

The Impossibility of a Bayesian Liberal?

Forthcoming, Journal of Politics

William Bosworth (LSE)

Brad Taylor (USQ)

7/8/2020

ABSTRACT

Aumann's theorem states that no individual should agree to disagree under a range of assumptions. Political liberalism appears to presuppose these assumptions with the idealized conditions of public reason. We argue Aumann's theorem demonstrates they nevertheless cannot be simultaneously held with what is arguably political liberalism's most central tenet. That is, the tenet of reasonable pluralism, which implies we can rationally agree to disagree over conceptions of the good. We finish by elaborating a way of relaxing one of the theorem's axioms that arguably lends itself to a coherent account of political liberalism, namely the condition of indexical independence.

The Impossibility of a Bayesian Liberal?

Liberalism has treated the need for an agreement to disagree as axiomatic from its very beginning. From John Locke's essays on religious toleration through to Rawls's idea of an overlapping consensus of comprehensive doctrines, liberals bracket disagreement about the good life (be it secular accounts of happiness, dignity, self-realisation, species-being, autonomy, greatness, or religious accounts of doing God's will) in the hope a reasonable agreement about the nature of social justice and other political values will be forthcoming. We agree to disagree on the good in order to work out the just terms of coexistence *given* the disagreement.

Political liberalism calls this reasonable pluralism (Rawls, 2005). In multicultural societies individuals with different views of the good life do in fact coexist. Political liberals put great stock in the underlying rules of this coexistence being reasonably acceptable. This requires elaborating a standard of 'reasonable acceptability' that, given pluralism, remains staunchly impartial between different comprehensive doctrines of morality providing they are not antithetical to living in a society with rules that everybody reasonably accept. As Rawls explains,

"Citizens realise they cannot reach agreement or even approach mutual understanding on the basis of their irreconcilable comprehensive doctrines. In view of this, they need to consider what kinds of reasons they may reasonably give one another when fundamental political questions are at stake... Central to the idea of public reason is that it neither criticizes nor attacks any comprehensive doctrine, religious or nonreligious, except insofar as that doctrine is incompatible with the essentials of public reason" (Rawls, 2005: 441).

The process of working out the right answer to these political questions is described as the process of public reason (Quong, 2011; Rawls, 2005: 440-91). The move to engage in this kind of public reason can be construed as an agreement to disagree on moral comprehensive doctrines.

This paper is a contribution towards analytically refining what we might consistently say about this agreement to disagree. Despite Robert Aumann's (1976) famous agreement theorem concluding that a kind of agreement to disagree is logically irrational for Bayesians, political liberals have been surprisingly silent on the theorem's implications. We suggest the theorem refines our understanding of political liberalism by identifying errors in a number of its elaborations. In particular, the theorem implies public reason's idealized constituency of "rational and reasonable" agents (Rawls, 2005: 51) is logically impossible for the more salient interpretations of the idealization.

Our strategy is to show the logical possibilities for avoiding Aumann's strong conclusion concerning agreeing to disagree over the good, if plausible, should also apply to public reason. That is to say, it is equally rational to agree to disagree on the content of public reason (on the nature of the original agreement to disagree) as on the nature of the good, on any disagreement over this agreement, and so on in an upwards regress. The assumptions of common priors, negligible communication costs, sincerity, and non-ambiguity are, if anything, less likely to hold for public reason than for moral deliberation.

The reason for agreeing to disagree over things like religion and metaphysics in our comprehensive doctrines is therefore undermined at the level of public reason, or at the level of what Brian Barry calls "second-order morality". There appears no immediate

way of relaxing the idealization to justify the move from first-order morality to second-order morality. Charles Larmore's (1987, see also Quong, 2011: 204-211) prescription for individuals to retreat to the common ground "to converse on the basis of the rest of their beliefs" in the face of disagreement over conceptions of the good, which "to a large extent... will remain unresolved", also turns out to be either too optimistic or too pessimistic. Thomas Nagel's (1987) idea of epistemological impartiality, where policy is justified from "a position from nowhere", underdetermines why we should not first use this "position from nowhere" to update our 'positions from somewhere'. If individuals reason from the impartial and plural grounds political liberalism assumes, Aumann's theorem suggests they would come to a rationally informed agreement about the nature of the good if it ever were a position to come to a rationally informed agreement about the content of public reason.

There is, however, one tacit condition required for Aumann's theorem to work that we think makes sense to relax at the level of moral deliberation about the good, *but not* at the level of public reason. This is the condition of indexical independence. If we assume claims about the good in fact include some kind of indexical information, it makes sense to bracket the disagreement and move up a level to public reason in line with political liberalism.

Aumann's Theorem

Aumann's (1976: 1236) argument is exceptionally simple,

THEOREM. If two people have the same priors, and their posteriors for an event E are common knowledge, then these posteriors are equal.

Two people, 1 and 2, are said to have *common knowledge* of an event E if both know it, 1 knows that 2 knows it, 2 knows that 2 knows it, 1 knows that 2 knows that 1 knows it, and so on.

Although it is entirely rational for Bayesian agents exposed to different information to independently come to different conclusions, this disagreement should not persist once it becomes common knowledge.¹ The intuition here is that if you believe another's stated belief to be a sincere and a competent reflection of the evidence they have been exposed to, then it provides a neat summary of that evidence. It is important to note here that it is the posterior judgement which provides the information, and there is no need for the underlying evidence to be communicated. If I believe you to be rational and sincere, the fact that you disagree with my judgement tells me that you must have been exposed to some compelling evidence, even when I have no idea what that evidence is.

In simple examples of the sort found in probability textbooks it is easy to see how this works. Suppose that you and I take turns privately drawing balls from an urn, which we are told contains some combination of black and white balls. If our task is to estimate the proportion of black balls, we should take the evidence of our own eyes into account. If upon drawing four balls you find that three of them are black, you ought to estimate the overall proportion of black balls, and the probability of drawing a black ball next, at 3/4.

¹ It should be noted Kelly (2019) argues the practical epistemic import of disagreement is much lower than suggested by Aumann's theorem and its extensions. Lederman (2015) also shows that it is possible to preserve the common priors assumption while allowing for disagreement among rational agents, and takes this to cast doubt on interpretations of Aumann's theorem which emphasises common priors as the only contested assumption. Although the fact of persistent disagreement in the real world may not represent an action-guiding revelation of irrationality, we argue below that Aumann's theorem has important implications at the level of idealisation appropriate to political liberalism.

This estimate reflects the evidence you have so far been exposed to, and as a Bayesian agent you will be willing to update this belief as new evidence arrives.

But implicit in the assumptions of Bayesian rationality and common knowledge is the idea that your eyes are no more reliable than mine in this respect. Suppose that after the initial drawing of four balls that prompted you to give an estimate of $3/4$, you and I publically share our estimates with one another. If I declare an estimate that $1/4$ of the balls are black, you can conclude that my drawing went quite differently to yours. If you know I'm a sincere and rational Bayesian you can conclude that on my drawing I saw one black ball and three white balls. You should not rationally give any extra weight to the evidence of your own eyes simply because they are stuck to your head, and so public pronouncement should be just as powerful as a second drawing of your own and you must rationally reduce your estimate to $1/2$. For the same reason, my initial estimate of $1/4$ will be updated to $1/2$ once I know that your initial estimate is $3/4$. In this simple example, agreement should be reached in one step. For more complicated cases, an equilibrium will still only be reached once disagreement has been resolved.² This reasoning process has also been generalized to non-probabilistic beliefs (Bacharach, 1985; Samet, 2010).

So the factual judgement of another person you know to be rational provides a powerful piece of evidence since it is based on their rational judgement of the evidence they have collected. Unless you have reason to believe that your evidence is better than theirs, you have no reason to privilege your own judgement over theirs. The point relevant to our argument is that disagreement is only possible in this abstract sense when Bayesian

² See Bonanno and Nehring (1997) for a more detailed review of this literature.

agents give their own beliefs – whether prior or posterior – priority over the beliefs of others. In other words, disagreement can only persist if there is epistemic partiality.

This is precisely the ideal of epistemological restraint political liberalism presumes of agents engaged in public reason. Nagel (1987: 230) describes it accordingly,

“The idea is that when we look at certain of our convictions from outside, however justified they may be from within, the appeal to their truth must be seen merely as an appeal to our beliefs, and should be treated as such unless those beliefs can be shown to be justifiable from a more impersonal standpoint.”

If liberals are already committed to this restraint Aumann’s theorem suggests there is no immediate reason to bracket their disagreement about the good and proceed to higher-order questions of the appropriate limits of toleration of posterior beliefs about the good. If they end up agreeing on the nature of the good, the question of political compromise dissolves away.

In *Political Liberalism*, however, Rawls suggests reasonable agents will accept there are significant burdens of judgment associated with agreeing on the good. This is what, for him, ultimately grounds reasonable pluralism and the move from a distinctly moral comprehensive doctrine to the freestanding political nature of public reason. These Rawlsian ‘burdens’ dominate the political liberalism literature, but what is striking is the number that are out of keeping with Aumann’s result. Rawls claims the complex and conflicting nature of scientific evidence, disagreement about the significance of evidence, and the path-dependency of belief formation all conspire to make broad agreement on the good impossible. Aumann’s theorem nevertheless suggests these burdens provide no reason for rational individuals to agree to disagree. If the burdens lead rational individuals to different prior beliefs about the good, the differences should be treated as

evidence to adjust posterior beliefs both ways until they are equivalent. There can, in other words, be no disagreement that is both rational and reasonable (*pace* Leland and van Wietmarschen, 2012; Peter, 2013).

While moral deliberation is seen as subject to the burdens of judgement, the constituency of public reason is typically taken as a simulation of a more or less idealized constituency of “rational and reasonable” agents (Rawls, 2005: 52; Quong, 2011: 143-4; Gaus, 2011: 276). Aumann’s theorem suggests these agents should therefore agree on the nature of the good. We could of course resist the view that the constituency of public reason will be Bayesian rational in the relevant sense by relaxing the precise level of idealization. But if we relax the Aumannian conditions for agents in public reason deliberating about the good, we also need to do it for the other areas of their deliberation. Doing so will extend the burdens of judgement to deliberation over justice as well and undermine the point of moving to public reason in the first place.

The only way to justify the political liberal’s agreement to disagree therefore is to *specifically* show the conditions Aumann’s theorem assumes hold asymmetrically between first-order moral deliberation and second-order political deliberation. That is, to show at least one condition for Bayesian rationality might hold for somewhat idealised political deliberation but not first-order morality. The conditions include

- (1) Individuals will have had common priors relevant to event E at some point.
- (2) There are no costs associated with communicating information and updating beliefs to establish the common knowledge associated with E.

(3) Posteriors are communicated sincerely, unambiguously, and relatively precisely.

(4) The content of the agreement is truth-theoretic.

(5) The information communicated does not contain indexical content.

These conditions largely overlap with the Rawlsian burdens of judgement. For instance, Rawls (2005: 57) suggests the “significant complexity” of moral deliberation justifies divergent judgement in light of the disparate nature of “citizens’ total experiences”. Aumann assumes over such a possibility with the condition of common priors (1) and (2) costless (non-complex) communication. Rawls (2005: 56-7, c and e) also suggests the vague and ambiguous nature of moral decisions will rationally justify divergence. Such a possibility is excluded with condition (3).

Insofar as convergence to some form of agreement (even if it is only a convergence on a meta-consensus) is assumed to be the end-point of public reason, there must be a condition that holds for public reason but not for deliberation about the nature of the good. We will now sketch five possible ‘outs’ that roughly capture the logical space of Aumann’s theorem. We suggest all but one are implausible. Only one of the conditions in which it is rational for Bayesian agents to accept an agreement to disagree can be plausibly said to hold for moral and not political deliberation. That is, there is only one condition that it is both plausible to relax and renders the ideal constituency of public reason of rational and reasonable agents logically possible.

(1) Common priors

The most contested assumption of Aumann's result is condition (1), that of common priors. Aumann (1998) sees (1) as formalising the constraint on rationality that differences in belief must ultimately derive from differences in information. Two people exposed to exactly the same information should have the same beliefs. In reality differences in upbringing and culture will mean this is not the case, as Rawls (2005: 56-7, d) points out, but in this section we suggest such differences cannot be justified from an ideally rational point of view. If society or nature randomly allocates personality traits, which alter the way we process information, there is no reason to systematically favour our own priors over others.

As far as we can see, all forms of public reason assume the same common priors condition as Aumann at some point. These priors are sometimes described as common sense, as with Rawls (2005: 441), "We are to appeal only to presently accepted general beliefs and forms of reasoning found in common sense, and the methods and conclusions of science when these are not controversial." Common priors are necessary for such appeals. Larmore (1987: 54) argues that in the face of disagreement over conceptions of the good agents should retreat to a common ground and continue "to converse on the basis of the rest of their beliefs". The idea is that public reason will require common priors to identify this common ground.

Is there any justification for an asymmetry here with first-order moral deliberation? It may appear plausible that we will never have common priors sufficient to establish a common posterior on the nature of morality. Quong (2011: 204-212), for instance, argues there is a common justificatory framework to resolve disagreements over the

subject matter of public reason, whereas there is no such framework for foundational disputes over the good. A Catholic and an atheist supposedly have no common framework to appeal to (nothing the other would treat as good evidence) to resolve a disagreement over whether or not it is, say, true that faith is good. If this turned out true and there were indeed no common prior beliefs to work from, it would quite plausibly justify the liberal agreement to disagree over the good.

Yet the common prior assumption is more wide-ranging than is often presumed such that it applies to first-order morality as much as second-order public reason. Hanson (2006) demonstrates that among rational individuals “uncommon priors require origin disputes.” Rational individuals who agree enough about the origins of their priors will mean that they should agree enough for identical posteriors. He suggests,

“Beliefs about the origins of priors also seem relevant to the rationality of priors. For example, if you learned that your strong conviction that fleas sing was the result of an experiment, which physically adjusted people’s brains to give them odd beliefs, you might well think it irrational to retain that belief.” (Hanson, 2006: 320)

It is a violation of what Hanson calls pre-rationality to allow one’s prior to depend on such contingent circumstances that cannot be said to be rigged towards tracking the truth. If two rational individuals disagree due to uncommon priors, Aumann’s argument and the Bayesian imperative of epistemic impartiality is simply pushed back a level. If rational individuals agree enough about the origins of their priors, then they will agree enough for identical posteriors.

It is reasonable to expect the common ground assumed by political liberals to be sufficient to agree on these origins. Hanson argues, “If priors are transmitted culturally via children copying visible adults, standard theories about individual variations in such culturally transmitted belief tendencies offer little support for the idea that some children are better able to select the most truth-tracking cultural elements from among the available cultural transmissions” (Hanson, 2006: 326). This might be said to justify some of the claims made by genealogists of morality. As a rule, these genealogies are decidedly non-liberal. While Nietzsche’s *Genealogy of Morals* is open to interpretation, he roughly suggests that realizing the historical origin of egalitarianism is a capitulation to slave-like resentment gives credence to prior notions of Aristocratic and elitist virtue. While Nietzsche’s history is sketchy at best, this is the kind of intellectual strategy Hanson’s result may suggest. Instead of bracketing different conceptions of the good like slave and aristocratic morality, we should come to an agreement concerning their origins and then, through this process, update our priors about the good. Bayesians do this by treating comprehensive doctrines that arise in more favourable truth-tracking conditions as better kinds of evidence than those that do not.

So it is difficult to argue uncommon priors can rationally justify an agreement to disagree about the nature of the good. Rather than move up a level to second-order morality, we should move back with the constraint of pre-rationality to discover the historical basis for differences in prior belief. Rational Bayesians should be able to work out historical grounds for updating their belief towards unanimity in light of these differences. The conceptions of the good held in communities conducive to truth tracking should be given more weight than those that are not and Bayesian agents should update until posteriors are equivalent. A group’s conception of the good would be counted as better

evidence, for example, if it arose historically in a culture where deliberation and free expression flourished than in societies conducive to the Stockholm Syndrome.

(2) Communication and computation costs

It is clearly impractical to imagine that we will be able to convince everybody of our own conception of the good. Even if we were to assume there is in fact a rational way to convince others of our view of the good life, it seems *prima facie* sensible to suggest the costs of communication for the information-sharing necessary to trigger Aumann's strong conclusion will be too great. The difficulty of satisfying condition (2) for moral deliberation suggests (2) would be the salient condition to justify political liberalism if it was somehow easier to do so for public reason.

Rawlsians, for instance, draw a contrast between the complexity of conceptions of the good whose scope ranges over a large range of states of affairs (amounting to a comprehensive doctrine) and the relative simplicity of public reason. The scope of public reason is often construed to be narrower, restricted to constitutional essentials or what Rawls calls "the basic structure" of society (Rawls, 2005: 11, 13, 257-88). It is taken to be 'coarser-grained' (List and Valentini, 2020) and easier to communicate compared with the finer granularity and greater complexity of moral reasoning. This amounts to a burden of judgement (Rawls, 2005: 56-7, d) for moral deliberation.

From a Bayesian point of view though it is important to keep in mind exactly what needs to be communicated in order for updating to occur. Although the reasoning upon which conceptions of the good are based is highly complex, communication of a discrete belief

about the good is much simpler. The preference utilitarian, for example, need not explain to the religious fundamentalist exactly *why* they hold this belief about the good but simply that they *do* hold it. Bayesian updating requires the communication only of posterior judgments, not priors or the evidence on which such judgements are based.

If the preference utilitarian updated their belief in light of the religious fundamentalist's belief that following the will of God is good, however, this would be a radical concession. It would require updating priors over a whole host of other comprehensive beliefs due to the complexity and granularity of conceptions of the good. These new updates would then ideally be communicated back to the fundamentalist and one might think the updating and communication process will need to spread over their entire comprehensive doctrines. This is to be contrasted with Rawls's (2005: 375) "freestanding" question of the content of the overlapping consensus. Updating here will plausibly be far more contained and may therefore seem significantly less costly.

Scott Aaronson (2005), however, has extended Aumann's result in a way that suggests the difference in communication and computational costs will be relatively minor for ideal Bayesian agents. He proves the number of messages agents will need to communicate in order to approximate agreement for expectations over a random variable will be relatively small regardless of complexity.³ Of course, this can be done by packing so much – so many granules or 'bits' – into each message that they become too complex to compute. As Aaronson (2005: 635) puts it,

"Independently of communication cost, it seemed obvious that computation cost would provide another barrier to agreement. For just to calculate her opinions,

³ Specifically, $1/(pe^2)$ messages will suffice to have agents' assessment agree within ϵ with probability at least $1 - p$ over their prior (where ϵ and p are parameters).

Alice might need to take expectations over sets of [possible states of the world] of size $2^{\Omega(n)}$ that are constantly being updated in response to messages from Bob.”

Yet his result also blocks this line of reasoning. Aaronson (2005: 640) proves the same number of messages also suffice for what he calls “summary messages” that sufficiently summarise the expectations of the agents with only 2 bits. He suggests this complexity reduction is crucial for drawing “a ‘real-world’ moral from Aumann’s theorem” (Aaronson, 2005: 635).

We should stress though that Aaronson’s results are heavily idealised and concern expectations on a quantified variable. Political liberals may resist, first, by suggesting beliefs about the good are not quantified in this way. They may be fundamentally qualitative, for example. But if they are the burden is on political liberals to explain why the results should not hold qualitatively when they do so clearly quantitatively. Others might point to the extremely idealised nature of the results, suggesting they would not hold in non-ideal settings. But the burden is again on them to show what about the non-ideal environment would mitigate the results. If it is the insincerity, ambiguity or vagueness associated with communicating conceptions of the good, which we will turn to now, we have little reason not to expect exactly the same for public reason.

(3) Sincerity, Ambiguity, and Vagueness

We group the assumptions of sincerity, ambiguity, and vagueness (3) into a single section since they are, if anything, more troubling for agreement-seekers in public reason than in moral deliberation in a way that is perhaps starker than (1) and (2). With public reason, individuals have a set conception of the good. They actively have an interest in vetoing

policies that cut against it and rejecting vetoes on policies they endorse. With these incentives, their sincerity is easily compromised. A partisan is likely to be more motivated to be insincere in order to elicit consent during the process of public reason than moral deliberation. When it is taken for granted that partisans are dealing with people who do not agree with their view of the good, it is unclear what incentives there are for avoiding strategic communication.

Differing conceptions of the good will likely mean we attach different senses to our moral terminology. Different religious points of view usually associate different extensions to the term 'justice', for instance. When we engage in moral deliberation we often try and reason individuals from their perspective to our own. There are incentives to engage them unambiguously on their own terms. For public reason there is a contest over how policy claims are framed. Terms like 'freedom', 'equality', and 'justice' are sometimes described as essentially contested (Gallie, 1953), meaning individuals with different viewpoints will associate different extensions with the terms. Given public reason takes plural viewpoints for granted, there is no incentive for individuals to engage their audience on their audience's terms and we can therefore expect more ambiguity than less. The goal is often merely consent, even if that consent is non-substantive. We might think of Tony Blair's tactics during the Good Friday Agreement in Northern Ireland as a case in point. Unionists agreed to allow Sinn Fein to be eligible for parliament on the condition they 'gave up violence', in Blair's wording. It turned out that the Unionists thought this meant that Sinn Fein had agreed to decommission their weapons, which was not the case. Blair was nevertheless successful in eliciting consent. Ambiguity was useful for the kind of agreement sought in this instance as it is useful for the kind of agreement aimed at in public reason.

Ambiguity is one of Rawls' (2005: 57, e) burdens of judgement. The vagueness of moral deliberation is another. There are a number of influential theories that suggest the vagueness of terms like 'rich', 'bald', and 'tall' will be inevitable (e.g. Williamson, 1994). Whether a hard case counts as 'rich' or not will therefore be epistemically indeterminate. No amount of Bayesian updating will overcome divergence of judgement for vague predicates and could justify an agreement to disagree. Yet, again, while it is arguably plausible to suggest public reason will trade in vague predicates like 'illegal', 'person', and 'freedom' (Dowding and Bosworth, 2018), it is difficult to say the same about predicates like 'is good' or 'is bad'. While we may disagree over hard cases for vague predicates, like whether Trump's 2016 Travel Ban was illegal or whether a fetus constitutes a person, there will be large-scale agreement on the vast majority of cases. Indeed, we can usually reduce the discussion to a single metric like dollars in a bank account, hairs on the head, or height in centimeters. No such reduction or consensus exists with rival conceptions of the good. There are certain unsavory groups who believe the death of infidels a good thing where the infidels believe the spread of their own religion the only true good. This is not disagreement over hard cases; it is disagreement over virtually all cases.

(4) Agreement is truth-theoretic

Some proponents of political liberalism do not see the agreement they espouse as an agreement of belief, but as something more like the agreement involved with consent or even strategic coordination (see Weitham, 2010; Thrasher and Vallier, 2015; 2018; Kogelmann and Stich, 2016). Public reason then would appear *not* to satisfy (4) in aiming at truth-theoretic agreement. While the ideal constituency of public reason will be both rational and reasonable, the standard of reasonableness is not treated as equivalent to rationality (Rawls, 2005: 48-54). The condition of rationality and the

rational updating of belief may only be applicable to moral rather than public reason in this sense. Insofar as public reason concerns reasonableness it could be said to bypass the constraint of truth-tracking beliefs. Such interpretations of public reason put pressure on the idea that political liberalism is aiming at truth-theoretic agreement. This would be an obvious asymmetry with first-order moral deliberation, thereby rendering the agreement to disagree on the good rational.

Many of Rawls's claims appear consistent with this reading. The starkest evidence for this is that he says agreement through the process of public reason will be insensitive to the concept of truth. He proposes, "that in public reason comprehensive doctrines of truth or right be replaced by an idea of the politically reasonable addressed to citizens as citizens" (Rawls, 2005: 441) and makes it more perspicuous in terms of his so-called political constructivism, saying it

"does not use (or deny) the concept of truth; nor does it question that concept, nor could it say that the concept of truth and its idea of the reasonable are the same. Rather, within itself, the political conception does without the concept of truth" (Rawls, 2005: 127)

This suggests public reason is not truth-seeking. It is not a pooling of information to approximate a true representation of public right, say; rather, it is purely procedural or strategic. This undermines the cut of Aumann's theorem because, as stipulated with (4), the process of Bayesian updating is a process that updates beliefs about what is *true*. It is not clear the conclusion of Aumann's theorem holds even with this weaker claim given individuals engaged in public reason will not be motivated solely by the veracity of their beliefs.

Cohen (2009), however, argues that public reason will still track some form of truth, what he calls the “political conception of truth”. He argues,

“The reason that the concept of truth has no place cannot be that the claims made by a political conception are not truth-apt; those claims must be truth-apt, even if the political conception itself somehow abjures the concept of truth. They must be, if there is to be a common ground of argument under conditions of doctrinal disagreement.” (Cohen, 2009: 17)

The goal of public reason is to update this common ground by refining it until there is consensus on a theory of justice. This common ground must take the theory to be *true*. Even if this were all false and public reason somehow eschewed all need for agreement on what is true, it would all be for nothing if we could not then agree on the result of the procedure. That is to say, we need to work towards equivalent posteriors concerning the content of the theory of justice that parties deemed reasonable. So we will need rational agreement at some stage of the public reasoning process, and it is this agreement that is privileged over agreement about the good life. Even if we concede that public reasoning is not truth-sensitive, we still require truth-theoretic agreement over the *content* of the agreement to disagree (be that content what is consented to or what is strategically signaled). It is this latter truth-theoretic agreement that must be thought to satisfy the Aumannian conditions better than moral deliberation over the good.

An alternative option for establishing an asymmetry regarding (4) is to deny moral claims about the good are truth-theoretic. This might on the face of it look to endorse a ‘boo-hurrah’ theory or close non-cognitive analogue, where moral discourse ultimately boils down to what A.J. Ayer described as ‘emotive ejaculation’. But even logical positivists usually do not deny that individuals *think* their moral view is true. It comprises something they believe they know about the world. Even if it turned out moral deliberators could

only justify thinking X is good by appealing to their preferences, they still would claim that their view that ‘X is good or valuable’ is true. There is nothing in Aumann’s theorem that guarantees what is agreed to *has to be* true. Nor even does it guarantee that the moral opinion is substantive. All it guarantees is that ideal Bayesians will ultimately have equivalent posteriors concerning whether or not a moral assertion is true if those posteriors are common knowledge.

(5) Indexical Independence

So asymmetries between moral deliberation and public reason on conditions (1)-(4) fail to justify the agreement to disagree that underpins political liberalism. We nevertheless believe there is a way to preserve the spirit of political liberalism while remaining faithful to Bayesian rationality. It is rational to agree to disagree over views of the good if such views are taken to contain indexical information. That is, relaxing (5), the condition of indexical independence. This is a similar move to (4) but is importantly different in that it preserves the truth-theoretic nature of moral deliberation.

Indexicals are concepts like ‘I’, ‘me’, ‘here’, and ‘now’ that are dependent on speaker-context for their content. They have a linguistic character (Kaplan, 1989), which is often called their meaning, but that character only determines the semantic content for each token as a function from context. We all grasp what individuals mean by ‘the good’ – its linguistic character – even when *we* associate different contents with the concept (see Dreier, 1990; 2009; Khoo and Knobe, 2018; and Bosworth, 2020 for a close analogue). Insofar as we agree not to treat the individuals’ use of ‘good’ as an error, ‘the good’ is structurally identical to more conventional indexicals.

The statement ‘I am living in the United Kingdom’ is informative. Were somebody not to initially believe it and then hear my utterance and shared the relevant priors, they would update their posterior belief about where I live if they believed I was sincere. Yet they would not update the belief that *they* live in the United Kingdom, despite it being the content of *their* belief ‘I am living in the United Kingdom’. They do not form a belief about *their* location from *my* indexical statement. This is a special feature of indexical information. For my audience to use my assertion as evidence towards the thesis that *they* live in the United Kingdom would be self-evidently absurd. So, we may conclude, Aumann’s theorem does not apply to indexical information.⁴

It may seem like we are making too much of the condition of indexical independence. While there may well be indexical statements, it is perhaps plausible that the information contained in them can be reduced to regular run-of-the-mill propositions where it cannot be rationally agreed to disagree over. It is perhaps easy to believe the information that is communicated by ‘I am living in the United Kingdom’ is just the same as the information communicated by ‘Will Bosworth is living in the United Kingdom’. This, however, is not the case. Indexical propositions are *essentially* indexical. Removing the ‘I’, the ‘here’, the ‘this’, etc. from propositions lobotomizes information that cannot be captured with non-indexical terms. This to say that there is information content that is captured by indexicals that cannot be described independent of explicit indexicality.

⁴ The formal idea in terms of possible worlds semantics is that the content of essentially indexical beliefs is irreducibly first-dimensional (see Bosworth 2020: 298-303, 308-9) whereas the content of beliefs in Aumann’s proof must be expressible only in terms of second-dimensional content. As Aumann (1976: 1236-7) says, “Included in the full description of a state w of the world is the manner in which information is imparted to the two persons. This implies that the information sets $P_1(w)$ and $P_2(w)$ are indeed defined unambiguously as functions of w , and that these functions are known to both players.” Aumann assumes here possible “contexts of utterance” are drawn from the “circumstances of evaluation”, to use Kaplan’s (1989) phrasing. This collapses the two-dimensional distinction required to capture the semantics of indexicals.

John Perry's famous thought experiment in his paper 'The Problem of the Essential Indexical' (1979) captures the intuition behind this claim. Imagine you are going about your shopping, strolling down the aisles of a supermarket with a trolley. You suddenly come across a trail of sugar from – you surmise – an unfortunate shopper who unbeknownst to them has a hole in their sugar bag. You follow that trail to try and warn them. After a while, the sugar doubles up. There are now two trails. The shopper is returning to where they were before. You quicken your pace. Soon there is another fresh trail, and then another, and another, and eventually it dawns on you: you are the shopper with the hole in their sugar bag. This information is essentially and explicitly indexical. One might claim that you can re-describe that information by substituting the indexical here for your proper name. Say, 'I am Will Bosworth' implies the information conveyed by 'I am the shopper with the hole in my bag' is also conveyed with 'Will Bosworth is the shopper with the hole in their bag'. But this is not quite right: say I have amnesia. I might not be aware I am Will Bosworth. The assertion 'I am Will Bosworth' would be informative to me, but not 'Will Bosworth is Will Bosworth'. The assertion 'Will Bosworth is the shopper with the hole in their bag' would certainly be informative, but not quite as informative as 'I am the shopper with a hole in their bag'.

The justificatory process for this kind of first-person knowledge is semantically the same as the justificatory process for knowledge about the good. Take the claim 'It is good to avoid recessions'. The way to tease out the non-indexical content is to simply ask, 'why?' The appropriate response here is a fact that links recessions to a more basic good. So we might say recessions increase unemployment and unemployment is bad. But what is wrong with unemployment? We could raise the fact that unemployment decreases the likelihood an individual will report satisfaction with their life. But this is only a salient fact because of the underlying normative idea that it is good for individuals to feel

satisfied with their life. But why is it? This regress continues until it bottoms out with a basic claim, something like ‘Satisfaction with one’s life is just good’. ‘But why?’ a persistent interlocutor might ask. At some point in the regress the only honest response is, ‘It just is!’ Similarly, a persistent interrogator of my claim that I am Perry’s shopper might return, ‘Okay, you have shown that Will Bosworth is the shopper with a hole in their bag and Will Bosworth is the current speaker, but how do you know you are the speaker?’ At some point ‘I just know I am!’ is the only recourse to such insistent questioning, as it was with ‘Satisfaction with one’s life is just good simpliciter’. One of the only ways to accommodate this, short of discrediting all first-person knowledge and knowledge of the good as baseless, is to treat the end of both justificatory chains as essentially indexical content.

We propose this is enough to suggest liberals can treat a claim like ‘Pain is bad’ as an indexical statement. The information ‘Pain is bad’ is information over which it would be an *error* to update one’s belief in light of becoming aware that other Bayesians believe it. This is true even when we share common priors. No amount of historical analysis of our origins (Hanson, 2006) will change it. Something resembling this line of thought might be gleaned from Brian Barry in *Justice as Impartiality*. It is worth quoting a fragment from his book at length for it is instructive as much for what it gets wrong as for what it gets right,

According to Nagel, if you make a belief of yours a basis for public policy, ‘it must be possible to present to others the basis of your own beliefs, so that once you have done so, *they have what you have*, and can arrive at a judgment on the same basis.’ It is this requirement that Nagel asserts is violated by convictions whose ‘source... is personal faith or revelation’. But if somebody believes my report of a certain experience, he has what I have, in the only sense that is relevant. Suppose I have a pain and describe it to you with some precision (intense, throbbing, and located six inches below and to the right of the navel, say). Obviously, I

still have the pain and you don't acquire it. In that sense, then, you do not have what I have. But if the issue is what causes this pain, then you do have what I have. And if you are a doctor while I am not, your view about the cause will probably be better founded than mine.

I wish to argue that there is a precise analogy between what can be said about a pain and what can be said about a religious revelation. Suppose I hear a voice in my head which claims to be that of God, or some other religious personage. If I report its content faithfully to you, then you have what I have in the relevant sense. We can both form views about the probable validity of this experience as a genuine religious revelation. In principle, there is no reason why I should be more impressed by experience than you by my report of it, so long as you do not believe that I am deliberately deceiving you. It is a piece of information whose significance we both have to evaluate, and we may legitimately arrive at different conclusions about it. The phenomenon is perfectly general: faced with the same form books or the same meteorological data, two people can reasonably make different predictions about which horse will win or what the weather will be. This is what we mean by calling something a matter of judgement. The one thing we should agree about is that nobody is entitled to claim certainty. (Barry, 1995: 180-1)

Barry calls this view 'scepticism'. The label of 'scepticism', as far as we can see, is intended to capture the last two sentences, which license his basic claim that "no conception of the good can justifiably be held with a degree of certainty that warrants its imposition on those who reject it" (Barry, 1996: 165). This is strange given his analogy to the patient in pain: it appears to suggest the patent falsehood that the patient can never be certain that they are in pain. Aumann's theorem suggests the last two sentences are plainly mistaken. It is irrational to have conflicting judgments when presented with the same meteorological data and form guides (of course, individuals may bet differently depending on their levels of risk-aversion).

We can nevertheless save the argument in the rest of the fragment by taking claims about the good to be indexical. A better way to think of the analogy between pain and revelation is with the thought experiments associated with phenomenal experience like Frank Jackson's (1986) Mary in the room. Mary is a brilliant scientist living in a black

and white room from birth a couple of centuries from now. The physical sciences are now complete and Mary has memorized every last physical facts about colour, e.g. that different colours amount to light at different wavelengths. She is also a good Bayesian and can communicate with people outside the room. She learns that she is in a black and white room, that there are more colours than those that exist in her room, and has updated her priors in light of reports from Bayesian agents living on the outside about what the experience of these colours is like. One day the door of Mary's room opens, she leaves, and sees the colour red for the first time. Despite her prodigious – bordering on omniscient – knowledge, upon seeing red she learns something new. She is acquainted with the colour red for the first time, like people are supposedly acquainted with direct revelation from God, or with the knowledge that pleasure is a value that cannot be reduced to a calculus of pain, and so on. No amount of non-indexical information prepared her for this the new information, 'So *this* is redness'. Mary can be near certain of the veracity of her new knowledge about colour, as can the patient about their knowledge of their pain, as can individuals of the value of happiness and freedom. The idea is that no procedure designed to establish common knowledge could move rational individuals on their *basic* conceptions of the good as much as they could not move Mary on her experience of redness.

Reasonable Rejection and Idealization

This is the key to Barry's interpretation of political liberalism (see also Scanlon, 1982; Dowding, 2011). He advocates justice as impartiality, where a constitution is just if and only if individuals motivated by a desire for agreement cannot reasonably reject it from reasons grounded by *their* fundamental and subjective conceptions of the good. What is

reasonable and what is not is determined in the court of public reason. Political liberalism then is not based on assumptions justified by one or another view of the good, but rather on a desire to freely and equally agree on how to live given the content of our respective references to the good are not equivalent (Barry, 1995: 164-8).⁵ This is quite explicitly consistent with a relaxation of (5). So long as we reject Barry's label of scepticism and take indexicality seriously, his approach amounts to a coherent way of preserving liberal beliefs in the shadow of Aumann's theorem.

If we are to take the relaxation of (5) seriously though, we have to set the right level of idealization for the constituency of public reason. Barry's (1995: 195-207) so-called 'a priori' approach to justice as impartiality appears, as things stand, the closest to getting the idealization mix right. The a priori approach involves philosophically scrutinizing the justifications from an ideal point of view for potential vetoes grounded by the basic conceptions of the good actually held by non-ideal members of the public.

Barry (1995: 104, 207-13) counterbalances this a priori approach with what he calls the empirical approach. Yet the empirical approach gets the idealization mix wrong by being excessively non-ideal. It looks to the policies and principles endorsed by countries that are open to free and equal deliberation (e.g. the Netherlands and Scandinavia – Barry, 1995: 106, 108) as an approximation of the content of justice as impartiality. The idea is that citizens will have had the chance to both publicize and work through all their potentially reasonable grievances. Many deliberative democrats endorse the rough approach (Cohen, 1989; see also Weale, 2013). Yet if we relax (5), individuals can only be counted within the constituency of public reason if their comprehensive doctrines

⁵ The desire to freely and equally agree on the terms of cooperation is inclusive enough to gain a realistically large convergence on the good of implementing the outcomes of public reason (Barry, 1995: 164-8; Scanlon, 1982: 116-7; see also Rawls, 1999: 517).

hold basic views of the good to be indexical. This is plausibly the case only for a very small minority of individuals. It is unlikely, for example, somebody holding a Christian conception of the good would deem a Muslim believing that faith in Allah is good to be anything other than mistaken. Yet to justify an agreement to disagree the Christian would need to accept that the Muslim's fundamental belief is, from the Muslim's perspective, not itself mistaken. While only a minority of individuals may believe conceptions of the good are indexical in this sense, we can still simulate public reason amongst an ideal constituency of agents all of whom do. It would be an error then for political liberalism to endorse outsourcing the process of reasonable rejection to deliberative mechanisms. It would lend itself to a constituency that is too inclusionary for agents to rationally accept an agreement to disagree over the good.

Gerald Gaus' justificatory liberalism (1996; 1999) also appears to roughly accommodate the result but is likewise a shade too non-ideal. Principles of free speech and liberal toleration are victoriously justified, Gaus argues, in the sense that individuals cannot consistently infer reasons to "defeat" (i.e. reject) the principle from within their own belief system. This works in the abstract, but, like Barry, some of Gaus' epistemological assumptions require adjustment. Gaus' (1999: 38-62) relativism over reasons, for instance, appears to commit him to the idea we should agree to disagree not just over basic views of the good, but also over what those views imply. While relaxing (5) renders it rational to agree to disagree over grounding views of the good like whether the absence of pain is fundamentally good, it is still irrational to agree to disagree over factual inferences like whether unemployment causes pain. To suggest the truth of the latter should be relative to an individual's belief system would be to implausibly suggest the assignment of different truth-values to 'I am Will Bosworth' rationalizes the assignment of different truth-values to 'Will Bosworth is the shopper with a hole in their bag'. While

we may agree to disagree over the essential indexical ‘pain is bad’, we should not over conditionals like ‘if pain is bad, unemployment is bad too’. The constituency of public reason should be further idealized then for agents to reach agreement on the truth-value of factual claims like ‘employment causes pain’.

While the empirical approach to justice as impartiality and justificatory liberalism are a little too non-ideal to avoid the cut of Aumann’s theorem, there are a number of interpretations of political liberalism that are far too ideal. The consensus approach to public reason, for example, idealizes the constituency to such a degree that agents engaged in public reason all reason alike (see Quong, 2011). Agents in the constituency of public reason, by this account, reason in a way that does not appeal to actual differences over the good. Aumann’s theorem nevertheless shows this to be irrational. Ideal agents should either agree on a conception of the good or relax (5). Where (5) is relaxed and there is a realistic amount of variation between basic views of the good, agents will have quite varied reasonable vetoes of social rules. These vetoes will therefore have varied justifications that other agents with rival conceptions of the good would not themselves make. So while it seems irrational for the constituency of public reason to agree to disagree over which social rules *are* reasonably rejected, relaxing (5) makes it quite rational to agree to disagree over which social rules *it would be good to* reasonably reject. Which social rules are reasonably rejected will be determined by the multifarious non-ideal conceptions of the good when they are simulated in the (somewhat) idealized process of public reason. The social rules it is good to reject will be relative to each agent’s basic conception of the good.

Conclusion

Indeed, with perfectly rational and reasonable agents committed to reaching a fair compromise among competing views we would expect to see complete agreement on policy as well as the basic structure (Taylor and Bosworth, 2020). Bayesian updating with complete epistemic impartiality would remove any disagreement on questions of fact (including the factual question of the moral commitments of each party to public reason). This would preclude disagreement on which policies and institutions are justified by public reason while allowing differences in judgments about which policies and institutions would be good as judged by the individual themselves.

If Bayesian agents in public reason have purged all their factual and logical differences, however, it may be objected the conceptions of the good that continue to divide them will not look anything like the comprehensive doctrines that make up contemporary civic discourse. The view that following God's will is fundamentally good, for instance, presumably assumes that God exists. If an atheist and theist are Bayesian they should not agree to disagree on the point of God's existence. Does this imply that either all religious or all secular comprehensive doctrines should be excluded from the simulation of public reason? We do not think so. There is nothing irrational about an individual revising their understanding of their *basic* conception of the good *post hoc* upon updating their beliefs about its implications. Most religious doctrines, for example, tend to stress that *faith* in God is also a fundamental good. This faith demands representation in the court of public reason as much as any fundamental religious claim.

Somebody raised on J.S. Mill's *On Liberty*, for another example, might believe autonomy is always good and justify their belief by the supposed fact that autonomy increases happiness. But say agents in public reason rationally agree that autonomy does not

always do so. It is impossible to say the Millian is mistaken then to react and update their essential indexical belief to *also* include autonomy as a discrete basic good side by side with happiness. They might then supplement this with further essentially indexical beliefs concerning whether various trade-offs between autonomy and happiness are fundamentally good or not.

This may appear intellectually arbitrary. It is, however, just to pass the intellectual buck to public reason. Political liberals should not make strong discriminations between reasonable and unreasonable basic conceptions of the good because they must accept these conceptions will be indexical and therefore impossible to determine whether mistaken or not. But this is far from relativism: there is much that can be said to prove it unreasonable to use faith in God to veto, say, extending the franchise to atheists in public reason. Representatives of this view would not accept identical reasoning from rival conceptions of the good to exclude theists, so cannot be said to provide an acceptable justification.

References

Aaronson, S. (2005) 'The complexity of agreement', *Proceedings of the Thirty-Seventh Annual ACM Symposium on Theory of Computing*, ACM, pp. 634–643.

Aumann, R.J. (1976) 'Agreeing to Disagree', *The Annals of Statistics*, 4: 6, pp. 1236-1239.

Aumann, R. J. (1998) 'Common priors: A reply to Gul', *Econometrica*, 66: 4, pp. 929–938.

Bacharach, M. (1985) 'Some extensions of a claim of Aumann in an axiomatic model of knowledge', *Journal of Economic Theory*, 37: 1, pp. 167–190.

Barry, B. (1995) *Justice as Impartiality*, Oxford: Clarendon Press.

Bonanno, G., & Nehring, K. (1997) 'Agreeing to disagree: A survey', *Document Prepared of an Invited Lecture at the Workshop on Bounded Rationality and Economic Modelling. Theory*, forthcoming.

Bosworth, W. (2020) 'An interpretation of political argument', *European Journal of Political Theory*, 19: 3, pp. 293-313.

Cohen, J. (1989) 'Democratic Equality', *Ethics*, 99, pp. 727-51.

Cohen, J. (2009) 'Truth and Public Reason' *Philosophy and Public Affairs*, 37: 1, pp. 2-42.

Connolly, W.E. (1974) *Terms of Political Discourse*, Princeton, NJ: Princeton University Press

Dowding, K. (2013) 'The Role of Political Argument in Justice as Impartiality', *Political Studies*, 61: 1, pp. 67-81.

Dowding, K. and Bosworth, W. (2018) 'Ambiguity and vagueness in political terminology: On coding and referential imprecision', *European Journal of Political Theory*, forthcoming.

Dreier, J. (1990) 'Internalism and Speaker Relativism', *Ethics* 101: 1, pp. 6–26

Dreier, J. (2009) 'Relativism (and expressivism) and the problem of disagreement', *Philosophical Perspectives*, 23, pp. 79-110.

Gallie, W.B. (1956) 'Essentially Contested Concepts', *Proceedings of the Aristotelian Society*, 56, pp. 167-98.

Gaus, G. (1996) *Justificatory Liberalism: An Essay on Epistemology and Political Theory*, New York: Oxford University Press.

Gaus, G. (2006) "The Rational, the Reasonable and Justification", *Journal of Political Philosophy*, 3 (3): 234-258.

Hanson, R. (2006) 'Uncommon priors require origin disputes', *Theory and Decision*, 61: 4, pp. 319–328.

Jackson, F. (1986) 'What Mary didn't know', *Journal of Philosophy*, 83: 5, pp. 291-295.

Kaplan, D. (1989) 'Demonstratives,' in *Themes from Kaplan*, ed. J. Almog, John Perry, and H. Wettstein (Oxford: Oxford University Press).

Kelly, T. (2019) 'The Epistemic Significance of Disagreement', in: Fantl, J., McGrath, M., Sosa, E. (Eds.), *Contemporary Epistemology: An Anthology*. John Wiley & Sons, Ltd, pp. 249–264.

Khoo, J. and Knobe, J. (2018) 'Moral Disagreement and Moral Semantics', *Nous*, 52: 1, 109-143.

Kogelmann, B. and Stich, S. (2016) 'When Public Reason Fails Us: Convergence Discourse as Blood Oath', *American Political Science Review* 110 (4): 717-30.

Larmore, C. (1987) *Patterns of Moral Complexity*, Cambridge: Cambridge University Press.

Lederman, H. (2015). 'People with Common Priors can Agree to Disagree', *The Review of Symbolic Logic*, 8(1), 11-45.

Leland, R.J. and van Wietmarschen, H. (2012) 'Reasonableness, Intellectual Modesty, and Reciprocity in Political Justification', *Ethics*, 122: 4, pp. 721-747

List, C. and Valentini, L. (2020) 'What Normative Facts Should Political Theory Be About? Philosophy of Science meets Political Liberalism' in *Oxford Studies in Political Philosophy*, ed. D. Sobel, S. Wall, and P. Vallentyne. (New York: Oxford University Press).

Nagel, T. (1987) 'Moral Conflict and Political Legitimacy', *Philosophy and Public Affairs*, 16, pp. 215-40.

Nagel, T. (1991) *Equality and Partiality*, Oxford: Oxford University Press.

Perry J (1979) The problem of the essential indexical. *Nous*, 13, pp. 3–21.

Peter, F. (2013) 'Epistemic Foundations of Political Liberalism', *Journal of Moral Philosophy* 10 (5): 598-620.

Quong, J. (2011) *Liberalism Without Perfection*, Oxford: Oxford University Press.

Rawls, J. (2005) [1993] *Political Liberalism*, New York: Columbia University Press.

Samet, D. (2010) 'Agreeing to disagree: The non-probabilistic case', *Games and Economic Behavior*, 69: 1, pp. 169–174.

Scanlon, T. (1982) 'Contractualism and Utilitarianism', in Amartya Sen and Bernard Williams (eds.) *Utilitarianism and Beyond*, Cambridge: Cambridge University Press.

Taylor, B. and Bosworth, W. (2020) Agreeing to Disagree Politically, *New Political Economy*, forthcoming.

Thrasher, J. and Vallier, K. (2015) 'The Fragility of Consensus', *European Journal of Philosophy* 23 (4): 933-54.

Thrasher, J. and Vallier, K. (2018) 'Political Stability in the Open Society', *American Journal of Political Science*, 62 (2): 398-409.

Weale, A. (2013) *Democratic Justice and the Social Contract*, Oxford: Oxford University Press.

Weithman, P. (2010) *Why Political Liberalism?: On John Rawls's Political Turn*. Oxford: Oxford University Press.

Williamson, T. (1994) *Vagueness*, London: Routledge.