Title: **Value as Fully Rational Desire: an analysis of everything and nothing**

Author: **Y. S. Lo**

Value as Fully Rational Desire: an analysis of everything and nothing

Abstract: This paper demonstrates that Michael Smith's analysis of normative value as fully rational desire is a necessary but empty truth, telling us nothing particular about value or anything else.

Over the last two decades, many have challenged Michael Smith's now well-known rationalist analysis of (normative moral) value as fully rational desire. But until now, no one has been able to produce a direct knock-down argument against it – despite its apparent implausibility. Here I provide such an argument, demonstrating that Smith's analysis is empty, telling us nothing particular about value or anything else.

Briefly, Smith's analysis says:

S.   P is valuable if and only if for any agent X, if X is fully rational then X would desire P.[i]

P can be a proposition, an event, or a state of affairs. For Smith, P takes the form 'that agent Y performs action φ-ing in circumstances C'. But exactly what form P takes does not make any difference to my argument here.

The central claim of Smith's moral rationalism is that **S** is an analytic and so necessary truth. Obviously, whether that is the case depends on what the crucial term 'fully rational' means. According to Smith, to be fully rational is to be in the state of '*wide* Rawlsian reflective equilibrium'.[ii]   What does that mean? Rawlsian reflective equilibrium is usually understood in terms of *beliefs*. As such, it is a state in which an agent's beliefs are maximally coherent and unified among themselves. But Smith has a 'wide' notion of rationality,[iii] according to which, 'achieving reflective equilibrium may also be a goal in the formation of underived *desires*'[iv] and 'our desires will exhibit maximal coherence and unity when they are desires for the various things that we would believe valuable at the limit of rational inquiry'.[v]   In other words, to reason is not just to sort out one's beliefs, but – further – to sort out one's desires according to one's beliefs. We are fully rational just if both our beliefs and desires are fully rational: our beliefs are fully rational just if they are in a state of Rawlsian reflective equilibrium; our desires are fully rational (or, what Smith calls, 'maximally coherent and unified'[vi] or

'systematically justifiable'[vii]) just if they are desires for what we fully rationally believe to be valuable.[viii] Apart from holding a 'wide' conception of rationality, Smith also holds a 'non-relative conception', according to which 'it is part of what we mean by "rational" that fully rational people converge in their desires'.[ix] That is, not only is it the case that 'under conditions of full rationality we would all reason ourselves towards the same conclusions [i.e., beliefs] as regards what is to be done'; but it is also the case that 'via a process of systematic justification of our desires we could bring it about that we converge in the desires that we have'[x]. In other words, 'fully rational agents would all have the same desires'.[xi] Accordingly, Smith's conception of full rationality can be summarized as follows:

> Agent X is fully rational if and only if (1) X has all and only the beliefs that any agent would have in the state of Rawlsian reflective equilibrium (call them 'RRE-beliefs' for short) and (2) X has all and only the desires for things that X believes to be valuable.

Given Smith's conception of full rationality, his analysis of value, **S,** can be spelt out as follows:

S. P is valuable if and only if for any agent X, if (1) X has all and only RRE-beliefs and (2) X has all and only the desires for things that X believes to be valuable, then (3) X would desire P.

Now, according to Smith, the process of reaching a Rawlsian reflective equilibrium is a *truth tracking* process, in that the RRE-beliefs that one would acquire at the limit of this infinite process would be all relevant true beliefs.[xii] Hence, **S** can be further spelt out as follows:

S. P is valuable if and only if for any agent X, if (1) X has all and only relevant true beliefs and (2) X has all and only the desires for things that X believes to be valuable, then (3) X desires P.

For simplicity, let us use the term 'feature F' to stand for the feature of being valuable. Then, **S** can be rewritten as follows:

S. P has feature F *if and only if* for any agent X, if (1) X has all and only relevant true beliefs and (2) X has all and only the desires for things that X believes to have feature F, then (3) X desires P.

I am now going to demonstrate to what extent **S** is, as Smith tells us, an analytic and necessary truth.

First, suppose the left-hand-side of the biconditional statement, **S,** is true — i.e., P has feature F. It follows that it is a relevant true belief that P has feature F.[xiii] So, for any agent X, if (1) is true (i.e., X has all and only relevant true beliefs), then X believes that P has feature F. Now, if (2) is also true (i.e., X desires all and only the things that X believes to have feature F), then it further follows that (3) X desires P. In other words, for any agent X, if both (1) and (2) are true, then (3) is true. That is, we have logically arrived at the right-hand-side of **S.** To complete the proof for **S,** however, we need to go the other way too.

There are only two possibilities. Either there exist some agents who satisfy both conditions (1) and (2) or there exist no such agents. Let us consider the first possibility and see what follows from the right-hand side of **S.** For any agent X, (2) implies that if X desires P then X believes that P has feature F. Also, for any agent X, (1) implies that if X believes some statement then it is true. So, for any agent X, (1) and (2) together imply that if X desires P, then P has feature F. It follows that for any agent X, if X satisfies (1), (2) and (3), then P has feature F. Are there any such agents who satisfy all three conditions? Under the possibility that we are considering, there exist some agents who satisfy both (1) and (2). But, given the right-hand-side of **S,** those agents who satisfy both (1) and (2) also satisfy (3), which means that there are agents who satisfy all three conditions. So, it follows – as a matter of logic – that P has feature F. That is, we have logically arrived at the left-hand-side of **S.** This is arguably sufficient to show that the analysans (i.e., the right-hand-side) of **S** implies its analysandum (i.e., the left-hand-side).

But what about the alternative possibility, which we have not considered? That is the possibility that there exist no agents who satisfy both (1) and (2). In cases where that is so, the universal conditional (that for any X, if (1) and (2) then (3)), which constitutes the analysans of **S,** will be automatically true simply because the antecedent of the conditional is not satisfied. But this is where the problem lies. Presumably, no one putting forward a conceptual analysis where the analysans is a universal conditional would want to say that the analysandum is automatically true in cases where nothing happens to satisfy the antecedent of the conditional. In particular, Smith would presumably not want to say that everything is automatically valuable in situations where no one happens to be fully rational. Furthermore, a Smithian analysis of *dis*value would say something like this: P is

disvaluable if and only if for any agent X, if X satisfies conditions (1) and (2) then X desires the prevention of P.[xiv] But if we were to consider the possibility that no agent satisfies both (1) and (2), then, for reasons just given, the verdict of the analysis of disvalue would contradict that of the analysis of value. In short, given the logical structure of Smith's analysis, consideration of the possibility that nothing satisfies both (1) and (2) should be regarded as irrelevant to deciding the truth or falsity of the analysis.

Putting the point another way: Smith could quite reasonably restrict the range of possible worlds for his analysis to all and only those worlds in which both (1) and (2) are satisfied, and then say that when he claims that **S** is a necessary truth, he is speaking loosely; strictly speaking, what he means is that **S** is a necessary truth *relative to that restricted set of possible worlds* — i.e., it holds in all the worlds belonging to that restricted set. In other words, **S** has a restricted kind of necessity, which, I suppose Smith could argue, is good enough to sustain his central notion that **S** is an analytic and necessary truth.[xv] The clear fact that **S** is not logically necessary (i.e., it does not hold in all logically possible worlds) does not present a genuine problem for Smith. Or so at least he might argue.

Leaving restricted necessity aside, however, Smith's analysis now confronts a lethal problem. According to Smith, **S** is an analysis of *value*. As I have logically demonstrated above, **S** is indeed an analytic and necessary truth (relative to a restricted set of possible worlds). But no part of the demonstration actually depends upon what feature F really is or what concept it falls under. This means that feature F is arbitrary! The analyticity and necessity of **S** hold *regardless* of the identity of feature F. It can be the feature of being valuable, as it is originally supposed to be. But it can equally well be the feature of being approved of by Chairman Mao, the feature of being conducive to making money, the feature of being a fake Vermeer, the feature of helping to make life difficult for wives who have multiple admirers, or indeed *any* feature you like — **S** will still be analytically and necessarily true.[xvi] Whatever **S** might analyze, it certainly does not analyse any concept of value in any substantive way, let alone 'the concept of value that lies at the heart of morality'.[xvii]

NOTES

[i] See M. Smith, 'Dispositional Theories of Value', *Proceedings of Aristotelian Society, Supplementary Volume* 63 (1989): 89-111; M. Smith, *The Moral Problem* (Oxford: Blackwell, 1994); J. Bigelow and M. Smith, 'How not to be Muddled by a Meddlesome Muggletonian', *Australasian Journal of Philosophy* 75 (1997): 511-27.

[ii] Bigelow and Smith, 'How not to be Muddled by a Meddlesome Muggletonian', p. 523 (emphasis added).

[iii] Ibid. p. 526.

[iv] Smith, *The Moral Problem*, p. 159 (emphasis added).

[v] Bigelow and Smith, 'How not to be Muddled by a Meddlesome Muggletonian', p. 526.

[vi] Ibid. p. 523.

[vii] Ibid. and Smith, *The Moral Problem*, pp. 159-161.

[viii] Cf. Bigelow and Smith, 'How not to be Muddled by a Meddlesome Muggletonian', p. 523.

[ix] Smith, *The Moral Problem*, p. 166.

[x] Ibid. pp. 165-166.

[xi] Ibid. p. 175, Cf. p. 173.

[xii] Smith (Ibid. pp. 155-61) endorses an extended version of Bernard William's notion of full rationality. William's idea is that a fully rational agent has all relevant true beliefs and no false beliefs, and makes correct inferences, whereas Smith adds that a fully rational agent also desires just what she believes to be valuable.

[xiii] The question here is whether P has feature F. Any belief that is relevant to the question is a relevant belief. Since any belief is relevant to the question whether the belief is true, the belief that P has feature F is a relevant belief relevant here.

[xiv] This follows directly from **S** and the assumption that P has disvalue iff the prevention of P has value.

[xv] Also see note 17 below.

[xvi] Let F be any feature, p be any proposition, x be any agent, y be any thing, 'Dxy' stand for 'x desires y' and 'BxFy' stand for 'x believes that Fy'. The logical structure of **S** is this: $\forall F \forall p(Fp \leftrightarrow \forall x \forall y (((Dxy \leftrightarrow BxFy) \& (BxFy \leftrightarrow Fy)) \rightarrow Dxp))$. **S** holds in all possible worlds in which it is true that $\exists x \forall y((Dxy \leftrightarrow BxFy)\&(BxFy \leftrightarrow Fy))$. Those are worlds where omniscient F-lovers live. For different Fs, **S** has different restricted scopes of necessity. But there is no reason for thinking **S** is 'more necessary' (so to speak) in the case where F is the feature of being valuable than in the other cases.

[xvii] Bigelow and Smith, 'How not to be Muddled by a Meddlesome Muggletonian', p. 515.