

Constructivist Contractualism and Future Generations

Gustaf Arrhenius & Emil Andersson

Gustaf.Arrhenius@iffs.se

Emil.Andersson@filosofi.uu.se

Abstract

In constructivist contractualist theories, such as Rawls's, principles of justice should mirror beliefs that we all, in some sense, share. One would then arrive at principles that everybody could, in that sense, accept. However, what to count as beliefs that we all share depends on how we delimit the set of people that make up the "we". Thus, for constructivist contractualism, the question of whom to assign a part in the justification procedure and whom to exclude, and how to justify these inclusions and exclusions, are of crucial importance.

Keywords: contractualism, future generations, constructivism, intergenerational justice, John Rawls, Bruce Ackerman.

I

A principle of distributive justice has to tell us, among other things, to whom to distribute the relevant benefits and burdens, the *recipients* of the principle, and to whom to assign responsibility for the distribution, the *agents* of the principle.

The determination of the recipients and agents of a distributive principle is the classical assignment problem in moral and political philosophy. Many long-discussed questions arise from this problem: the status of foetuses, foreigners, future generations, and animals are examples of the problem of delimiting the domain of recipients; children, intellectually disabled, and ill-informed people are examples of the problem of delimiting the domain of responsible agents.

Assume that the way to justify principles of justice is that we all, in a metaphorical sense, sit down and discuss the principles. The principles that we all agree upon are then the justified principles. An advantage of such an approach is that it is metaphysically non-committal and non-mystical. The principles don't have their origin in "an independent moral order", nor are they "required by natural law, or by a realm of values known by rational intuition" Rawls (2005), p. 22). Instead, the correct principles of justice are, as John Rawls put it, "conceived as agreed to by those engaged in" social cooperation (Rawls (2005), p. 23).

The relevant kind of agreement must, of course, be understood as an agreement under appropriate conditions. We have to make sure that everyone is treated fairly, are given an equal say in the decision, and so on. Still, there is one fundamental problem that must be solved before we can start: Who is eligible to take part in the meeting, who are "we all"? Contemporary people in a well-ordered liberal state? All the people on earth? All the human beings that will ever exist? This is what we shall call the *new assignment problem*.¹

The answer that we give to the new assignment problem will determine whose beliefs and opinions which will be taken into consideration. Only the beliefs of those that are included in the relevant "we" will count, and hence our decision on this issue will fundamentally shape both the hypothetical meeting and its result. In what follows we shall discuss how this problem arises for one influential version of contractualism and the implications of different answers to the new assignment problem.

II

Roughly, a contractualist political theory can be divided into three levels. The first level consists of the substantive political principles, the second of the justification model or procedure that justifies the substantive principles. The third level consist of the reason why we should choose a certain justification procedure, the justification of the justification model so to speak.²

One way to construct a justification model is to start out from some non-contingent, indisputable, non-moral features of the human predicament. The aim is to justify a moral and political order by appeal to something that nobody can reject, irrespective of one's beliefs and values. Examples of such attempts are Hobbes's absolute sovereign chosen by self-interested rational agents in a pre-social situation and Gauthier's moral code "generated as a rational

¹ There is a structural similarity between this problem and the recently much discussed boundary problem in democratic theory. For the latter, see e.g., Arrhenius (2005), (2018); Goodin (2007); Miller (2009).

² This way of structuring a political theory does not, of course, rule out the possibility of a dynamic relation between the levels, like in Rawls's idea of a reflective equilibrium. See Rawls (1971); Tersman (1993).

constraint from the non-moral premises of rational choice” (Hobbes (1973) and Gauthier (1986), p. 4).³

Theories of the type above are sometimes called “foundationalist” (Ripstein (1987), p. 116). This term has been used rather loosely among political theorists and their usage is certainly different from that of epistemologists. For the purpose of this paper, we shall take “foundationalism” to mean the view that there is a set of facts that holds true for all different cultures and societies, and normative principles can only be justified by appeal to such facts.⁴

One might also further restrict the first clause above to only cover non-moral facts. This would be appropriate for capturing Hobbes’s and Gauthier’s theories but it is arguably not necessary. One could claim that there are some moral facts that are true in every society. Scanlon’s contractualism can be seen in this light. Although Scanlon refers to “informed, unforced, general agreement” in his canonical statement of the theory, the main work is done by the idea that one can find principles “that no one could *reasonably* reject”, where “reasonably” is a moralized notion that supposedly is the same in all societies (Scanlon (1998), p. 153, our emphasis. See also pp. 192-194).

One of the consequences of the foundationalist way of reasoning is that it does not matter who takes part in the justification procedure. Whatever people we assign to this procedure, the resulting normative principle would be the same because the result is not, or so the argument goes, affected by contingent features like culture, conceptions of the good and so on. Indeed, it is enough that one (arbitrarily chosen) individual deliberates. Hence, no new assignment problem on the justificatory level will occur for these theories.⁵

There have been attempts in political philosophy to move away from the foundationalist method of constructing a justification model. Instead of starting out from non-contingent premises, these theories focus on beliefs that we all implicitly hold, in some sense. The most prominent anti-foundationalist political theory of this kind is Rawls’s later political interpretation of his contractualist theory. Another prominent example is Bruce Ackerman’s dialogic theory. Roughly, the building blocks of the justification model in these theories are justified insofar as they mirror implicit beliefs that we all, in some sense, share (Rawls), or if different sets of beliefs of different groups converge to a consensus on these matters

³ Cf. Ripstein (1987), p. 116 fn. 1.

⁴ The domain of “all different cultures and societies” is somewhat imprecise but it will suffice for the present purpose.

⁵ One might, however, argue that there is a kind of assignment involved in the specification of the justification procedure. Hobbes and Gauthier specify a rationality criterion, and Scanlon a reasonability ditto, which excludes certain beings from having an impact on the justification procedure. This is not the assignment problem that we shall focus on in this paper, but we shall briefly comment on it in section IV below. Another issue for the foundationalist is that it might be possible to draw different conclusions from the set of non-contingent premises, a kind of underdetermination of the data so to speak. We shall not pursue this issue further here.

(Ackerman). Henceforth, we shall refer to anti-foundationalist theories of Rawls's and Ackerman's kind as *constructivist* contractualist theories, or constructivist contractualism for short.

Two clarificatory remarks regarding our understanding of constructivism are in order. *First*, we assume that all versions of constructivism conceive of the correctness (or reasonableness, as the case may be in Rawls's later writings) of the resulting normative principles as being in some sense *constituted* by the fact that they are the result of the justification model. The role of the justification model is thus not epistemic; it is not supposed to assist us in identifying some independently valid principles, and to justify us believing in these principles. *Second*, we consider constructivism to be anti-foundationalist in virtue of its reliance on the implicit beliefs in some particular society. It is possible to construe constructivism differently, by viewing the "materials of construction" – the fundamental ideas that justify the justification model – as having a different status. For instance, it might be claimed that these fundamental ideas are true in the realist sense, with the result being a kind of foundationalist constructivism.⁶ Though such a view is possible, it does not qualify as a form of constructivism in the sense that we are concerned with in this paper.

Although the theories of Rawls and Ackerman differ in many aspects, the justification model of both theories can be thematised as a conjunction of three separate issues: constraints on what information we can make use of in the justification procedure, i.e., information constraints; delimitation of the topic of justice, i.e., the objects of the justification procedure; and criteria delimiting who is granted a part in the justification procedure, i.e., the agents of the justification procedure. Examples of the first are Rawls's veil of ignorance and Ackerman's principle of Neutrality. Objects under discussion are basic institutions for Rawls and more or less everything for Ackerman. The assignment of agents in the model and its consequences for the justification of the resulting principles is one of the topics of this paper. Roughly, Rawls's subjects are contemporaneous rational and reasonable people in a well-ordered society; Ackerman's subjects are members of a society that can raise a claim.

In constructivist contractualist theories, the justification model can change depending on whose "implicit beliefs" that are taken into account or who have beliefs that make it possible to reach a consensus upon certain issues. The diverse beliefs of different people can affect what "we" can consent upon, or what to count as an implicit belief. Thus, for Rawls's and Ackerman's theories, the questions of whom to assign a part in the justification procedure and of whom to exclude, and how to justify these inclusions and exclusions, are of crucial importance.

⁶ Scanlon appears to conceive of his contractualism as being constructivist in this way (Scanlon (2014), pp. 96–98). For a general discussion on this topic, see Street (2008).

There are at least two important groups that are excluded from Rawls's and Ackerman's convention: non-citizens and future people. These people are excluded in two ways: One does not take their beliefs into account in the construction of the justification model and these people are not directly represented in the model.

We shall concentrate on the latter group in this paper, not because that is a more important group than the former but rather because all the problems that arise with non-citizens, with one exception, also arise with future people, but not vice versa.⁷ We shall use future generations as a probe to uncover a general problem for constructivist contractualism.

We shall assume that future people will not differ radically from us when it comes to rationality and mental capacity. Other beings are excluded from the justification procedure just on these grounds: children, intellectually disabled people, and nonhuman beings. There are interesting philosophical problems hidden in such exclusions, but we shall not consider them here.⁸

III

Rawls's original position is intended to represent people as free and equal beings. The resulting principles of justice are those which people "would consent to as equals when none are known to be advantaged by social and natural contingencies". The original position is an "expository device" which "sums up the meaning" of our notions of fairness and "helps us to extract their consequences" (Rawls (1971), pp. 19–21).⁹ The rationale for the original position is that it embodies an adequate account of equality and that it generates fair principles.

The original position involves a set of information constraints on the parties: they do not know their place in society, their class position, social status, fortune, natural assets, abilities, intelligence, strength and the like. Nor does anyone know her conception of the good, particulars of her rational plan of life or special features of her psychology such as her aversion to risk. The parties do not know the particular circumstances of their society like the political and economic situation or the level of civilisation and culture. Nor do they know what generation they belong to (Rawls (1971), pp. 136–137).

⁷ The exception is the regulation of cross-border migration (although if time travel became possible, this difference would cease to hold).

⁸ For an in-depth discussion of both of these issues, see Nussbaum (2007). Interestingly, Rawls does not exclude children and infants from his justification procedure, since he holds that everyone with the capacity (developed or not) for moral personality are to be included (Rawls (1971), p. 509). Ackerman, on the other hand, explicitly excludes children since they lack the skills necessary to participate in neutral dialogue (Ackerman (1980), pp. 96, 140). Though there are potential problems with both approaches, we shall not address them here.

⁹ Cf. Kymlicka (1990), pp. 62–63.

How are these constraints justified? The fundamental idea behind Rawls's political interpretation of justice as fairness is that it is a *political conception*, not a metaphysical or an epistemological conception, as he puts it (Rawls (2005), pp. 10, 97). A political conception of justice is, according to Rawls, independent of controversial comprehensive doctrines of the good; applies to the basic structure of society, the "ground rules" of the public domain; is or can be shared by citizens regarded as free and equal; draws solely upon ideas implicit in the public political culture and the public tradition of their interpretation (Rawls (2005), pp. 11–15).

The two last clauses explain the information constraints in the original position. In short, the information constraints to be imposed on the convening people are justified insofar as they mirror the implicit beliefs that these people have about what it is to view people as free and equal persons in a liberal democratic regime:¹⁰

... since justice as fairness is intended as a political conception of justice for a democratic society, it draws *solely* upon the basic intuitive ideas that are embedded in the political institutions of a constitutional democratic regime and the public tradition of their interpretation. Justice as fairness is a political conception in part because it starts from within a certain political tradition. (Rawls (1985), p. 225)

There is, however, a tension in Rawls's veil when it comes to information about what generation one belongs to. On the one hand, the parties do not know the generation to which they belong. On the other hand, Rawls uses what he calls the "present time of entry interpretation", that is, the parties do know that they are contemporaries and existing people and in that sense know to which generation they belong.¹¹

How does Rawls make these two considerations compatible? The answer is that he assumes that not knowing what stage of civilisation one's society has achieved amounts to the same thing as not knowing what generation one belongs to (or, for that matter, not knowing what society one belongs to) (Rawls (1971), p. 287). But, as Rawls himself acknowledges, this equivalence does not hold all the way (Rawls (1971), p. 140). If the parties know that they are the present people, then there is nothing that prevents them from dismissing any kind of duty to posterity. The parties cannot affect what previous generations have done, so why not just adopt a principle that no one has a duty to future generations?

¹⁰This also holds for information that is allowed to penetrate the veil, like the five ideas of the good: goodness as rationality, primary goods, permissible comprehensive ideas of the good, political virtues, and the good of a well-ordered society. Rawls explicates these ideas of the good in Rawls (1988). Other building blocks of the original position are listed in Rawls (1971), pp. 146–147.

¹¹Rawls (1971), p. 140. As far as we know, Rawls never presents any clear-cut reason for this interpretation. He seems to think that the only alternatives to this interpretation are to take all actual (people who exist or will exist) or all possible persons into account in the contract and that would be to "stretch fantasy too far" (Rawls (1971), pp. 139, 146).

Rawls provides two different accounts of how to avoid such an outcome. In *Theory*, he adds a motivational assumption:

What is essential is that each person in the original position should *care about the well-being* of some of those in the next generation . . . Moreover for anyone in the next generation, there is someone who cares about him in the present generation. Thus the interests of all are looked after and, given the veil of ignorance, the whole strand is tied together. (Rawls (1971), pp. 128–129, our emphasis)

While contemporary people are represented as free and equal in the justification model by means of constraints on information, future people are represented by letting information penetrate the veil: the parties know that they care for somebody in the next generation. Present people are taken into consideration as free and equal because the parties are not allowed to know what kind of persons they are; future people are taken into account because the parties know that they will have a benevolent interest in the well-being of at least one person in the next generation.

Could we respect future people as free and equal persons by having a sufficiently benevolent interest in their well-being? Consider Richard Hare's "ideal sympathiser". This person treats all people's interests as if they were her own interests. Hare asks us to internalise the viewpoint of each person affected by our acts and treat all viewpoints, including our own as equally important, or, as Hare puts it, we should treat all viewpoints as our own.¹² So perhaps the original position could represent future people as free and equal persons by means of Rawls's motivational assumption by introducing another heuristic device inside the original position, a restrained version of the ideal sympathiser: the parties have to be ideally sympathetic to at least one person in the next generation.

Being risk-averse, the persons behind the veil of ignorance would then choose to maximise the position of the worst off in both generations and implement a savings plan to fulfil this goal.¹³ This savings plan would then limit the scope of the difference principle between generations and "no representative man in any generation of the most disadvantaged can complain of another for not doing his part" (Rawls (1971), p. 293).

There are, however, several problems with this solution. Firstly, the way to maximise the position of the worst off in the next generation could very well be to exploit so many non-renewable resources that the worst off in the next-next generation would have to suffer, and so

¹² Note that Hare is not promoting the ideal sympathiser as a heuristic device and he is not making use of any benevolence assumptions. Rather, he argues that this procedure follows from the logic of imperatives and a person's commitment to be moral. In Hare (1984) he suggests, however, that the ideal sympathiser is the best way to interpret the idea of equal consideration.

¹³ Cf. Rawls (1971), pp. 129, 292.

forth. How could Rawls avoid this conclusion? He could simply assume that we are ideal sympathisers with one person in every forthcoming generation. Already in its original formulation, however, Rawls's motivational assumption is a stretch, and one might worry that the above revision would "stretch fantasy too far" (Rawls (1971), pp. 139, 146. Cf. fn. 11).

Nevertheless, Rawls probably had another solution in mind. Recall the clause in the motivational assumption which said that "given the veil of ignorance, the whole strand is tied together". Assume that we have three persons that belong to three subsequent generations, for example a grandfather, a father and a daughter. The grandfather would now take his granddaughter into account not because he cares for her as much as he cares for himself, but because he cares for his son as much as he cares for himself and he knows that his son cares for the daughter as much as the son cares for himself. This transitive chain of sympathy can then be generalised over all future generations. Rawls hints towards this solution when he says that "[t]he parties are regarded as representing family lines, say, with ties of sentiment between successive generations" (Rawls (1971), p. 292).

The addition of this kind of motivational assumption has been criticised in the literature. Jane English, in an early criticism, noticed that Rawls's own arguments against assuming benevolence in the original position are applicable in this case as well (English (1976), pp. 92–93). Against characterizing the parties as being moved by benevolence, Rawls argues that such a motivational assumption is not only unnecessary (as mutual disinterest in combination with the veil of ignorance achieves the same result), but also in conflict with grounding the theory on weak assumptions.

Benevolence would also bring with it additional complexity, as we would have to characterize the nature and strength of such desires in great detail. Hence, as Rawls suggests, reasons of simplicity also speak in favour of avoiding such a characterization of the parties (Rawls (1971), pp. 148–149). If Rawls is right about this, then these considerations would speak against his proposed solution to the problem of future generations as well.

Another line of criticism has been advanced by Brian Barry, who claimed that this proposed modification of the motivation of the parties is objectionably ad hoc; a modification solely justified by the need to solve the problem of future generations but lacking any independent rationale (Barry (1977), p. 279).

Here we would like to point out yet another problem with this solution, relating to the constructivist method of appealing to our implicitly shared beliefs. Notice first that taking future people into account by means of the motivational assumption is not to take them into account as free and equal persons. Rather, they are taken into account instrumentally. The parties don't care about future people directly as free and equal persons but instead because these people might, through a transitive chain of interactions, affect the well-being of somebody whose well-

being they care about as free and equal persons. To care for certain people as a means for other ends is, of course, not to care for them as free and equal beings, as ends in themselves.

Since the representatives in the original position take present and future people into account for different reasons we can now ask: Is this distinction between contemporary and future people an implicit or latent belief in our political culture?

There is a park outside our window. Let us say that we decide to plant a bomb there, and this bomb kills several children. We all think that this act is wrong and unjust but we can have different reasons for thinking so. For example, those with consequentialist inclinations might think the act is wrong because it diminishes people's well-being whereas those with deontological inclinations, such as Rawls, might think it is wrong because it violates people's rights. Now consider the case in which we plant a time-bomb in the park. The bomb explodes some generations later and several children are killed. Why would this act be wrong? Whatever reasons people invoked to condemn the former act above, we believe most people would invoke the same reasons again. They hold the plausible Same Reasons View:

The Same Reasons View: The fundamental reasons which determine an act's normative status do not change because the act affects future people rather than present people, other things being equal.¹⁴

If we were to accept Rawls's first account of justice between generations, we would have to reject this view. Though the rights of future people are violated in the second case as well, we would have to say that the fundamental reason that grounds the rights of these future people differs from what grounds the rights of present people: the rights of present people are grounded in their status as free and equal, while the rights of future people are grounded in their being instrumental in negatively affecting the well-being of someone whose well-being the parties in the original position cares about. But taking future people into account in this instrumental way is at odds with the widely shared belief in the Same Reasons View and is therefore incompatible with the constructivist method of justifying the justification model by an appeal to our implicitly shared beliefs.

¹⁴ The *ceteris paribus* clause in the formulation is meant to imply that the acts have the same consequences for people, for example killing or harming, performed with the same intentions, and affects people to whom the agent stands in the same relevant relations. Notice, for example, that we don't intend this principle to rule out the possibility of giving more weight to the interests of people to whom we have special relations as compared to strangers. According to common-sense morality we are morally permitted, or obliged, to give some kind of priority to our own interests, to our families, to our friends, to our patients, to fellow-citizens, and so forth. However, if we include such considerations in our moral theory, the other things wouldn't be equal when an act affects a person to whom we have such a special relation rather than a stranger. This would not change the fact that we would invoke the same fundamental reason when our act affects people, present or future, to whom we stand in no special relation.

Ackerman's solution to the problem of future generations also comes into conflict with the Same Reasons View. Just as Rawls's original position, Ackerman's justificatory device – the neutral dialogue – is framed by a number of different constraints such as a rationality, consistency, and neutrality constraint (Ackerman (1980), pp. 4, 7, 11, 35, 35, 38, 40, 41). Why should you as a citizen accept these constraints? Ackerman's idea is that we can reach a common consent on these constraints even among people who disagree about most other things. According to Ackerman, there are different paths to accept these constraints: scepticism, experimentalism, autonomy, and so forth. Groups with different fundamental ideas about justification or values can find their own arguments that support Ackerman's constraints on the dialogue:

... I do not imagine that I am defending an embattled citadel on the fringe of modern civilization. Instead, I am pointing to a place well within the cultural interior that can be reached by countless of pathways of arguments coming from very different directions. (Ackerman (1980), p. 12)

Ackerman is thus an anti-foundationalist in that he does not aspire to ground his theory in non-contingent features of the human situation. He argues that the constraints imposed on the people in the dialogue are only justified insofar as they are part of a common ground that we all can accept, an "overlapping consensus" to use Rawls's terminology.

A necessary condition for citizenship in Ackerman's society is dialogic performance: "a citizen is (by definition) someone who can properly claim the right to be treated as a fellow member of the political community" (Ackerman (1980), p. 74). It follows that a person has to exist to be part of the justificatory dialogue, so how can his theory take future people into consideration? Ackerman has a promising suggestion: We have to be prepared to answer the claims they are going to raise when they have achieved citizenship. Thus, a "citizenry devoted to an ongoing liberal conversation must take steps *now* to assure that . . . delegitimizing dialogue do not take place later" (Ackerman (1980), p. 113, my emphasis. Cf. p. 111). This obligation would in a neutral dialogue yield, according to Ackerman, that the present generation has to assure that the next generation "will obtain a quantity of manna [the distributive currency in his theory] *no less than that guaranteed to members of the present generation*" (Ackerman (1980), p. 113).

As we noted above, the best way to assure that the next generation is well-off can very well have as a side-effect that people in the further future will be worse off. We will never enter into a dialogue with these people because when they exist, we do not, and vice versa. Consequently, if we were to deplete resources at the expense of people in the further future, then we do not need to worry about any dialogic embarrassment because no dialogue will take

place with these people. Hence, Ackerman needs to add some extra premise to explain why we have any obligations to people in the further future.

A possible solution, analogous to Rawls's amendment of his theory described above, is to claim that we have not left enough manna to the next generation if they cannot both consume the same amount of manna as we started out with and fulfil their dialogic duties to the subsequent generation. Let us say that we started out with one grain of manna each and we leave 100 000 grains to the next generation. Assume further that the two next generations will consist of 75 000 people.¹⁵ The next generation will consume one grain each and can only reproduce 25 000 grains, that is, there will only be 50 000 grains left for the next-next generation. If we could foresee that the next generation only could reproduce 25 000 grains, then they could complain to us that we have not guaranteed them equally as much manna as we started out with, because they cannot both consume the same amount of manna as we did and avoid dialogic embarrassment with the generation succeeding them. Consequently, if we want to avoid delegitimizing dialogue with the next generation, we have to take steps now to assure that the next generation both can consume the same amount of manna as we started out with and avoid delegitimizing dialogue with the generation succeeding them. This can be generalised to every subsequent generation.

There is a structural similarity between this solution and Rawls's motivational assumption. On Rawls's view the rights of people in next-next generation are grounded in their being instrumental in negatively affecting the well-being of someone whose well-being the parties in the original position care about. Similarly, on Ackerman's view the rights of people in the next-next generation are grounded in their being instrumental in affecting our ability to avoid delegitimizing dialogue with the next generation. Both methods generate a solution that more or less amounts to the same as taking people in the further future directly into account, without actually doing so. Both methods negate the Same Reasons View, and both methods are therefore in conflict with the constructivist method of appealing to our implicitly shared beliefs.

IV

In his later work, Rawls proposes a different solution regarding how to take future generations into account in the justification model that seems to respect the Same Reasons View. Rather than changing the motivation of the parties, he proposes that they are to choose a principle of just savings that they would want all previous generations to have followed. The idea being that

¹⁵ So we are here setting aside the great challenge of how Ackerman's theory could deal with changes in the population size. See, e.g., Arrhenius (forthcoming), (2000).

whereas the parties know that they are contemporaries, they do not know their place among generations. The correct principle of just savings is thus, on this view, the principle that the members of *any* generation would adopt as the one that they and all other generations are to follow (Rawls (2001), p. 160, (2005), p. 274).

It is far from clear whether this solution is satisfactory. For instance, it has been argued by Steve Gardiner that this way of solving the problem runs into problems once we consider cases involving the very first generation (which has no generations preceding it) as well as cases where the choices made by earlier generations determines how many future generations there will be (Gardiner (2009), pp. 110–114).¹⁶

Here, we shall once again consider the viability of this solution in the light of the constructivist method of appealing to the shared beliefs in current democratic societies. Even if this solution is an improvement over the earlier one, there are problems emanating from the reliance on the beliefs of current people when dealing with future generations.

A tension in Rawls's theory with regard to future people appears when we consider the justification of the justification model. When the parties in the original position take future people's interests into account they will use the conception of primary goods that reflects the implicit beliefs of the present generation. Since it is the beliefs of these presently existing persons that count when we are to determine the account of primary goods, the parties will not have access to the notion of primary goods that will be endorsed by future generations.

As a political conception, primary goods are grounded in the ideas implicit in the public political culture and the public traditions of interpreting these ideas. That these ideas are implicit in our political culture does not guarantee, however, that this will be the case in the future. Arguably, the conception of primary goods that one can extract from the public political culture two hundred years ago is rather different from what one can extract today. Moreover, the public traditions of interpreting these ideas have changed during this century and are likely to continue to change.

At any rate, there seem to be only two possible positions here: Either one holds that the ideas implicit in the public political culture change over time or one holds that they are constant. One natural way of reaching the latter conclusion is to claim that there are interests that all people share irrespective of their cultural belonging. But in that case, constructivist contractualism would collapse into foundationalist contractualist theory: we would justify our theory by appealing to facts that hold true for every possible society (more on this below).

Are there any problems for Rawls with the former view, that the ideas implicit in the public political culture change over time? Well, it seems to run into problems with another

¹⁶ For a different take on the viability of this solution, see Attas (2009).

cherished notion of political liberalism that Rawls thinks follows from the implicit beliefs in contemporary political culture. An important trait of political liberalism of Rawls's kind is to try to be *neutral*, in some sense of the word, between competing claims about what is valuable in a life. Provided that a modern society contains a plurality of conceptions of the good, there is a need for an account of the needs of citizens that is suitably independent of these conceptions, and thereby neutral in regard to them. Political liberalism thus "seeks common ground – or if one prefers, neutral ground – given the fact of pluralism" (Rawls (2005), pp. 180, 192).

How can such common ground be uncovered? Rawls hopes to do so by "working from fundamental intuitive ideas implicit in the public political culture" (Rawls (2005), p. 192). His account of primary goods is thus suitably neutral by being based on the conception of free and equal persons that he assumes is widely shared in a democratic society. The primary goods – basic rights, liberties and opportunities, income and wealth, and the social bases of self-respect – is an account of things that everyone need "as free and equal persons", and can therefore be accepted by everyone who shares this idea of persons (Rawls (2005), p. 180). Neutrality in regard to competing claims regarding what is valuable in life is thus secured by constructing the account of primary goods on the basis of shared beliefs.

An account of primary goods that is not in this way based on common ground would presumably be classified as non-neutral. Consequently, to use one's own account of primary goods without it being shared in the requisite way would be a breach of neutrality when acting politically. But if implicit ideas change over time, this is precisely what the present generation will be guilty of doing once they make use of their own conception of primary goods when taking the interest of future generations into account. Hence, Rawls's "common ground" neutrality is not neutral between generations.

Interestingly, the problem that we have here identified is not directly rooted in who are assigned as agents, or represented agents, in the justification model. It may plausibly be claimed that by excluding information about which generation the parties belong to and having them select a principle of just savings that all generations are to follow, the parties are in effect acting as representatives for all generations. But even if future generations are being represented in such a way, there is a problem related to the assignment of the persons whose implicit beliefs are taken into account in the justification of the justification model. Though it may be thought that these will coincide on constructivist views, we here see that this need not be the case. Even if future generations are represented in the justification model, the model itself is constructed from the beliefs of present people.

As we said, the new assignment problem may often coincide with another assignment problem that may affect foundationalist contractualism too: who should be assigned as agents in the justification model? For example, in Gauthier's bargaining model, one has to consider

how to represent future generations in the bargaining process.¹⁷ But foundationalists are not committed to justifying their justification model by reference to the implicit beliefs of some prior “we”, so they need not worry about the new assignment problem.

Since this new assignment problem arises as a consequence of the constructivist reliance on presently shared beliefs, it is not surprising that it arises in the case of Ackerman’s theory as well. As previously mentioned, the distributive currency in Ackerman’s theory is not primary goods, but what he refers to as “manna”. Firstly, how do we decide what counts as manna and how to amalgamate different goods into this overarching good of manna? Ackerman assumes that his notion of “manna” can model any good whatsoever. This involves two controversial assumptions: a) that we have an uncontroversial notion of what counts as a resource, “a shared understanding of the range over which the notions ‘good’ and ‘goods’ can move”, as Flathman puts it (Flathman (1983), p. 360); b) that all goods are commensurable and can be amalgamated into one good.

Secondly, Ackerman assumes that there is an uncontroversial way of measuring manna: physical quantity. This measure is then used in the neutral dialogue which yields the conclusion that all citizens have the right to equal initial shares of manna, including future generations. However, Ackerman’s criticism of utilitarianism on this point, “. . . [T]he problem with utilitarianism is . . . its effort to evaluate distribution rules by how much “good” they produce. *Any* such effort requires a specification of the good that will be contested by *some* citizens who insist on measuring their good by a different yardstick, one that gives them more manna than their competitors”, holds equally well with respect to his own yardstick (Ackerman (1980), p. 49).¹⁸

Ackerman could modify his theory and say that all the features he ascribes to manna are not obvious truths but something which all citizens could reach a consensus upon. This is quite improbable, we surmise, and Ackerman would need further revisions of his concept of manna to reach such a consensus. At the end of the day, it is likely that he would end up with something similar to Rawls’s more developed theory of primary goods. Consequently, it will share with this theory the problem noted above, that our notion of primary goods may not be what future people consider as such.

V

¹⁷ For a discussion, see Arrhenius (1999).

¹⁸ See Fishkin (1983) for criticism of Ackerman on this point.

How might a proponent of constructivist contractualism respond to our arguments? One possible counterargument is that we underestimate the ability of a political culture to contain implicit ideals that, when established, endure over time and which nevertheless allow for significant historical variation.¹⁹ For instance, the ideal of free and equal persons might be an ideal that is firmly entrenched in current liberal democratic societies. Even if understandings of the ideal vary over time, the implicit ideal itself endures. If there are ideals of this kind, it might be claimed, then the theory can handle future people in an adequate way after all.

There are at least two replies to this kind of counterargument. The first is that even if we assume that there are some implicit ideals that remain largely unchanged over time, significant historical variation regarding the interpretation of these ideals is enough to generate problems for the theory. In particular, our argument regarding neutrality and primary goods is perfectly compatible with the underlying implicit ideal enduring over time. As long as this ideal allows for different interpretations, so that it can be used to derive different accounts of primary goods, our argument remains intact.

In order to further explain this response, recall that Rawls's account of primary goods is based on his conception of persons (or citizens) as free and equal. To be able to reach a specific account of primary goods (e.g. the relevant rights and liberties, and their relative weights) it is necessary for Rawls to give this general idea of the freedom and equality of persons more determinate content. Hence he claims that citizens are free in three respects: (i) They are independent of the particular conception of the good that they possess at any given time, in the sense that they are capable of revising it or abandoning it completely. (ii) They are "self-authenticating sources of valid claims", and thereby entitled to make claims on their institutions. (iii) They are capable of taking responsibility for their ends, and to adjust their claims accordingly (Rawls (2005), pp. 29–34).

The point here is that even if we were to assume that the general idea of persons being free and equal in a fundamental moral sense endures as an implicit ideal, it is both possible and highly likely that the interpretation of this ideal varies quite significantly over time. Rawls's particular explication of the freedom of persons need not be the one that is affirmed by future generations, as they may reject or reinterpret one or more of (i)-(iii). They may thus affirm a quite different interpretation of the freedom and equality of persons, which in turn leads them to a different specification of the primary goods.

As an example of a different interpretation of the freedom of persons, consider the communitarian critique of Rawls's conception of the person. Michael Sandel, for instance, holds that a significant part of our identity is given by our final ends and commitments, which are

¹⁹ We are grateful to Steve Gardiner for raising this issue.

provided by the community with which we identify. It is therefore a mistake, he claims, to characterize us as free in the Rawlsian way. Indeed, if we were to consider ourselves as free in such a way we would not be “ideally free” persons, but rather reduce ourselves to persons “wholly without character, without moral depth” (Freeman (2007), pp. 304–306; Sandel (1982), p. 179). Regardless of the soundness and strength of this critique, it nicely illustrates the very real possibility of significant disagreement on this issue.

Let us now proceed to our second reply, which is based on the observation that even if it is possible for a political culture to contain enduring ideals of this kind, it is surely also possible for such a culture to change more dramatically. That is, for its implicit ideals to be abandoned, and for new ones to take their place. It is for instance surely possible for a liberal democratic society to change into a non-liberal authoritarian one, and for its ideals to change accordingly. As long as such changes are possible, it does not matter whether they are unlikely. Even unlikely events are enough to give rise to the theoretical problems that we have here presented, and hence it will not do to respond by pointing out that political cultures may – or that it is highly likely that they will – contain implicit ideals that endure over time. Moreover, with the current political development and with the increasing pressure that climate change will put on liberal democracies, such changes don’t seem unlikely at all.

In order to offer a more robust response to our arguments a proponent of constructivist contractualism might go further and claim that it is a mistake to apply the theory to the kind of cases that we have here discussed. Such an argument may appeal to Rawls’s own remarks regarding the reach of his theory, as he says that justice as fairness is to be applied to a “modern constitutional democracy” and holds that “[w]hether justice as fairness can be extended to a general political conception for different kinds of societies existing under different historical and social conditions . . . are altogether separate questions” (Rawls (1985), p. 225.). On the basis of such remarks it might be claimed that constructivist contractualism is a significantly limited theory in that it does not only take our implicit beliefs as the starting point, but also that it is only applicable to societies such as ours.²⁰ That is, it operates on the basis of the implicitly shared beliefs of a liberal democratic society, in order to generate principles *for such a society*. It is thus a mistake, according to this line of reasoning, to try to apply the theory to the case that we described above, where a society ceases to be a liberal democracy.

²⁰ Cf. Rorty’s “Ethnocentrism”, Rorty (1989). Rorty makes surprisingly fast transitions from “liberal democrats” via “us, present day liberal democrats”, “us, Western liberal democrats”, to “us, the inhabitants of the rich, North Atlantic democracies”. See Comay (1986). Perhaps the main problem with Rorty’s political philosophy is a kind of dogmatic dichotomy: Either the foundationalism of traditional political philosophy or the ethnocentric beliefs and traditions of a temporary and local “we”. That the liberal democratic tradition could contain beliefs that contradict any temporary and local ethnocentrism, as argued above, is a thought that seems not to have occurred to Rorty.

We consider this suggested limitation of the theory to be questionable. A plausible contractualist theory about justice should be able to handle our duties to future generations not only in cases where a certain political culture endures relatively unchanged over time, but also in cases where more radical changes take place. This point may be most obvious if we focus our attention to our duties to individuals who just happens – through no fault of their own – to find themselves at different points in the development of a society.

Consider the following case: Lucky Ada is born into our society at a future time t_1 when it is still a liberal democracy, whereas unlucky Bea is born into our society at a future time t_2 when our society has changed into an illiberal authoritarian system (with a corresponding change in implicit ideals). If both Ada and Bea have the same moral status – if they both should be considered as free and equal persons – it appears highly implausible to suggest that our duties of justice towards Bea somehow evaporates due to the change that occurs between t_1 and t_2 . If it would be unjust to make Ada suffer as a consequence of present people's overuse of natural resources, then it would also be unjust to make Bea suffer by similar means. Both Ada and Bea should be considered as free and equal persons regardless of the society that they happen to find themselves in, and it appears highly implausible to claim that the societal circumstances that Bea finds herself in make it the case that her equal moral status give rise to no duties of justice towards her. Any plausible theory of justice should be able to accommodate that fundamental insight. If not, it would arguably be in conflict with currently implicitly shared beliefs and, again, a clear violation of the Same Reasons View.

It is an interesting question whether there are any plausible arguments in favour of limiting a constructivist contractualism to liberal democracies only. As indicated above we are sceptical of this idea but unfortunately cannot discuss it further here. Instead, let us note that even if some plausible argument for such a restriction could be made, the resulting limitation would most likely not be enough to rebut the arguments above. As already noted, those arguments merely require that the implicit ideals can be interpreted in different ways. In order to adequately respond to these arguments, it would thus be necessary to limit the theory to the case where not only the implicit ideals remain the same, but the same is true about their interpretation. Such a drastic limitation appears moot, and not one that a proponent of the view would like to defend, we surmise.

Finally, let us comment on yet another possible response to our argument. Someone who wants to defend the lack of neutrality in the case of future generations may suggest that such a limitation can be supported by Rawls's view on justice between peoples. The theory that Rawls presents in his *The Law of Peoples* (1999) is, he claims, to be understood as “the *foreign policy* of a reasonably just *liberal* people”. The idea, Rawls emphasizes, is not to prescribe principles *for* other kinds of societies. Rather, the aim is to determine how a liberal people ought to relate to

other peoples, including “how far nonliberal peoples are to be tolerated” (Rawls (1999), p. 10). In doing so, we are to consider the point of view of such nonliberal peoples, and what is reasonable to accept from their perspective (Rawls (1999), p. 58). But in doing so, we still operate with the ideals and principles of a just liberal society. As Rawls puts it, the fundamental question for the law of peoples is how liberal societies “are to conduct themselves toward other societies from the point of view of their *own* political conceptions” (Rawls (1999), p. 121).

It is tempting to think about future generations in an analogous way. In determining how we, the present generation, should behave in relation to future generations, we need only do so from our point of view. Just as the inhabitants of a liberal society should primarily be concerned about extending their own convictions so as to cover their relations to other peoples, we should rely on our own implicitly held beliefs when we determine how to relate to future generations. Future generations are akin to a different people, though temporally rather than geographically separated from us. If that is the case, perhaps the lack of neutrality identified above can be justified after all.

For our purposes here, we may set aside the various objections to approaching the problem of global justice as a matter of the foreign policy of a just liberal people. Rather than entering that debate, we would like to challenge the proposed analogy between future generations and other peoples. From the point of view of Rawls’s constructivism, the major problem with such an analogy is that it is in direct conflict with one of the fundamental ideas that is claimed to be currently implicitly shared. This is the idea of society as a fair system of cooperation *over time*. The idea of a fair system of cooperation is, Rawls explains, to be conceived of as a society “existing in perpetuity: it produces and reproduces itself and its institutions and culture over generations and there is no time at which it is expected to wind up its affairs” (Rawls (2005), p. 18). If this is so, then future generations cannot be handled as analogous to different peoples. Future generations are members of *our* society, conceived of as existing over time. To treat them in a different way would, assuming that Rawls is correct, be in conflict with current implicit beliefs.

It might be objected that this response only works in cases where the societal changes over time are not too great. In cases where the change is sufficiently radical – as for example when a liberal society transforms into a non-liberal authoritarian one – we should say that the old society has ceased to exist, and a new one has taken its place. It is appropriate in such cases, one might suggest, to treat future generations as analogous to other peoples.

It is an interesting question whether Rawls’s argument in *Law of Peoples* can be extended to future non-liberal societies in a fruitful way. However, this argument can only justify non-neutral treatment of future generations in cases involving radical societal change and is in that respect quite limited in its reach. The problem of non-neutral treatment in the absence of radical

societal change would still remain. Hence, this line of reasoning cannot provide a complete solution to the problem we have presented.

VI

In constructivist contractualist theories, such as Rawls's and Ackerman's, the implications of the justification model can change depending on whose implicit beliefs that are taken into account or who have beliefs that make it possible to reach a consensus upon certain issues. The diverse beliefs of different people can affect what we can consent upon, or what to count as an implicit belief. Thus, for constructivist contractualism, the questions of whose beliefs should count in the construction of the justification model and of whose to exclude, and how to justify these inclusions and exclusions, are of crucial importance. This new assignment problem, as we call it, becomes especially intractable for constructivist contractualism when we consider future generations. We have argued that there is no satisfactory way for constructivist contractualists to incorporate the beliefs of future generations and stay true to their methodological commitments.²¹

²¹ We would like to thank Vuko Andric, Andrea Asker, Paul Bowman, Krister Bykvist, Tim Campbell, Bob Goodin, Göran Duus-Otterström, Julia Mosquera, Gerald Lang, Shlomi Segall, Orri Stefánsson, Folke Tersman, and especially Steve Gardiner for very helpful discussions. Thanks also to the audiences at the Political Philosophy Workshop, Århus University, October 2019; and the participants at the Institute for Futures Studies' PPE-seminar, October 2019, for useful questions and comments. Financial support from Riksbankens Jubileumsfond is gratefully acknowledged.

References

- Ackerman, B. (1980). *Social Justice in the Liberal State*. New Haven: Yale University Press.
- Arrhenius, G. (forthcoming). *Population Ethics: The Challenge of Future Generations*. Oxford University Press.
- Arrhenius, G. (1999). Mutual Advantage Contractarianism and Future Generations. *Theoria*, 65(1), 25–35. <https://doi.org/10.1111/j.1755-2567.1999.tb00112.x>
- Arrhenius, G. (2000). *Future Generations: A Challenge for Moral Theory*. Retrieved from <http://www.diva-portal.org/smash/record.jsf?pid=diva2:170236>
- Arrhenius, G. (2005). The Boundary Problem in Democratic Theory. In F. Tersman (Ed.), *Democracy Unbound: Basic Explorations* (pp. 71–95). Stockholm: Stockholm University. Filosofiska institutionen.
- Arrhenius, G. (2018). The Democratic Boundary Problem Reconsidered. *Ethics, Politics & Society: A Journal in Moral and Political Philosophy*, 2018(1), 89–122. <https://doi.org/10.21814/eps.1.1.52>
- Attas, D. (2009). A Transgenerational Difference Principle. In A. Gosseries & L. H. Meyer (Eds.), *Intergenerational Justice* (pp. 189–218). Oxford: Oxford University Press.
- Barry, B. (1977). Justice Between Generations. In H. L. A. Hart, P. M. S. Hacker, & J. Raz (Eds.), *Law, morality, and society: essays in honour of H. L. A. Hart* (pp. 268–284). Oxford: Oxford University Press.
- Comay, R. (1986). Interrupting the Conversation: Notes on Rorty. *Telos*, 1986(69), 119–130.
- English, J. (1976). Justice Between Generations. *Philosophical Studies*, 31(2), 91–104.
- Fishkin, J. S. (1983). Can There Be a Neutral Theory of Justice? *Ethics*, 93(2), 348–356.
- Flathman, R. E. (1983). Egalitarian Blood and Skeptical Turnips. *Ethics*, 93(2), 357–366.
- Freeman, S. Richard. (2007). *Rawls [Elektronisk resurs]*. London: Routledge.
- Gardiner, S. M. (2009). A Contract on Future Generations? In A. Gosseries & L. H. Meyer (Eds.), *Intergenerational Justice* (pp. 77–118). Oxford: Oxford University Press.
- Gauthier, D. P. (1986). *Morals by Agreement*. Oxford: Clarendon.

- Goodin, R. E. (2007). Enfranchising All Affected Interests, and Its Alternatives. *Philosophy and Public Affairs*, 35(1), 40–68.
- Hare, R. M. (1984). Rights, Utility, and Universalization: Reply to J. L. Mackie. In R. G. Frey (Ed.), *Utility and Rights* (p. 245). Minneapolis: University of Minnesota Press.
- Hobbes, T. (1973). *Leviathan* (10th edition). New York: Collier Books.
- Kymlicka, W. (1990). *Contemporary Political Philosophy: An Introduction* (Second edition). Oxford [England]: New York: Clarendon Press; Oxford University Press.
- Miller, D. (2009). Democracy's Domain. *Philosophy & Public Affairs*, 37(3), 201–228.
- Nussbaum, M. C. (2007). *Frontiers of Justice: Disability, Nationality, Species Membership*. Cambridge: Harvard University Press.
- Rawls, J. (1971). *A Theory of Justice*. Cambridge, Mass.: Belknap Press.
- Rawls, J. (1985). Justice as Fairness: Political not Metaphysical. *Philosophy & Public Affairs*, 14(3), 223–251.
- Rawls, J. (1988). The Priority of Right and Ideas of the Good. *Philosophy and Public Affairs*, 17(4), 251–276.
- Rawls, J. (1999). *The Law of Peoples: With, the Idea of Public Reason Revisited*. Harvard University Press.
- Rawls, J. (2001). *Justice as Fairness: A Restatement* (E. Kelly, Ed.). Cambridge: Harvard University Press.
- Rawls, J. (2005). *Political liberalism*. New York: Columbia University Press.
- Ripstein, A. (1987). Foundationalism in Political Theory. *Philosophy & Public Affairs*, 16(2), 115–137.
- Rorty, R. (1989). *Contingency, Irony, and Solidarity*. Cambridge: Cambridge University Press.
- Sandel, M. J. (1982). *Liberalism and the limits of justice*. Cambridge: Cambridge University Press.
- Scanlon, T. (1998). *What We Owe to Each Other*. Cambridge, Mass.: Belknap.
- Scanlon, T. (2014). *Being realistic about reasons*. Oxford: Oxford University Press.

- Street, S. (2008). Constructivism about Reasons. In R. Shafer-Landau (Ed.), *Oxford Studies in Metaethics Volume 3* (pp. 207–245). Oxford: Oxford University Press.
- Tersman, F. (1993). *Reflective Equilibrium: An Essay in Moral Epistemology* (Doctoral Thesis). Stockholm University, Stockholm.