, not because *M* cannot be true (is inconsistent) but because the content of *M* is inconsistent with my believing it. And thus to believe *M* is to violate a norm of rationality, to aim for exclusively true beliefs. The moral is that there is more to rationality norms concerning belief than mere consistency.

Returning to the case of credences, consider

C: $A \wedge cr(A) = \epsilon$

for some small ϵ. This has the following Moore-like property: C cannot be true if given high credence (in fact, any credence bigger than ϵ). For suppose $cr(C)$ is large. The credence in C of a rational agent is at most the credence in each conjunct; hence $cr(A)$ must be at least as large as $cr(C)$. But if C is true, then its second conjunct is so too: thus $cr(A) = \epsilon$, which is a contradiction. Just as in the classi-

[6] Williams (1994) gives an account along these lines.

doi: 10.1111/j.1467-9264.2011.00322.x

cal Moore case, it is irrational to give C high credence (although C could be true), because in doing so one puts oneself in a situation where something one gives high credence to must be false.

Now one who gives C high credence violates SR. For

$$cr(C) = cr(A \wedge cr(A) = \epsilon) = cr(cr(A) = \epsilon) \times cr(A \mid cr(A) = \epsilon).$$

If this is to be high, both factors in the last expression must be high, and if $cr(A \mid cr(A) = \epsilon)$ is high then SR is violated.

In these Moorean cases, it is crucial that the credence (or belief) functions are of the same person. No problems whatever arise from my giving C high credence when '*cr*' refers to someone else's credence function. Thus Christensen errs when he writes 'the epistemic respect we owe to ourselves is not ultimately different in kind from the epistemic respect we owe to others' (2007, p. 336).

Being in a Moorean state involves violating SR, or equivalently, obeying SR guarantees avoidance of Moorean irrationality. This alone does not *establish* SR as a rationality norm, because some principle weaker than SR may be enough to avoid the Moore's paradox cases. But these cases do show that at least *some* degree of insight into one's own beliefs is, contra Christensen, constitutive of rationality. The DBA further shows that, in fact, so is the strong degree of insight encapsulated in SR.[7]

Centre for Logic and Analytic Philosophy
Institute of Philosophy
University of Leuven
Kardinaal Mercierplein 2
3000 *Leuven*
Belgium
jacob.chandler@hiw.kuleuven.be

School of Humanities
University of Glasgow
Glasgow G12 8QQ
UK
adam.rieger@glasgow.ac.uk

[7] Jake Chandler was supported by a post-doctoral research grant from the Research Foundation, Flanders (FWO). Adam Rieger thanks Igor Douven and the Formal Epistemology Project at the Catholic University of Leuven for a Visiting Fellowship, and the AHRC for an award under the Research Leave scheme.

doi: 10.1111/j.1467-9264.2011.00322.x

References

Bovens, Luc 1995: '"P and I Will Believe that Not-P": Diachronic Constraints on Rational Belief'. *Mind*, 104, pp. 737–60.

Christensen, David 2007: 'Epistemic Self-Respect'. *Proceedings of the Aristotelian Society*, 107, pp. 319–37.

Milne, Peter 1991: 'A Dilemma for Subjective Bayesians—and How to Resolve It'. *Philosophical Studies*, 62, pp. 307–14.

Popper, Karl 1959: *The Logic of Scientific Discovery*, 2nd edn. New York: Basic Books.

Samet, Dov 1997: 'On the Triviality of High-Order Probabilistic Beliefs'. Game Theory and Information 9705001, EconWPA. http://ideas.repec.org/p/wpa/wuwpga/9705001.html.

Skyrms, Brian 1987: 'Dynamic Coherence and Probability Kinematics'. *Philosophy of Science*, 54(1), pp. 1–20.

Sobel, Jordan Howard 1987: 'Self-doubts and Dutch Strategies'. *Australasian Journal of Philosophy*, 65, pp. 56–81.

van Fraassen, Bas C. 1984: 'Belief and the Will'. *Journal of Philosophy*, 81, pp. 235–56.

Williams, John N. 1994: 'Moorean Absurdity and the Intentional "Structure" of Assertion'. *Analysis*, 54(3), pp. 160–6.

Proceedings of the Aristotelian Society, Vol. CXI, *Part* 3
doi: 10.1111/j.1467-9264.2011.00322.x