Johanna Thoma

Bargaining and the impartiality of the social contract

Article (Accepted version) (Refereed)

Bargaining and the Impartiality of the Social Contract

Johanna Thoma, University of Toronto

Abstract

The question of what a group of rational agents would agree on were they to deliberate on how to organise society is central to all hypothetical social contract theories. If morality is to be based on a social contract, we need to know the terms of this contract. One type of social contract theory, contractarianism, aims to derive morality from rationality alone. Contractarians need to show, amongst other things, that rational and self-interested individuals would agree on an impartial division of a cooperative surplus. But it is often claimed that contractarians cannot show this without introducing moral assumptions. This paper argues that on the right understanding of the question contractarians are asking, these worries can be answered. Without relying on moral assumptions, the paper offers a novel derivation of symmetry, which is the axiom responsible for the impartiality of the most famous economic bargaining solutions appealed to by contractarians.

1 Introduction

The question of what a group of rational agents would agree on were they to deliberate on how to organise society is central to all hypothetical social contract theories. If morality is to be based on a hypothetical social contract, then we need to know what the terms of this contract are. One type of social contract theory, sometimes called contractarianism, aims to derive morality from rationality alone. What contractarians need to show is that the rules that we commonly think of as constituting morality would be agreed upon by rational and purely self-interested individuals in a pre-moral context. In particular, contractarians have aimed to show that rational and self-interested individuals would divide the fruits of cooperation in a fair or impartial way. After bargaining theory developed as a branch of game theory in the 1950s, a number of contractarians followed Braithwaite's (1955) suggestion and used solutions from economic bargaining theory to try and show this. Gauthier's *Morals by Agreement* (1986) was the most prominent and elaborate such attempt. But the game theoretic analysis of the bargaining problem also gave rise to the worry that no determinate and satisfactory answer can be given to the question of what rational, self-interested individuals would agree on in a pre-moral context.

Any attempt at answering this question seems to be subject to one of three criticisms: First, moral considerations may already have played a role in the determination of the outcome of bargaining. This is a problem for contractarians, because it means they failed their project of deriving morality from rationality alone. Second, the outcome of bargaining may give weight to factors that are morally arbitrary, and thus fail a requirement of impartiality. This is a problem for contractarians because they will have failed to derive anything that is recognizable as morality from their non-moral assumptions. If these two problems are avoided, a third one seems to arise: There may be no unique rational outcome of bargaining. Without making any normative assumptions, or arbitrary assumptions about the bargaining process, the answer to what rational agents would agree on in bargaining may have no determinate answer. These criticisms were most forcefully made by Sugden (1990) and, more recently, Thrasher (2014).

This paper argues that there is a way of understanding the bargaining problem that contractarianism sets itself, according to which all three problems can be avoided. According to this understanding, we have the resources to show that there is a unique outcome of bargaining, which can be derived without making any moral assumptions, but which results in a rule for distributing the fruits of cooperation which is impartial, and hence fulfils an important requirement of what we think of as morality.

The first key to my resolution of the problem is to show that no standard non-cooperative game can be used to model the process of bargaining if we make the following two assumptions: Firstly, the process of bargaining is something that is itself to be agreed on; And secondly, bargaining is completely costless. The problem that is sometimes identified with this setup is that under these conditions, rational agents may never reach agreement. It is here that I think a proper understanding of the question contractarianism is asking is needed, and this is the second key to my resolution of the problem. The hypothetical question we should understand contractarianism as asking is not 'What would happen if rational agents were to bargain?', but 'What would rational bargainers agree on, provided they reach agreement?'. Agents are entitled to ask that question if they take agreement amongst rational agents to at least be *possible*, and if they already consider it rational to cooperate. If that is the question, then we can justify bargaining solutions by hypothesising an agreement point, and asking what this agreement has to be like given that it commanded universal agreement from rational agents. I will show that, in agreement with all the candidate fair bargaining solutions, any such agreement will have to be Pareto efficient, symmetric, and thought of by all agents as unique. This goes a long way towards showing that a principle that captures an idea of impartiality can be derived from rationality alone, and hence that the contractarian project is not hopeless.

2 The Contractarian Project

Contractarians like Gauthier, Narveson or Buchanan have traditionally aimed to derive morality from rationality alone. To achieve this project, two important conditions must be met: our starting point must be non-moral, and the principles that derive from our theory must be recognisably moral, which, at a minimum, means that they must be sufficiently impartial. Contractarianism takes the moral sceptic seriously. It aims to justify morality to somebody who takes all her reasons to be grounded in rational self-interest. The founding insight of contractarianism is the following: Mutual advantage is possible if agents manage to constrain their pursuit of self-interest somehow. And so constraint of the pursuit of self-interest is ultimately in an agent's interest. Contractarians argue that the constraints that self-interested individuals thus have reason to follow turn out to be what we commonly think of as moral constraints.

The founding insight that constraint of self-interest makes possible mutual advantage goes back at least to Thomas Hobbes' *Leviathan* (1651/2010). Today it is often interpreted in game theoretic terms, in particular in terms of the Prisoner's Dilemma.[1] If each individual maximises her personal gain in a Prisoner's Dilemma situation, the overall outcome will not be Pareto efficient: In fact, everybody could be made better off if the individual agents could all commit themselves to cooperate. And there are many other types of interaction that are structured such that a constraint on individual utility maximisation can lead to an outcome that is mutually beneficial, such as the production of public goods, or the use of pooled resources. The more general point is that in strategic interactions, there is no guarantee that Nash equilibria (which are usually taken to be the individually rational outcomes of a game) will be Pareto efficient.

[1] See Gauthier (1969) for a reading of Hobbes' Leviathan in game theoretic terms.

Contemporary contractarianism is about how the benefits from cooperation in interactions that are so structured can be reaped, and how they should be divided. Seeing that the benefits to all from mutual constraint are large, Gauthier argued that reason commands agents to give up what he calls the 'presumption against morality':

> In so far as [the parties] would agree to constraints on their choices, restraining the pursuit of their own interests, they acknowledge a distinction between what they may and may not do. As rational persons understanding the structure of their interaction, they recognize a place for mutual constraint, and so for a moral dimension in their affairs. [. . .] Agreed mutual constraint is the rational response to these structures. Reason overrides the presumption against morality. (Gauthier 1986, p.9)

Note that Gauthier is not saying yet that it follows from the recognition that mutual constraint is needed to reap the benefits of cooperation that it is rational for agents to constrain their pursuit of self-interest. If we are happy to ascribe rationality to collective agents, then in some sense it would be rational for 'us' to cooperate. But that need not give the individual a reason to cooperate yet.

This is why the contractarian argument must proceed in two steps: At the first step, we ask what division of the gains from cooperation it would be rational for the involved parties to agree on. This is the question bargaining theory is supposed to give the answer to. At the second step, the contractarian wants to show that it is rational for agents to act according to such a hypothetical agreement. It is at this second step that Gauthier provides his much-contested theory of constrained maximisation. It aims to show that individuals in fact have a reason to do their part to make cooperation possible, and hence de facto give up the pure pursuit of the greatest satisfaction of their self-interest.

What I am interested in here, however, is the first step of the argument. It is here that contractarians can use bargaining theory in order to answer the question of how rational agents would agree to divide a cooperative surplus. In line with the contractarian project, the agreement should be imagined to take place in a *pre-moral* context. Bargaining, to the contractarian, is also *hypothetical*: We do not in fact find ourselves in a pre-moral context, and we may not be perfectly rational. Moreover, cooperation is not usually preceded by an explicit agreement. And so we imagine agents who are about to cooperate wondering what division of the cooperative surplus they would have agreed on had they bargained in a pre-moral context under conditions of full rationality, and common knowledge thereof.

To answer this question, economic bargaining theory seems well-suited. For the contractarian project to succeed, the answer should correspond to a division that is in some sense impartial and could be described as moral.

3 Axiomatic Bargaining Theory

Economic bargaining theory concerns the way (rational) agents divide, or should divide goods amongst themselves, when all agents would be worse off if no agreement was reached. This branch of game theory thus appears to have obvious relevance for contractarians. Two different approaches can be distinguished: an axiomatic approach, and a strategic one. The axiomatic approach specifies some characteristics the outcome of bargaining should have in the form of a number of axioms, and derives a unique bargaining solution from those axioms. The strategic approach specifies more precisely the process of bargaining: most commonly, it models it as an offer-counteroffer non-cooperative game, where the notion of a Nash equilibrium or a refinement thereof is used to solve the game. This section provides an introduction to the axiomatic approach, which is the approach I ultimately want to defend. The next section will turn to challenges to the contractarian use of bargaining theory.

The most famous axiomatic bargaining solution was offered by Nash (1950), whose characterisation of the bargaining problem is still the starting point for most of bargaining theory. Focusing on a two agent case, let us assume two agents a and b have utility functions u_a and u_b. Each could adopt a number of different strategies, and each combination of strategies affords them each a certain utility. We can then define a feasible set of pairs of utilities (or outcomes) that the two agents could bring about. Call this the bargaining set B, which we assume to be closed. This set can be represented as an area in two dimensions, as in Fig. 1. When agents can randomise over strategies, the bargaining set is always convex, because agents can always achieve an outcome on the line connecting two other feasible outcomes by randomising over the strategies that generate those. A bargaining problem is now defined by a bargaining set and a disagreement point d. The disagreement point is the set of utilities that the bargaining parties would walk away with if no agreement was found.

Axiomatic bargaining solutions aim to specify what set of utilities rational agents would agree on, using nothing but the bargaining space and the disagreement point as input: They are functions F on (d, B). To find such a function, we take for granted that agents will never agree on a set of utilities that affords either of them a utility smaller than that of the disagreement point. So we can focus on the area S in Fig. 1.

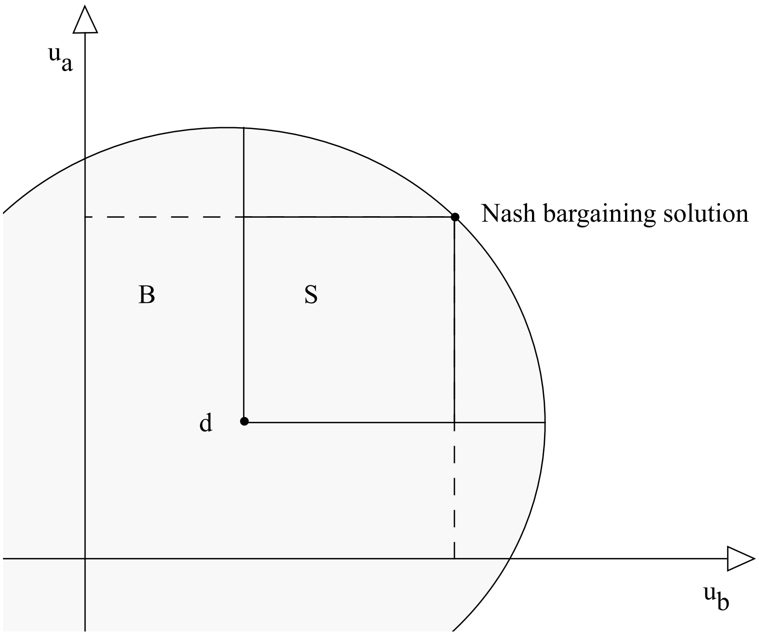

Figure 1: A bargaining problem and the Nash bargaining solution

The Nash bargaining solution can now be derived from four axioms:

1. **Invariance with respect to linear, strictly increasing affine transformations:** Let $A(.)$ be a positive linear transformation. Then,

$$f(A(d), A(B)) = A(f(d, B)).$$

   What this means is that if we rescale each agent's utility function, the bargaining solution should also merely be rescaled in the same manner. Importantly, this implies that the bargaining solution cannot depend on any interpersonal comparisons of utility.

2. **Pareto efficiency:** For every (d, B), there is no $y \in B$ such that for every agent i,

$$y_i >= f_i(d, B) \text{ and } y \neq f(d, B).$$

So, at the bargaining solution, it must be impossible to make one bargainer better off without making the other worse off. Graphically, what this means is that the bargaining solution must be on the right and top boundary of S, which is often called the Pareto frontier.

3. **Independence of irrelevant alternatives:** Suppose (d, B_1) and (d, B_2) are different bargaining problems such that $B_1 \subset B_2$ and $f(d, B_2) \in B_1$. Then,

$$f(d, B_2) = f(d, B_1).$$

   The intuition behind this is that alternatives which would not be chosen were they available should not make a difference to what is in fact chosen.

4. **Symmetry:** Suppose S is symmetric in the sense that there exist utility operators such that $(u_a, u_b) \in S$ if and only if $(u_b, u_a) \in S$. In this case, the solution f should be of the form $(x,\, x)$, that is, a point on the line $u_a = u_b$. So in symmetric situations, the agents should each receive the same amount of utility.

Nash proved that the only function that satisfies all four axioms is the one that maximises the so-called Nash product $(u_a - d_a) \times (u_b - d_b)$ subject to $u_a, u_b \in S$. Graphically, the Nash solution can be found where the largest possible rectangle in S touches the Pareto frontier, as in Fig. 1.[2]

The second most famous bargaining solution is the Kalai-Smorodinski, or KS bargaining solution (see Kalai and Smorodinsky 1975). It shares Nash's representation of the bargaining problem, and Axioms 1, 2, and 4. However, it replaces Axiom 3 with an axiom we will call monotonicity. To define it, we first need to define the ideal point c: For each bargainer, it specifies the maximum utility they may receive if the other bargainer were to receive just the disagreement utility. These utility levels are not jointly feasible, and so lie outside the bargaining set, as can be seen in Fig. 2. Monotonicity now requires the following:

[2] Lensberg (1988) showed that the Nash bargaining solution can also be derived when we substitute an axiom called 'stability' for the independence of irrelevant alternatives. This axiom is often considered more plausible than independence of irrelevant alternatives, and claims that, in a multiple person bargaining game, the bargaining solution must be such that, if we hold the utility allocation to some players fixed and apply the bargaining solution to the subgame involving the remaining players, those players must be allocated the same amount as when the bargaining solution was applied to the original bargaining game.

3*. **Monotonicity:** For any two bargaining problems (d, B_1) and (d, B_2) such that $B_1 \subset B_2$, it should be the case that, for every agent i,

$$f_i(d, B_2) >= f_i(d, B_1).$$

The intuition here is that no player should get less as the bargaining space is expanded.

Kalai and Smorodinsky proved that the unique bargaining solution that satisfies 1, 2, 3* and 4 is the one that lies at the intersection of the Pareto frontier and a line connecting the disagreement point with the ideal point, as shown in Fig. 2.

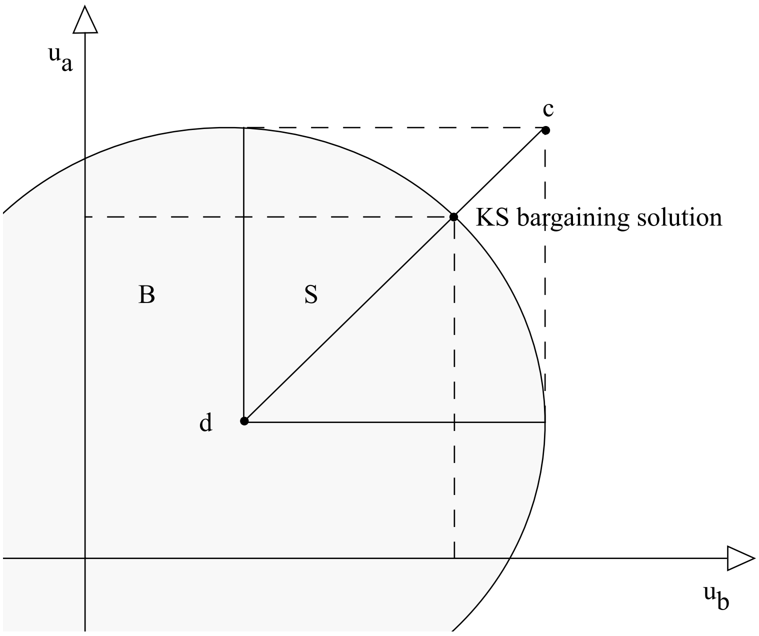

Figure 2: The KS bargaining solution

The bargaining solution Gauthier presents in chapter 5 of *Morals by Agreement* is formally equivalent to the KS solution, at least in bargaining that involves two parties (see Gaertner 2006, p.142 ff.). He thinks that the outcome of rational bargaining will be at the point of 'minimax relative concession', that is, at the point where the greatest concession made by any of the bargainers is smallest. Gauthier sets up the problem much like Nash does, with a bargaining space and a disagreement point. He also appeals to a 'claim point' which is equivalent to the ideal point in the KS solution. A relative concession is now defined in the following way: Take any utility pair u in B. The relative concession that would be involved in adopting u as a solution is for each bargainer i the difference between the claim point utility c_i and u_i, relative to the difference between c_i and the disagreement utility d_i, or $(c_i - u_i)/(c_i - d_i)$. In a two-agent bargaining problem, the minimax relative concession will also be found at the point where the line between d and c crosses the Pareto frontier.

While there are some interesting differences between the KS and Nash bargaining solutions, what I want to focus on here is the general approach they embody, and the axioms they both share, in particular the symmetry axiom.[3] The question I want to address is how bargaining solutions of the type just described could help the contractarian project of deriving morality from rationality. The next two sections raise problems for contractarians who want to employ the bargaining solutions in this way. In Sections 6 and 7, I will provide a defence on behalf of the contractarian.

4 Norms or Partiality

A look at the standard interpretations of bargaining theory suggests that the contractarian project is hopeless from the start. Contractarians want to use bargaining theory to answer the question of what division of a cooperative surplus rational agents would agree on in a pre-moral context. This concern means that they cannot share what has become the dominant interpretation of axiomatic bargaining theory. This interpretation sees axiomatic bargaining solutions not as predictions of what rational agents will agree to do, but as attempts to capture what fairness in bargaining consists in. In fact, it is common for economists to interpret axiomatic bargaining solutions as those that would be picked by an impartial third party. Take, for instance, Pfingsten and Wagener (2003):

> In a cooperative bargaining problem, an element out of a set of feasible utility vectors for a group of agents has to be selected. The selected outcome is regarded as the cooperative agreement reached by the agents. In the axiomatic approach, the outcome further is interpreted as the recommendation which an impartial arbitrator who holds certain normative positions would give as to how the bargaining problem should reasonably be solved. (p. 360)

[3] This is especially apt given the fact that Gauthier later gave up his original bargaining solution and endorsed the Nash bargaining solution (see Gauthier 1993).

If this is what axiomatic bargaining solutions do, they are clearly inappropriate for the contractarian project. Contractarians want to answer the question of what agreement rational agents will come to, not the question of how an impartial, and morally motivated third party will arbitrate.

The other main interpretation of bargaining theory, in fact its original one, does focus on the question of predicting what the outcome of actual bargaining will be. But given that interpretation, economists have found it problematic that axiomatic bargaining theory treats the process of bargaining essentially as a black box. Axiomatic bargaining solutions formulate conditions on what the outcome of bargaining will be like, but do not say how the parties will arrive at that outcome.

For those who think axiomatic bargaining theory is essentially about predicting what rational agents would actually agree on in bargaining, the ambition is to show that their preferred bargaining solution corresponds to the Nash equilibrium in various different non-cooperative games that model the process of bargaining. Nash himself declared this to be the goal of bargaining theory (see Nash 1953). Hence this ambition is usually referred to as the 'Nash program'. On this interpretation, what I have called the strategic approach above is in fact preferable, but it is hoped that an axiomatic bargaining solution can offer unification or a shorthand for treating many different bargaining problems. However, in the following, I want to show that the results from the strategic approach are not promising for the contractarian project.

The overwhelming popularity of the Nash bargaining solution amongst economists seems to stem from a model developed in Rubinstein (1982). It can be used to show that the Nash bargaining solution corresponds to the perfect equilibrium in an infinitely repeated offer-counteroffer game amongst rational agents (see Binmore, Rubinstein, and Wolinsky 1986). In this game, one agent starts by suggesting a division of some good. The other player can either accept, in which case the game ends, or make an alternative offer for the first player to consider, and so on. If bargaining takes time, and the bargainers discount future utility, the first player will immediately offer a division which offers the second player just enough to not continue negotiations – which, if both agents discount utility at the same rate, and if the time each round takes approaches 0, will be the Nash bargaining solution.[4]

This last qualification is important, however, and often ignored.[5] The per-

[4] There are also non-cooperative implementations of the KS solution, but they involve much more complicated procedures for bargaining. See, for instance, Moulin (1984).

[5] An unpublished manuscript by Joseph Heath made me aware of this point.

fect equilibrium in Rubinstein's model will generally be biased towards the more patient agent if the bargainers discount future utilities at different rates. So in a way, Rubinstein's bargaining game supports only what is sometimes called the *generalised* Nash bargaining solution. This solution includes parameters for each bargainer representing their bargaining advantage and can reproduce such a bias. That the outcome from bargaining must be a generalised Nash bargaining solution follows from Axioms 1-3 above (see Binmore 2007, pp. 476 ff.). Only when the symmetry axiom is added does the Nash solution (and thus a unique solution) result. We could hence say that it is the symmetry axiom in particular which is not supported by Rubinstein's bargaining game.

The dependence on patience in Rubinstein's bargaining game shows that the bargainers' equal rationality may not cancel all advantages in bargaining, as Nash (1953) suggested. And it is bad news for contractarianism: While there are no hidden moral assumptions in this reconstruction of bargaining, the result does not turn out to be impartial in a way that resembles morality. If contractarianism employed Rubinstein's prediction of what rational agents will agree on, patient individuals would turn out to be systematically advantaged in cooperative schemes. But that does not look very much like morality, and contractarianism will have failed. Moreover, a dependence on patience is a feature that is most likely shared by all bargaining games where bargaining is potentially extended in time, but can be ended by agreement. All such bargaining games thus seem badly suited for the contractarian project.

One may think that games that have a predetermined and known end point could serve the purposes of the contractarian better. Such games may either be alternating in the way that Rubinstein's is, or involve simultaneous offers. Alternating offers games with a known end point again would not help the contractarian. In these games the last agent would always accept any offer that is minimally better than the utility at the disagreement point. In that case the other agent could end up with almost everything, which is hardly impartial. The other possibility is to model bargaining as a game that ends with simultaneous offers being made. Take, for instance, a game where agents get what they claim just in case the simultaneous offers made at the end of negotiations are compatible, or otherwise walk away with nothing. This is what is sometimes called a game of pure coordination. Sugden (1990) claims that such a game of pure coordination is in fact the best interpretation of the game that contractarians like Gauthier have in mind. An analysis of this game prompts him to argue that Gauthier is not warranted in the assumption that there will be a unique outcome of bargaining. Indeed, it can be shown that in such a game, there are infinitely many Nash equilibria, namely all divisions that are Pareto efficient (see Nash 1953, Schelling 1959). This again is bad news for contractarianism. The fact that there are many Nash equilibria means that no bargaining solution in particular is supported by this analysis. Many outcomes seem compatible with the players' rationality, and the bargainers then still face the problem of how to pick an equilibrium.

Maybe, however, the axioms of axiomatic bargaining theory can be understood as expressing what guides equilibrium choice in such cases. One way this could be done is if we understood the axioms as telling us what equilibria are focal in the sense that Schelling (1960) describes. Focal points specify the strategies that appear the most natural or salient to the agents, and capture what they expect the other agents expect them to do. If this is what axiomatic bargaining theory gave us, it would again serve a role which can not be reduced to the strategic approach. The problem for contractarians here is that, given their project, they would have to give some account of why their preferred bargaining solution is focal which does not appeal to a social norm that could be understood to be moral in any way. It is sometimes claimed that what helps agents to coordinate on an equilibrium in a pure coordination game is a common understanding of what would be the fair outcome (see Binmore 2005, p.23). But of course this is the kind of focality contractarians cannot appeal to. It presupposes an idea of fairness and hence bargaining would not take place in a pre-moral context.[6]

Now imagine a truly pre-moral context where there are no norms about how to divide gains from cooperation in general. Even before considering the particular axioms that actually go into the bargaining solutions, it is hard to see why there should be any rules as to what equilibrium is focal that hold independently of the details of the case. For instance, in some cases, there may be possible equilibria which involve everybody receiving an equal amount of resources, even though, due to differing utility functions and disagreement utilities this does not correspond to any of the bargaining solutions. The simplicity of this solution may make it focal for the bargainers. Without any kind of notion of fair division, it seems that what is focal will depend on such contextual features. And then the strategy of interpreting axiomatic bargaining theory as specifying what is focal in a pre-moral context seems hopeless from the start.

We have now canvassed the standard ways of modeling bargaining as a non-cooperative game, and the result holds little promise for the contractarian project. The solutions of these games seem to fall into three categories:

[6] For some contractarians, this may not be as much of a problem. Binmore (2005) and Skyrms (1996), for instance, provide an evolutionary account of the social contract which permits for the co-evolution of norms of bargaining and the social contract itself.

(i) The solution is not impartial in the way Gauthier wants it to be. This is the case for the most well-known models of the strategic approach, such as Rubinstein's, or when a context-dependent idea of focality is used for equilibrium selection.

(ii) The solution was derived from some moral assumptions, for instance when existing fairness norms are used to predict equilibrium selection.

(iii) There is no unique solution, such as, for instance, in a simultaneous offer game of pure coordination.

All three options imply that contractarianism fails. In the first case, it fails because what results from bargaining does not resemble what we understand as moral; in the second case, it fails because we do not start from non-moral assumptions; and in the third case, it fails because there is no definite answer to the question of what rational bargainers would agree on in a pre-moral context, and thus an important element in contractarianism would be missing. So the strategic approach, like the 'fairness interpretation' of axiomatic bargaining theory, also doesn't hold much promise that bargaining theory can deliver an impartial bargaining solution derived from rationality alone. All in all, it seems like contractarians can't use standard bargaining theory to produce a unique bargaining solution without appealing either to norms or producing a partial result.

5 Symmetry and Uniqueness

The argument of the last section proceeded without discussing the actual properties of bargaining solutions. Once we look at those, more specific arguments why the Nash and KS solutions are unsuited for the contractarian project apply. It has frequently been claimed that the axioms that go into these bargaining solutions do as a matter of fact express normative commitments that go beyond the rationality of the players. In a sense, the fact that a straightforwardly normative interpretation of bargaining theory became so popular later on is testament to the normative flavour of these axioms. In particular, symmetry – which is an axiom in virtually every bargaining solution – has been claimed to express some idea of equality, and is thus an axiom that is not legitimate in a pre-moral context as contractarianism envisages it. Schelling (1959) and Harsanyi (1961) argued early on that symmetry is not justified from the point of view of rationality. Recently, Thrasher (2014) argued that this axiom should have no place in contractarianism. The argument appears to be simply that under most conceptions of what happens in bargaining, many divisions of the cooperative surplus seem to be equally possible from the point of view of rationality. In fact, any division which gives each more than the disagreement utility seems rationally defensible in many games.

That it smuggles in normative commitments has also been claimed of what is often understood to be the functional equivalent of the symmetry assumption in Gauthier's theory, namely his equal rationality assumption, and what he takes it to imply (see, for instance, Goodin 1993). Gauthier claims that assuming equal rationality implies the following:

> Since each person, as a utility-maximizer, seeks to minimize his concession, then no one can expect any other rational person to be willing to make a concession if he would not be willing to make a similar concession. (Gauthier 1986, pp. 143-144)

This principle seems to go beyond what follows from common knowledge of rationality. Suppose I am bargaining with one other person, whom I know to be rational, in a simultaneous offer game as I just described it. There seems to be nothing irrational about not being willing to accept less than 80% of the cooperative surplus. If the other person were to claim only 20%, for instance, that would be the only rational thing to do. And it is not irrational for the other person to accept 20% given that I am taking 80%. This is what it then means to maximise utility. This just repeats again that in such a game, any Pareto efficient division is a Nash equilibrium and hence consistent with common knowledge of rationality. According to critics, what Gauthier seems to be describing is the idea that it would be in some sense unfair to demand of another person to give up more than one would be willing to give up oneself. But that is the kind of normative commitment that he wanted to avoid.

Given that many outcomes appear to be equally defensible in this type of bargaining from the point of view of rationality, Thrasher argues that even if the symmetry axiom didn't smuggle in normative commitments, it would still be arbitrary. He contends that the only reason the axiom is used is to artificially create uniqueness of the bargaining solution. We have seen above that this is indeed the effect it has. Without symmetry, Nash's axioms result in the generalised Nash solution, which has free parameters and hence does not determine a unique solution. Sugden (1990) provided a similar argument. He showed that symmetry can in fact be derived from uniqueness. Uniqueness itself is essential for contractarianism. Without uniqueness, contractarianism does not work: The question of what rational agents would agree to do in a pre-moral context would have no determinate answer. And such indeterminacy in the antecedent is commonly thought of as making any counterfactual with that antecedent false (see, for instance, Hajek 2007). But such a counterfactual plays a key role in contractarianism. Now while uniqueness is essential for contractarianism in this way, Sugden claims that it cannot simply be assumed, but rather must be demonstrated.[7]

6 The Contractarian Bargaining Problem

The lessons from standard interpretations of bargaining theory and from critics seem clear: A unique bargaining solution is either partial, or will be derived from norms. The symmetry axiom, in particular, is either normative, or an arbitrary device to create uniqueness. Still, I argue that contractarianism has the resources to respond to these criticisms. In the following, I want to argue that the symmetry axiom can be defended without appeal to moral assumptions. In particular, I want to show that the criticism I described in the last section is founded on a misunderstanding of the question contractarians like Gauthier are asking. Symmetry is defensible when the question is not 'What will happen when rational agents bargain?', but rather 'What division of the cooperative surplus will rational bargainers agree on, if they manage to reach agreement?'

As an initial response to the argument of section 5, I first want to argue that contractarians should in fact hold it to be impossible to model the bargaining process with a standard non-cooperative game. And this is because any such model would presuppose some specific procedure that bargaining will follow. How could a contractarian be justified to presuppose some bargaining procedure? Either she would have to simply impose it, or she would have to provide an argument for why rational agents would choose to follow that procedure. Imposing a particular procedure for bargaining seems arbitrary given the contractarian project. We have seen that different bargaining procedures will lead to different outcomes. The contractarian wants it to be the case that everybody has a reason grounded in self-interest to comply with the agreement the theory isolates. If the contractarian imposed some specific procedure for bargaining by which one agent receives less than he would have under some other, equally arbitrary, procedure, her claim to having given everybody such a reason grounded in self-interest alone would

[7] Sugden provides a convincing argument that what he calls the 'principle of rational determinacy' is not generally true. That is, it is not true in all games that the rationality of the agents guarantees that there is a single determinate outcome to the game. However, this is still consistent with it being the case that in the bargaining game, in particular, the principle holds.

be weakened. Why should any agent have a reason to follow this hypothetical agreement, rather than the hypothetical agreement that would have been struck, had a different, perhaps more favourable bargaining procedure been imposed? It seems more in line with the contractarian project to instead leave the choice of bargaining procedure up to the agents themselves. But in this case, no bargaining procedure seems like the unique rational choice. Given that different bargaining procedures lead to different outcomes, the choice of bargaining procedure itself then becomes part of the bargaining problem. Anybody who tries to model bargaining as a noncooperative game thus has nothing to hold on to.[8]

My solution to the problems raised in sections 5 and 6 requires another assumption about the bargaining procedure. And that assumption is that bargaining is completely costless, also in terms of time. Along with several contractarians, we assume that bargaining could in principle go on indefinitely without any costs. And so nobody is under any kind of pressure to come to an agreement early.[9]

Of course, the costlessness assumption is unrealistic: In real life, prolonged bargaining will turn out to be costly in a number of ways. Still, even though this situation is unrealistic, I don't think the idea of rational choice in this situation is incoherent in the way Kraus and Coleman (1991) argue it is. They argue that without there being any cost of bargaining, there is no basis for choice anymore. Since stopping to bargain to enjoy the fruits of cooperation can't be taken to be

[8] Arguably, Gauthier's *Morals by Agreement* leaves the bargaining procedure open to the agents. He does describe bargaining as a 2-stage process, whereby agents first claim their ideal utilities, and then offer concessions until a feasible division has been reached (p.131). The process of making concessions is not well described, however. It is not clear, for instance, whether the agents see what concessions everybody else made, whether they make concessions simultaneously, whether they have to concede anything each round, whether their concessions have to become larger every round of bargaining, etc. Later on, Gauthier claims that bargaining will follow Zeuthen's principle (proposed in Zeuthen 1930), which claims that the person with the smallest relative concession has to concede and make a larger concession each round. However, this is not a further description of the structure of the bargaining game, but rather a principle that Gauthier takes to be implied by the bargainers' equal rationality. Most of what Gauthier says about bargaining in fact follows the axiomatic approach in leaving the process of bargaining unspecified. He starts off much like Nash, by imposing some conditions on the final bargaining solution (p.130). It must be a point on the Pareto frontier, and it cannot require any interpersonal comparisons of utility. And his more formal derivation of the principle of minimax relative concession focuses on the properties of various outcomes, namely on whether some division of the cooperative surplus is one that all bargainers are rationally willing to entertain.

[9] Gauthier introduces this assumption, almost as an afterthought, on p. 156 of *Morals by Agreement.* This feature of Gauthier's set-up is ignored by Sugden, whose argument that uniqueness cannot be presupposed by Gauthier relies on a reading on which the agents play a game of pure coordination with a predetermined end-point (in which case there are infinitely many Nash equilibria).

an opportunity cost of bargaining (i.e. a cost involved in foregoing some good), it seems like what is bargained over couldn't be of value to the agents. It is true that on the model I propose, delaying getting one's share from bargaining is not an opportunity cost of bargaining. But agents are still aware that they will never get the benefits if they never agree, and they can take this to be a bad outcome. So it's not the case that the bargainers have nothing to gain from bargaining.

The fact that the costlessness assumption is unrealistic does also not necessarily speak against it. After all, we are also assuming that bargaining takes place in a pre-moral context, which is also something none of us ever face. There are two reasons one might give for why a costlessness assumption is appropriate for the kind of bargaining problem contractarians set themselves. Firstly, the decision of how the benefits from cooperation are to be divided up is so important that any cost of delay through prolonged bargaining is not in proportion to what is at stake. The second justification one might give is that contractarians want to isolate the concept of a purely rational bargain. When agents ask themselves what agreement they would have struck concerning how to divide up the cooperative surplus, what they are interested in is then only what results from the agents' equal rationality, and their respective interests insofar as these concern the good to be divided up. The costs to bargaining might then be judged irrelevant to the concept of a purely rational bargain.

If the costlessness assumption does not seem well-founded, the following at least shows this: Firstly, it is possible to derive a unique and impartial bargaining solution without making normative assumptions. After all, the costlessness assumption is not a normative assumption. And secondly, since the costlessness assumption is doing most of the work in the following derivation, more attention should be paid to it, both by contractarians themselves, and by critics. The contractarian project depends on a proper defence of this assumption.

There were two reasons why in section 4, modeling the bargaining process led to predictions that are not impartial: Firstly, bargaining is modeled by a procedure which advantages some agents; And secondly, patience generally helps the bargainer. Given the costlessness assumption, and the assumption that the bargaining procedure is itself to be agreed on, neither should be an issue for the contractarian. But if we grant these assumptions, what is going to happen in bargaining? If it is unclear how bargaining will proceed, there is no clear way to model bargaining with the tools of non-cooperative game theory. And so non-cooperative game theory cannot provide an answer anymore. Intuitively, at least, it is not clear anymore whether rational agents will ever agree in such a context. There seems to be no reason why the bargainers could not go on each claiming the full cooperative surplus indefinitely. It seems like we have removed any reason for the bargainers to stop bargaining. Things seem even worse when we consider the fact that there is no predetermined procedure for bargaining which could guide the bargainers to agreement. Indeed, the fact that they will have to settle on a procedure of how to bargain may set off an infinite regress: The bargainers will have to use some procedure to decide how to proceed in bargaining. If this procedure is to be accepted by all, all need to agree on how to proceed in bargaining. But how are they going to agree on that? And so on. This regress further raises doubts that rational agents will ever manage to come to an agreement.

So now the problem is no longer only that the outcome from bargaining may be indeterminate. Now one of the possibilities of what might happen doesn't even involve any agreement. In fact there is a worry that bargaining indefinitely is what is bound to happen in this situation. Or, foreseeing that there is never a reason to stop bargaining, the agents foresee that there will never be agreement and don't even start to bargain. So even though the bargainers do have something to gain from bargaining, their rationality may stand in the way of them ever receiving those benefits.

But now I think it helps to think again about the founding motivation of contractarianism: Mutual advantage is possible if agents manage to mutually constrain the pursuit of their self-interest. Thus mutual constraint is ultimately in everybody's interest. Contractarians like Gauthier take this to imply that it must be rationally possible for a group of agents to reap those benefits. A theory of rationality which makes it rationally impossible for agents to reap these benefits can't be the right theory of rationality. In line with this reasoning, one of the assumptions in Gauthier's derivation of the principle of minimax relative concession is that every agent thinks that there is a feasible outcome that everybody would be willing to entertain - that is, agree to provided everybody else does (p.143). This implies that agreement is actually possible - for instance if the agents have the same conception of what this feasible outcome is. Gauthier defends the assumption that each thinks agreement is possible as follows:

> For cooperation to occur, all must agree on an outcome (or joint strategy) represented by a feasible point in outcome-space. Any person who supposes that there is no feasible concession point on which rational persons can agree is denying that there is any way in which those rational persons can co-operate. But each recognizes that everyone finds it rational to co-operate. (p.143)

This assumption is crucial in Gauthier's derivation, and may even be his one departure from traditional rational choice theory in the bargaining context. What he is saying here is in effect that a failure to do one's part to make agreement possible will always be irrational. However flawless an agent's reasoning who thinks that agreement is impossible, her belief is not compatible with knowledge of the other agents' rationality. Rational agents must be willing to agree to at least one division that they think all others will also be willing to agree to. There must be something wrong with any theory of rationality that told us that agents take it to be impossible to agree, and thus never bargain.

Still, this assumption does not quite solve our problem yet. Even if all take agreement to be possible, it may still also be possible that they never agree, maybe because they have different conceptions of what a feasible agreement point would be. The question of what would happen if rational agents were to bargain under the circumstances described may still have no clear answer. Now I think a closer look at what kind of question contractarianism is asking helps. Remember that bargaining theory is supposed to only solve the problem of what terms of cooperation agents should use. It is not supposed to solve the problem of how it can be rational for agents to comply with the hypothetical agreement. And so we can assume that the person who is asking the question of what terms of cooperation to use is already willing to cooperate, convinced that non-cooperation is irrational. But if that is so, the question contractarianism is asking should not be 'what would the outcome of bargaining be?'. It should be, rather, 'what division of the cooperative surplus would we have agreed on?'. One way of reading this has the inquirer assume that agreement would definitely have been found. Now given the fact that it is possible that nothing is ever agreed on, we may think that the question is misguided. But there is another way of reading the question, namely as asking 'If we had come to agreement in bargaining, what would this agreement be?' This question is the appropriate one for contractarians to ask, because the person who asks the question of what agreement hypothetical bargaining would have led to is already willing to cooperate. All she wants to know is what precise terms of cooperation to abide by. She wants to know what the rational agreement is, and thus, if sometimes bargaining between rational individuals does not lead to agreement, this is irrelevant to her concerns.

7 A 'How Possible' Story about Agreement

If what I just described is the question we are asking, the classical axiomatic bargaining solutions again become defensible. The kind of argument we will have to provide given the setup of costless bargaining and the nature of the question we are asking is a 'how possible' argument. For the reasons given above, we can no longer specify some game and solve it for its Nash equilibria. Instead, we will have to hypothesise that there are divisions of the surplus on which agreement is possible, and ask what needs to be true of these divisions in order for agreement on them to be compatible with the bargainers' rationality and common knowledge thereof. This approach is more akin to the axiomatic approach in bargaining theory, and I want to show that this procedure can provide a rational, rather than a moral justification of the contested symmetry axiom of axiomatic bargaining theory.

Before providing a more formal argument, let me stress that it seems intuitive that if there will be agreement in a context of costless bargaining, the agreement will be in some sense impartial. As we have seen, it seems puzzling how agreement could be reached at all in this context. But if agreement is in fact reached, then an impartial division seems like the only plausible candidate. Any agent who takes herself to be disadvantaged relative to another in some proposed bargain can costlessly delay agreement indefinitely, in the hope of being the advantaged party at some point in the future. So it seems like agreement could never be reached on a partial solution. It is only an impartial solution that blocks this kind of reasoning. In the following, I want to show more formally that any agreement point in the set-up we described will have to be Pareto efficient, symmetric, and thought of by all agents as unique.

Let's assume there are n agents, and a bargaining space S of feasible utility allocations to the n agents. All agents are utility maximisers, and care only about their own utility. Both the bargaining space, and the players' rationality are common knowledge. Each agent has beliefs about what allocations are feasible as agreements, that is, that she thinks all agents would be willing to agree to. Let K_i be the set of utility allocations that agent i thinks are feasible as agreements. We have said that all agents think that agreement is possible, and so K_i must be non-empty for all i. Now utility maximisation and the setup of the problem demand the following principle:

> **R:** No individual i is willing to agree to a utility allocation $u = (u_1, u_2, \ldots, u_n) \in S$ if she thinks that there is another utility allocation $u' \in K_i$ such that $u'_i > u_i$.

In other words, no agent is willing to agree to a utility allocation if she thinks that a utility allocation which leaves her better-off is also feasible. The rationale for this principle is that if an agent thinks that agreement on some utility allocation is feasible, then there will at some point be the chance for her to in fact agree to that utility allocation. Given that there is no cost to delaying agreement, she is willing to wait for that opportunity whenever she can expect to gain from it. Common knowledge of rationality implies that every agent knows that all other agents follow **R**.

We now stipulate an agreement point $a^* = (a_1^*, a_2^*, \ldots, a_n^*) \in S$. It is a utility allocation that all agents have in fact agreed to. The first thing that we can now show is that a^*, if it is a feasible allocation point, must be Pareto efficient. Intuitively, the argument is that, if a^* is acceptable to all, everybody should also be willing to agree to a Pareto improvement to a^*. But then those who would benefit from the Pareto improvement should not in fact agree to a^*, but wait for the opportunity to agree to a Pareto improvement on a^*. If a^* is a feasible agreement point, it must thus be Pareto efficient.

More formally, suppose there was a utility allocation $a' \in S$ such that a move from a^* to a' would constitute a Pareto improvement. Then there must be at least one agent h for whom it holds that $a'_h > a^*_h$, while it is true for everybody that they receive no less in a' than in a^*. This is just what it means for a' to be Pareto superior to a^*. Seeing that everybody only cares about their own pay-off, which they want to maximise, everybody who receives no less in a' than in a^* would agree to a' if they agreed to a^*. And so everybody must be willing to agree to a' if they agreed to a^*: Agreement on a' is feasible if agreement on a^* is feasible. It is known that a^* is a feasible agreement point, and so a' must also be a feasible agreement point. Given that everybody's rationality is common knowledge, h must know all this. But then, by **R**, she could not rationally agree to a^*. She knows that a utility allocation where she does better than a^* is feasible. And hence she waits for the opportunity to agree to that utility allocation. Hence a^* is not a feasible agreement point. By assuming that a^* is both a feasible agreement point and not Pareto efficient, we have derived a contradiction. And so, if a^* is in fact an agreement point, it must be Pareto efficient. Given everybody's rationality is common knowledge, everybody knows that any possible agreement point must be Pareto efficient.[10]

The next thing that can be shown is that when agreement is reached, all agents must think of that agreement point as the only feasible agreement point. Evidently, once agreement on a^* is reached, it must be true that all agents believe that a^* is a feasible agreement point, that is, for all i, $a^* \in K_i$. But suppose that a particular agent k thought that a different agreement point $a^+ \neq a^*$ was also

[10] Note the similarity to the Coase Theorem (Coase 1960).

feasible. Now if she herself did better in one of the two utility allocations than in the other, then she could not think of both of them as feasible, since she herself would only agree to the one in which she does better. Suppose that $a^+_k > a^*_k$. In this case, k would violate $\mathbf{R}$ if she thought that a^* was feasible. k should never have agreed to a^* because she thinks that she can do better. Now suppose that $a^+_k < a^*_k$. In this case, learning that a^* is a feasible agreement point, according to $\mathbf{R}$, k can no longer think that a^+ is feasible: She herself would not agree to a^+, knowing that a^* is feasible. It remains to consider the case where $a^+_k = a^*_k$. Now, given that a^* is Pareto efficient, there must be at least one agent $j \neq k$ for whom $a^+_j < a^*_j$. But given that a^* has been agreed on, j also knows that a^* is a feasible agreement point, and, by common knowledge of rationality, k knows that j knows this. But if j knows that it is possible for her to get a^*_j, by $\mathbf{R}$, she would never agree to a^+. Knowing that, k can no longer think that a^+ is a feasible agreement point. It follows that once agreement is reached, everybody thinks that the actual agreement point is the only feasible agreement point.

We have shown that once agreement is reached, every agent must think of the agreement point that was reached as the only feasible agreement point. It does not follow from this that agreement is in fact only feasible on one particular agreement point.[11] We can hence not directly invoke Sugden's argument that derives symmetry from uniqueness. This is not a problem, however, since symmetry can also be derived from the belief that any agreement that is reached must be unique. Suppose that in a symmetric bargaining problem, the agreement point a^* wasn't symmetric, so that some agent g gets more than another agent l. Let a^- be the same as a^* except that the payoffs to g and l have been exchanged. By hypothesis, $a^-_l > a^*_l$. But then, by $\mathbf{R}$, l cannot believe that a^- is feasible, given that a^* has been agreed on. But is that belief justified, given that a^* is believed to be feasible? It seems not. If it was feasible to reach agreement on a^*, then it should also be feasible to reach agreement on a^-. Given g and l are both equally rational and have the same amount of utility to win or lose from bargaining, we can assume that l would agree to a^* if and only if g would agree to a^-. But if both players are aware of this, by $\mathbf{R}$, l should never agree to a^*, and g should never agree to a^-. And thus, given a symmetric bargaining space, the agreement point cannot be asymmetric.

The argument hinges on the assumption that if agents have a reason to believe that one allocation in a symmetric bargaining problem is feasible, then they should

[11] We only characterized what the agents' beliefs have to be like when agreement is reached, but not how the convergence of beliefs on what agreement is feasible is supposed to be achieved. Uniqueness would require that beliefs can only converge on one particular utility allocation, which we have not shown.

also believe that the ‘mirror-image’ allocation is feasible. I think that this is the truth in Gauthier’s remarks on what follows from the equal rationality of the agents. With that assumption in place, we can also state the argument as follows. We note that any asymmetric solution to a symmetric bargaining problem is distinct from its ‘mirror image’ allocation. But then there will be agents who do worse under one of the allocations than under the other and vice versa. These agents would only ever agree to the one in which they do better, but not the other. And so asymmetric solutions to symmetric bargaining problems cannot be feasible.

What we have now given is a ‘how possible’ argument for the symmetry of bargaining solutions, and to some extent, their uniqueness. We have assumed, with Gauthier, that agreement is possible, and that every agent takes it to be possible. We have then asked what must be true of any stipulated agreement point, given that agreement was reached under circumstances where there is no pressure on anybody to end the bargaining process early. Under these circumstances, everybody is willing to wait out for a different agreement under which they will receive more, provided they think such an agreement is feasible. Under these assumptions, we have established that any agreement must be Pareto efficient, symmetric, and thought of by everybody as unique. While this falls short of establishing the actual uniqueness of the agreement point, all that is needed to provide a full-fledged and determinate answer to the question of what division of the cooperative surplus rational agents would agree to is a defence of either the independence of irrelevant alternatives, monotonicity, or some other axiom that establishes a unique bargaining solution. But this should be the subject of intramural discussion between contractarians. What we have given here is a derivation of the axiom that is in large part responsible for the impartiality of the different bargaining solutions from purely non-moral assumptions. And so we defended the contractarian project against the charges described in sections 4 and 5. If infighting between different but similar bargaining solutions can be resolved, we can derive a unique agreement point which is impartial from assumptions that are non-moral.

8 Conclusions

Bargaining solutions such as the Nash or KS solutions are now mostly interpreted as formalising the notion of fairness in cooperation, and justified by testing them against our intuitions about fairness. Since contractarians need their bargaining solution to pin down what the only rational outcome in bargaining is without making any moral assumptions, their proposed bargaining solution requires a different kind of defence. That it may be difficult to provide such a defence became apparent in the first part of this paper: all standard non-cooperative games that have been used to model bargaining between rational individuals either produce no unique result, require normative assumptions to be solved, or have solutions that are not impartial enough to be the basis for a contractarian theory of justice.

This paper offered a defence of the contractarian project against this criticism. The argument relied on two key preliminary steps. The first is the assumption that hypothetical bargaining is open-ended: not only can the agents themselves decide what procedure to use in bargaining; bargaining is also completely costless so that agents have no problem continuing the bargaining process indefinitely. The second is that if we take it as a given that it must be rationally possible for us to cooperate in some form, contractarians can disregard the possibility that agents never reach agreement in the hypothetical bargaining process. Contractarians should be interested in what will be agreed on, given there is agreement. And so they can simply stipulate an agreement point. But if we can stipulate an agreement point, we can establish that any such agreement point must be symmetric, Pareto efficient, and thought of by the agents as unique. And thus the most normatively loaded axioms of the standard bargaining solutions have been derived from non-moral assumptions. The contractarian needs to show that the result of the agreement of rational and self-interested individuals will be impartial moral rules. The argument in this paper goes a long way towards showing that this can be done.

References

Binmore, K. (2005). *Natural Justice.* Oxford: Oxford University Press.

Binmore, K. (2007). *Playing for Real.* Oxford: Oxford University Press.

Binmore, K., A. Rubinstein, and A. Wolinsky (1986). The Nash Bargaining Solution in Economic Modelling. *The RAND Journal of Economics* 17 (2), 176-188.

Braithwaite, R. (1955). *Theory of Games as a Tool for the Moral Philosopher.* Cambridge: Cambridge University Press.

Coase, R. (1960). The Problem of Social Cost. *Journal of Law and Economics* 3 (1), 1-44.

Gaertner, W. (2006). *A Primer in Social Choice Theory.* Oxford: Oxford University Press.

Gauthier, D. (1969). *The Logic of Leviathan.* Oxford: Clarendon Press

Gauthier, D. (1986). *Morals by Agreement.* Oxford: Oxford University Press.

Gauthier, D. (1993). Uniting Separate Persons. In D. Gauthier, and R. Sugden (Eds.), *Rationality, Justice and the Social Contract.* Ann Arbor: University of Michigan Press.

Goodin, R. (1993). Equal Rationality and Initial Endowments. In D. Gauthier, and R. Sugden (Eds.), *Rationality, Justice and the Social Contract.* Ann Arbor: University of Michigan Press.

Hajek, A. (2007). Most Counterfactuals are False. unpublished manuscript, URL= http://philrsss.anu.edu.au/people-defaults/alanh/papers/MCF.pdf.

Harsanyi, J. (1961). On the Rationality Postulates Underlying the Theory of Cooperative Games. *Journal of Conflict Resolution* 5, 179-196.

Hobbes, T. (1651/2010). *Leviathan: Or the Matter, Forme, and Power of a Common-Wealth Ecclesiasticall and Civill.* Ed. by Ian Shapiro, New Haven: Yale University Press.

Kalai, E., and M. Smorodinsky (1975). Other solutions to Nash's Bargaining Problem. *Econometrica* 43 (3), 513-518.

Kraus, J., and J. Coleman (1991). Morality and the Theory of Rational Choice. In P. Vallentyne (Ed.), *Contractarianism and Rational Choice.* Cambridge: Cambridge University Press.

Lensberg, T. (1988). Stability and the Nash Solution. *Journal of Economic Theory* 45 (2), 330-341.

Moulin, H. (1984). Implementing the KalaiSmorodinsky Bargaining Solution. *Journal of Economic Theory* 33, 32-45.

Nash, J. (1950). The Bargaining Problem. *Econometrica* 18, 155-162.

Nash, J. (1953). Two-Person Cooperative Games. *Econometrica* 21, 128-140.

Pfingsten, A., and A. Wagener (2003). Bargaining Solutions as Social Compromises. *Theory and Decision* 55 (4), 359-389.

Rubinstein, A. (1982). Perfect Equilibrium in a Bargaining Model. *Econometrica* 50, 97-110.

Schelling, T. (1959). For the Abandonment of Symmetry in Game Theory. *The Review of Economics and Statistics* 41, 213 - 224.

Schelling, T. (1960). *The Strategy of Conflict*. Cambridge: Harvard University Press.

Skyrms, B. (1996). *Evolution of the Social Contract*. Cambridge: Cambridge University Press.

Sugden, R. (1990). Contractarianism and Norms. *Ethics* 100, 768-86.

Thrasher, J. (2014). Uniqueness and Symmetry in Bargaining Theories of Justice. *Philosophical Studies* 167 (3), 683-699.

Zeuthen, F. (1930). *Problems of Monopoly and Economic Welfare*. London: Routledge.