Paper accepted for

ECPR Joint Sessions, Workshop 10
Helsinki, 7 - 12 May 2007

RASMUS KARLSSON

Beyond the Rawlsian monologue
- a two-stage approach to intergenerational justice

Department of
Political Science
ABSTRACT

Recent works have advanced our theoretical understanding of the proper scope, shape, and currency of intergenerational contractual justice. But as attention is drawn to the task of formulating practical policy-options capable of providing a sustainable trajectory into the future, the monological reasoning of these contractual theories seems to offer little guidance.

In line with criticism articulated by Jürgen Habermas and Seyla Benhabib, the paper acknowledges the value of dialogical thinking in moral decision-making. Taking this argument to issues of intergenerational justice, it is suggested that instead of trying to distillate one single fair path to the future we should employ a two-stage contractual approach which allows a plurality of views about what the deep future should be like. By doing so, it may be possible to sidestep several notorious challenges to contractual theories, including different repopulation paradoxes, while reaffirming the fundamental political nature of intergenerational affairs.
Introduction

It is likely that each generation considers themselves to be living at a critical juncture in time. Yet, it is hard to argue that the stakes for humanity have ever been higher than in the past half century. While most treaties on intergenerational justice tend to focus on the risks entailed by modernity (such as global climate change, resource depletion, and ultimately thermonuclear destruction) it may be that the same transformative forces can be used to bring an end to poverty (Sachs, 2006), find advanced technological paths to climate stability (Hoffert, 2002), and finally foster a cosmopolitan sensibility which reflects our economic interdependence. Either way, we are now in a position to fundamentally alter the future of all human beings and that alone places an unprecedented moral and political responsibility on present generations.

Theories of intergenerational justice try to make philosophical sense of that responsibility. Early on, philosophers realised that many familiar notions of justice, equality, and utility are of little help as we turn to the future. Most disturbingly, the kind of systematic thinking that sparked the contemporary revival of political philosophy has proven to be exceptional vulnerable to issues of futurity (Laslett & Fishkin, 1992). Trying to tackle these vulnerabilities, the last decades have seen a growing academic interest in intergenerational justice, motivated not the least by the mounting evidence of global climate change (Page, 2006). Drawing on this work and the advanced theoretical vocabulary that has been developed within the contractual tradition, this paper will explore a new two-stage approach to intergenerational justice. By splitting the contractual situation into its procedural and substantial components, it is suggested that we can avoid certain philosophical paradoxes endemic to contractual thinking. While maintaining a Rawlsian monological mode of reasoning during the first procedural stage, the second substantial
stage employs a dialogical mode of thinking which acknowledges the fundamental political nature of intergenerational affairs. More specifically, the first procedural stage aims at constructing a norm of intergenerational conduct which no generation reasonably would reject whereas the second substantial stage is about debating the practical means of meeting that norm.

This two-stage approach to intergenerational justice flows from a pragmatic understanding of contractualism. In other words, I am not primarily concerned with constructing a water-tight metaphysical specification of the contractual situation but rather with developing better pragmatic tools that can be used in real-world deliberations on intergenerational affairs.

**Contractualism in intergenerational justice**

Following its wider ascendancy in modern political philosophy, contractual thinking has emerged as a promising approach for understanding our obligations to posterity. Even as it seems to offer important advantages over other competing ethical approaches (such as utilitarian consequentialism) it has nonetheless remained haunted by a few long-lived paradoxes, most notably the so called non-identity problem. Before I go on to describe what exactly it is that gives rise to these paradoxes, I will briefly review the concept of intergenerational contractual justice as it has evolved through the literature.

Unlike many other philosophical concepts, intergenerational justice does not come with an ancient history. One possible reason to this is that, in the pre-modern world, human actions had limited spatial and temporal extension. Though religious writings have always remained rife with apocalyptic undertones (O'Leary, 1994), the idea that man himself would bring an end to all human civilization was probably unthinkable to the ancients. Only in the last century, with the
development of nuclear weapons, did we begin to understand that our future as a species was by no means guaranteed and that our present actions were indeed likely to have far reaching consequences for the wellbeing of future generations.

When John Rawls published his magnum opus *A Theory of Justice* in 1971, he included a chapter on intergenerational justice which marked a significant contribution to the emerging philosophical debate on the topic. As often in political theory, much that has been written since can be seen as a response to the views that Rawls articulated. Rawls arrived at the question of intergenerational justice when trying to work out what would be a fair “social minimum” (Rawls, 1971, p. 284). The social minimum is the bundle of resources that a person needs to lead a minimally decent life in a particular society. Clearly, the level of such a social minimum is bound to depend on the historical circumstances and the overall wealth of that society. These factors are in turn determined by (a) what saving earlier generations have done and (b) how far “the present generation is bound to respect the claims of its successors” (Rawls, 1971, p. 284). Thought through, this meant that the very ability of a society to establish and preserve just institutions was dependent on to what extent other generations would be motivated to adopt a “just savings principle” (Wall, 2003).

This insight troubled Rawls since he had already specified elsewhere that “the persons in the original position know that they are contemporaries” (Rawls, 1971, p. 140) even if they would remain ignorant of what generation it was that they collectively belonged to (Rawls, 1971, p. 287). Being contemporaries they know that by the time they will come into existence “either earlier generations have saved or they have not; there is nothing the parties can do to affect it” (Rawls, 1971, p. 292). Similarly, they stand not to gain anything themselves by saving for successive generations.
One characteristic feature of justice as fairness is that the contractual parties are mutually disinterested, motivated neither by benevolence nor by envy or spite. Instead it is the alchemy of the original position, the state of unresolved identity, that makes it necessary for them to come up with principles that are acceptable to everyone from a position of equality. However, if the parties are allowed to know that they are contemporaries, nothing would prevent them from taking unfair advantage of their temporal position. And as indicated above, Rawls knew that the problem of justice between generations worked in both directions; not only is the present stage of civilization dependent on the actions undertaken by earlier generations, our own actions will also determine the range of options available to future generations and to what extent they will be able to uphold just institutions and the fair value of liberty.

In 1971, Rawls seemed to think that the only way each generation could be made to “carry their fair share of the burden of realizing and preserving a just society” (Rawls, 1971, p. 289) was by modifying the motivational assumption so that the parties would be regarded as head of families with ties of sentiment between successive generations:

“Thus imagining themselves to be fathers, say, they are to ascertain how much they should set aside for their sons by noting what they would believe themselves entitled to claim of their fathers” (Rawls, 1971, p. 289)

Naturally, this amounts to a substantial recasting of his original theory and, as Jane English demonstrated in her famous article, the switch from viewing the parties as individuals to viewing them as family lines comes at a high price (English, 1977).1 An alternative approach, advocated by Brian Barry among others, would be to reject what

---

1 Another possible objection to the 1971 account of Rawls’s theory is that it does not exclude one generation damaging another remote generation in a way that bypasses mutually successive generations (Barry, 1989, p. 193; Routley & Routley, 1978, p. 167).
Rawls calls “the present time of entry interpretation” and with it the postulate that all parties in the original position are contemporaries (Barry, 1977). At first this seems to be an attractive move and actually more in line with, what many of us thought was, the general theme of Rawls’s project, namely to use the original position to find out what is owed or justifiable to individual free and equal persons irrespectively of morally arbitrary factors such as nationality (Nussbaum, 2006, p. 264) or, in this case, date of birth.

However, even as Rawls eventually came to develop his views on intergenerational justice (Rawls, 1993, p. 20), later works such as *The Law of the Peoples* suggest that he was perhaps all along more interested in constructing justice for existing practice (James, 2005) and for one particular society (read the United States of America) than in finding principles that were truly universal in scope. Returning to the 1971 edition, Rawls in fact explicitly rejected an atemporal reading of his theory even as he recognized that the inclusion of “representatives from all actual generations” (Rawls, 1971, p. 291) would make it unnecessary for him to change the motivational assumption. Given to what extent the proposed modification undermined the overall simplicity and consistency of his theory, it is of course tempting to ask why he felt compelled to carry it out and not, as Barry wrote in a characteristic remark, “pursue the logic of his own analysis more rigorously” (Barry, 1977, p. 280)?

**The repopulation paradox**

It is instructive to note that Rawls thought that the question of justice between generations “subjects any ethical theory to severe if not impossible tests” (Rawls, 1971, p. 284). And his phrasing above, when he talks of “representatives from all actual generations”, indicates that he early on foresaw some of the problems associated with extending the original position so that all generations would be present behind
the veil of ignorance. Not only will it be difficult to draw neat cut-off lines for when the human lineage starts and ends respectively, we also know that the choices made at a hypothetical intergenerational meeting are going to influence how long the human tenure of the planet will be and consequently how many people will come into existence and when. As Barry lays out the puzzle:

“Isn’t there something incoherent in combining the idea that people in the original position are choosing among policies which will produce different total numbers of people with the idea that they know at the outset that they are all the people who ever have existed or ever will exist? It is surely a curious sort of choice if the results of it are already instantiated in the composition of the group of people doing the choosing!” (Barry, 1977, p. 271)

This formidable philosophical problem is one, though remote, relative of what in the literature is referred to as the repopulation paradox (Partridge, 1990), the non-identity problem (Parfit, 1984) or the case of the disappearing beneficiaries (Schwartz, 1978). More generally, the repopulation paradox stems from the observation that it is a very contingent matter which particular people get to exist.

We know that our identity is at least partially determined by our genetic profile. That genetic profile is in turn the result of the coming together of a singular egg and a singular sperm, a process which is highly sensitive to any alteration. Imagine that the *Titanic* would not have collided with an iceberg on that freezing night in April 1912. Then the news of the catastrophe would not have spread all over Europe and the Americas on the following day, not interrupting people in their daily routines, and, ultimately, leading to that different people were procreated. If a single disaster like the sinking of an ocean liner in 1912 seems to have such profound consequences, then consider:

---

2 This is different from saying that our genetic profile constitutes our identity, just consider the case of one-egg twins who certainly have different identities yet share the same genetic material.
“All long-run social policies affect the composition of the population to such an extent that after a relatively small number of generations (six to eight) the chances are infinitesimal that any individual who exists under the regime of a long-run policy would have existed if the policy had not been adopted, and conversely.”
(Galston, 1980, p. 252)

If this argument holds, it would have overwhelming implications for a hypothetical intergenerational meeting. For each possible choice made at the meeting there would be a corresponding set of people who owe their very existence to the “activation” of that future. And, as long as these people lead lives that are worth living, they would most certainly consider their own future to be preferable to any other future in which they do not exist (Carter, 2001, p. 440).

At first we may think that the veil of ignorance in itself would be of assistance here. Most recently, Jeffrey Reiman has taken up that thread by arguing that “future people are represented in the original position as the people who will exist in the future whatever particular individual traits they turn out to have, indeed, whatever particular individuals they turn out to be” (Reiman, 2007, p. 82). However, what Reiman and others tend to overlook is that the original position is essentially a device for representation. Because even if we were able to conceal all individual characteristic behind a “thick” veil of ignorance, we are still left with the problem that different choices at the meeting will produce different total numbers of people who are distributed differently throughout the time-space continuum. And any substantial policy, such as what amounts to a fair consumption rate of non-renewable resources, is bound to depend on how many future people we expect there to be and at what time. By the same token, any attempt to extend the veil so the parties do not know if they will exist (Mulgan, 2006, p. 43) is also fundamentally unsatisfactory.
A singular timeline

I believe that severe as the repopulation paradoxes above may seem, they are far less fatal to contractual theories than to other person-affecting theories of justice. Before I give further reasons for this optimism I will, in a related move, make an attempt to simplify the task at hand considerably by leaving the future and pointing our attention in the opposite direction.

Though probably uncontroversial to most people, we must first accept that time is “continuous, linear, unidirectional and irreversible” (Bell, 2003, p. 140). Whatever has happened in the past cannot be undone, at least as far as our present understanding of physics goes. This means that as we move further along the time-space continuum we leave behind only one singular fixated timeline. I take this to imply that, as we cannot change past historical events, it makes no sense to include past generations in a hypothetical contractual situation. Doing so would open the door to counterfactual reasoning such as the non-existence of the present. To exclude past generations does not mean that we have to reject all notions of posthumous interests (Callahan, 1987; Page, 2006, pp. 124-128), only that we do not have to concern ourselves with finding for instance a saving principle that is applicable to intergenerational exchanges taking place prior to the present (Rawls, 1971, p. 290).

Having adopted the present as a baseline we can define the scope of justice so that our moral community is confined to all present and future generations of human beings (Page, 2006, p. 51). I will not here address the question of how deep into the future we shall extend our

---

3 Rawls found the difference principle unsuitable for intergenerational justice on precisely the ground that “there is no way for later generations to improve the situation of the least fortunate first generation” (Rawls, 1971, p. 291).
moral community. Though it may take considerable time until natural genetic processes have changed our DNA so that it becomes warranted to speak of a new species, we may perhaps ourselves intervene at a much earlier stage (Fukuyama, 2003; Habermas, 2003).

A two-stage approach

As often in analytical philosophy, we have to be very careful with the words we use. In the case of a hypothetical intergenerational meeting we have this far talked about the decisions made at the meeting as “choices”. A choice however, presupposes different options to choose from. And it is here that is starts to get tricky:

“Of course, we may still say that the people in the original position reach an agreement, but this is entirely trivial since they are clones. All we are saying is that, faced with identical information and reasoning in an identical fashion, they arrive at identical conclusions” (Barry, 1995, p. 58)

This Rawlsian monologue means that there can not be a reasonable pluralism at a hypothetical intergenerational meeting. It is not like that some people in the original position will favour one future while others will favour another. Instead they will all favour one single trajectory into the future (coincidentally, the one in which they exist).

Let us now recall the fundamental singularity of the timeline. It implies that if we were to take up a vantage point deep into the future we would look back towards the present and see only one fixated timeline. And spread out along that timeline, there will be one single set of people who require us to fulfil our moral obligations to them. These obligations will be founded on justice and consist of certain perfect duties (Partridge, 1990, p. 44).

---

4 The matter is not dismissed because it is trivial but rather because it is part of a much larger theoretical discussion (Temkin, 2005).
What frustrates us is that the repopulation paradox prevents us from using a one-stage contract, of the kind envisaged by Rawls, to compare different futures. But let us now split the contractual situation, so that we separate its procedural and substantial components.

**The procedural component**

In international relations, the right to self-determination is a fundamental concept (Raic, 2002). Ideally, peoples and nations are given the right to sovereignty on the basis that they are considered the best judges of their own good (Jackson, 2005, p. 120; Kant, 2006, p. 70). What is often overlooked though is how well this right to self-determination fits into the intergenerational picture. Let me give a practical illustration by considering how the differences between a Frenchman and a Swede living today pale in comparison with the differences between a Swede living today and a Swede living three hundred years ago. Just take any issue like gender equality and we realize that the differences today are practically irrelevant compared to the differences over time.

Applying this example to the problem at hand, I believe that it is reasonable to argue that any intergenerational meeting would first and foremost favour the right to self-determination, giving each generation the means of pursuing their own particular conception of the good.

In the vocabulary of contractual theorists, we are here concerned with defining the *currency of justice* (Page, 2006, p. 54). Throughout the literature, different competing currencies have been proposed. Following Bryan Norton and in line with the idea of self-determination, I suggest that we should define the currency of justice as “options and opportunities for the future” (Norton, 1999, p. 118). More specifically, as discussed by Richard Howarth, we may think of a structured bequest package that “includes specific endowments of
reproduced capital, technological capacity, natural resources and environmental quality” (Howarth, 1997, p. 570). In the context of the two-stage approach to intergenerational justice this means that, in any contractual situation, the parties would favour a norm saying that each generation has the right to a fair and non-diminishing level of opportunity handed down through a structured bequest package from previous generations. To the extent that this is a norm which no generation would reasonably reject, independently of what substantial policies they may favour, we can consider the adoption of such a currency of justice to be the fair outcome of the first procedural stage.

The substantial stage

Initially, the norm arrived at during the procedural stage may seem to be of little help to our endeavour. We are still painfully aware that an intergenerational meeting cannot decide what substantial policy to pursue. However, we can. In fact, while acknowledging that the choice is to be made under certain ecological, technological, and social constrains, we are indeed free to choose between different futures and also politically debate their respective merits. For each future we imagine, different people – to whom we have moral obligations due to the singularity of the future – will come into existence. Each time we can ask ourselves that deceptively simple question: “what would we want the present generation to do if we were in the shoes of some future unborn generation?” (Epstein, 1992, p. 84). More specifically, by setting up the second substantial stage anew for each possible future, we can ask ourselves how well that trajectory corresponds to the norm arrived at during the procedural stage.

\footnote{Of course, we cannot “fixate” the timeline with any greater accuracy. In fact, due to the quantum reality of the universe (and somewhat more controversial, due to our own free will), not even a Laplacian demon would have that ability. Yet, we share the intuition that each individual now belonging to the “class of future people” will eventually have the same moral standing as anyone now living (Partridge, 2002, p. 81). It is that intuition that the contractual device tries to reflect as it embodies a conception of equality and spatial-temporal neutrality.}
To some, this may seem like cheating. However, I believe it to be fully consistent with our pragmatic reasons for using contractual thinking at the first place, namely to expand our moral sensitivity beyond the present. By probing different futures and speculate about how the level of opportunity is likely to be distributed, we have constructed an instrument for policy-evaluation which, even as it remains undeveloped, allows us to substantiate our obligations to posterity. While we maintain the Rawlsian monological reasoning during the procedural stage we do not try to distillate a single fair path into the future. This may seem as a failure for our theory. But as both Seyla Benhabib (Benhabib, 1992, p. 163) and Jürgen Habermas (Habermas, 1995) have argued, the monological model of moral decision-making favoured by Rawls may in fact be inferior to a “dialogical model of real-life argumentation among a plurality of participants endorsed by communicative ethics” (Sterba, 1994, p. 155). Though I, like James Sterba, believe that the two models of reasoning may serve different purposes and are surprisingly compatible, it is important to recognize that the question we are dealing with here is not just any question. In fact, to decide on a path into the future is the very definition of politics. It is therefore not surprising that we arrive at the substantial stage with a plurality of visions of the future, each informed by different empirical and normative positions.

To be successful, I believe that a theory of intergenerational justice should acknowledge this plurality yet offer a common set of methodological tools that can be used in a deliberative, or reflective, manner to give a voice to those unborn who, though definitely affected by current actions and choices, remain “unavoidably incapable of giving voice to their own concerns” (Goodin, 2003, p. 14). By constructing a norm of intergenerational conduct, the procedural stage provides exactly such a common yardstick that can be used when debating the merits of different political visions of the future.
So what?

When doing philosophy there is always that frightening question, so what? Why do we need all these theories to understand our obligations to posterity?

All over the world, a lot of practical political work is now being done to promote what is known as “sustainable development”. Though it may seem straightforward enough to simply try to reduce the ecological footprint of humanity, we are still faced with many difficult moral and political questions about our own place on this planet. Should we favour a deep ecological vision of the future in which a small human population live in close interconnectedness with nature? Or should we instead try to reduce our ecological footprint by radical technological innovation and ultimately space colonization (Karlsson, 2006)?

Independently of what vision we personally tend to favour we can expect future generations to have strong views on the decisions we make. Yet, a fundamental asymmetry is here at play. Though future generations may benefit or suffer from our political decisions, they cannot hold us accountable. Nor do they possess any bargaining power when it comes to influencing our actions. In fact they are only “virtually” represented as far as our moral sentiments permit. This is why a contractual theory aimed at enhancing those sentiments and intuitions becomes a priority.

If successfully communicated, hypothetical contractualism would offer an overarching ethical framework within which issues like sustainable development can be analyzed. It is my hope that the two-stage approach to intergenerational justice presented in this paper may contribute in making that political analysis somewhat less confusing.
References


Biographical note:

Rasmus Karlsson is a PhD Candidate in political science at Lund University, Sweden. His research interests traverse theories of intergenerational justice, sustainable development, and the temporal dimension of democracy.