

Reciprocity and the Case of Disability in Contractarian Theories of Justice

Name: Yunmeng Cai

Email: yunmeng.cai@nuffield.ox.ac.uk

Institutional Affiliation: DPhil candidate in Nuffield College, University of Oxford
Visiting scholar in Yale University

Abstract:

Contractarian theories of distributive justice often appeal to the notion of reciprocity to determine fair terms of social cooperation. Many also consider reciprocity to be an indispensable idea in theories of justice. However, the demand of reciprocity seems to imply that those unable to reciprocate, such as people with severe physical or mental disabilities, have to be excluded from considerations of justice.

In this paper, I argue that such tension results from confusion between two different roles the idea of reciprocity plays in generating principles of justice—the principle of reciprocity as a demand of justice and the norm of reciprocity as a feasibility constraint of justice. While reciprocity in these two senses can be aligned in the standard case where it guides cooperation among fully functioning contemporaries in a society, different implications start to emerge in special cases such as disability and intergenerational justice. In such cases, the question of how deviation from the norm of reciprocity may impact feasibility needs to be addressed.

I will then appeal to sociological game theory models which explain the evolution of social norms to examine the role of reciprocity as a feasibility constraint. I argue that, prevailing norms of reciprocity merely represent a soft feasibility constraint which may be overcome through proper institutional designs. With the establishment of risk sharing institutions, a more risk-concerned norm of reciprocity may be fostered, which allows a contractarian theory of justice to include people with severe disabilities while successfully addressing reciprocity-based concerns.

Keywords:

reciprocity; disability; contractarian theory of justice; social norms; risk-sharing

Reciprocity and the Case of Disability in Contractarian Theories of Justice

Contractarian theories often appeal to the notion of reciprocity in generating principles of justice. Many also consider reciprocity to be an indispensable idea in theories of justice. Barry, for example, suggested that “any theory of justice that tried to eliminate reciprocity would be doomed from the start.”¹ On the other hand, however, the demand of reciprocity seems to imply that those unable to reciprocate, such as people with severe physical or mental disabilities, have to be excluded from considerations of justice.² Responses to this concern range from the abandonment of the contract device, thus also dropping the idea of reciprocity,³ to the denial of duties of justice owed to such people⁴. For those who prefer to cover disability within the contract, some suggest that, while reciprocity is a demand of justice, it is not its only demand. Hence people with disabilities can be accommodated by other requirements of justice.⁵ However, it seems that such responses simply fail to address some essential concern of those who insist on the indispensability of reciprocity in a theory of justice.

¹ B. Barry, “Justice as Reciprocity”, in *Liberty and Justice: Essays in Political Theory 2* (Oxford: Clarendon Press, 1991), p. 235.

² I recognise that physical and cognitive impairments do not necessarily render a person incapable of participating in social cooperation and, in addition to his own impairments, he may partly be disabled by external social arrangements. In fact, I endorse this social model of disability by recognising disability as at least partly a socially imposed risk. But I do not intend to offer a detailed account of disability in this chapter. When I use the term “people with disabilities”, I only refer to it in a technical sense to cover those who have no capacity or prospect of reciprocation in any meaningful sense, thus being ineligible to be included in the scope of social cooperation according to social contract theories.

³ Nussbaum, for example, suggested to replace the contract doctrine with the capabilities approach. However, she acknowledged that, by dropping reciprocity as mutual advantage, such a theory would then need an account of “a more complex political psychology that makes it clear how a cooperation that is not understood in terms of mutual advantage alone will be sustained.” See M. Nussbaum, *Frontiers of Justice: Disability, Nationality, Species Membership* (Cambridge, Massachusetts: The Belknap Press of Harvard University Press, 2006), p. 153.

⁴ For example, some believe that people with severe disabilities are owed with humanitarian duties only.

⁵ Barry suggested that, while reciprocity regulates exchange activities, it says nothing about the initial distribution of resources. Therefore people with disabilities can be accommodated by a principle of equal access to resources or a principle of equal opportunities. Stark similarly proposed that people with disabilities can be taken care of by a separate need-based principle. See Barry, “Justice as Reciprocity”, pp. 235-8; C. A. Stark, “How to Include the Severely Disabled in a Contractarian Theory of Justice”, *The Journal of Political Philosophy* 15 (2007), 127-45.

In this paper I want to examine if the exclusion of people with disabilities is an inevitable implication of the contract doctrine. I will first look at contractarians' reasoning concerning the issue of reciprocity and disability. I suggest that we need to distinguish between two notions of reciprocity in their arguments, that is, reciprocity as a demand of justice and reciprocity as a feasibility constraint for theories of justice. I will argue that reciprocity as a demand of justice, when interpreted properly, can actually accommodate people with disabilities. However, such demands may be constrained by issues of stability. Rawls, for example, seems to believe that reciprocity as a norm forms such a feasibility constraint and therefore requires people with disabilities to be excluded from the initial contracting position.

So, in order to include people with disabilities in a social contract, the challenge is to reconcile their lack of prospect to reciprocate with the theory's reliance on reciprocal norms to make stable compliance possible. If a more generous norm of reciprocity than what is currently prevailing can be fostered, then the feasibility constraint may be overcome. I will finally examine this possibility by using sociological game theory models which aim to explain how social norms are evolved and realised.

Reciprocity as a Demand of Justice

In order to examine whether reciprocity-based exclusion is an inevitable consequence of the contract doctrine, what first needs to be understood is the actual requirement implied in the notion of reciprocity. Reciprocity is originally used to describe the practice of exchanging things with others for mutual benefit, thus reciprocity as mutual advantage. In this original context of bilateral exchange, the requirement of basing an act of exchange on reciprocity may be interpreted in several different ways. First, it may mean that both parties should be made better off through the exchange. Second, it may imply that participating in this exchange should

represent a rational choice for both parties. Third, it may sometimes also imply that both of them can regard the result of exchange as fair, that neither of them has exploited the other. Distinguishing between these interpretations is less relevant in this bilateral context because a voluntary exchange is likely to satisfy the three demands, especially the first two, simultaneously. However, when applied to a scheme of social cooperation, it becomes more relevant which interpretation is the actual requirement of reciprocity that constitutes a necessary condition of just social cooperation.

According to the first proposed interpretation, reciprocity as mutual advantage may simply be taken to mean that every participant in a scheme of social cooperation is better off compared to no social cooperation at all. Rawls, for example, often refers to society as a cooperative venture for mutual advantage and what he means is simply that “social cooperation makes possible a better life for all than any would have if each were to try to live solely by his own efforts.”⁶ If this is what reciprocity demands, then including people with disabilities in the scheme and sharing with them some fruits of social cooperation is unlikely to undermine the mutually advantageous result compared to the state of nature. But this does not seem to be what people have in mind when they argue that including those unable to contribute to social cooperation violates the idea of reciprocity.

If what reciprocity requires is rational cooperation according to the second interpretation, then, in the context of a multi-person scheme, it would demand every participant’s cooperation with every other participant to represent a mutually advantageous, thus rational, choice. This seems to imply that those with disabilities cannot be included in the cooperative scheme if they are incapable of making meaningful productive contributions. However, this conclusion assumes that production is the only form of social cooperation. As I have mentioned in the introduction,

⁶ J. Rawls, *A Theory of Justice* (Cambridge, Massachusetts: The Belknap Press of Harvard University Press, 1999), p. 109.

forms of social cooperation may include not only production but also more complicated mechanisms such as risk sharing. When a risk sharing mechanism is included in a scheme of social cooperation, it can be rational for participants to include those temporarily suffering disabilities. Barry calls this more complicated form of reciprocity “reciprocity as mutual aid”.⁷ As he explained, in the cooperation of mutual aid (which is one form of risk sharing), while those who contribute in a given period do not necessarily expect to claim repayment from the present beneficiaries, they expect to be reciprocated in the future when they need similar aid. However, reciprocity as mutual aid extends only to those who have a reasonable prospect of being able to reciprocate in the future. As acknowledged by Barry, it cannot cover those who are “chronic bad risks”.⁸ This means that, with the demand of reciprocity interpreted as rational cooperation, people who suffer severe and permanent disabilities have to be excluded even when risk sharing is incorporated in the scheme of social cooperation.

However, reciprocity as rational cooperation is too demanding an interpretation. If it is what reciprocity requires, few theories of justice would be able to meet the demand. Some may suggest that Gauthier’s theory of rational contractarianism implies reciprocity in this sense.⁹ But even Gauthier requires bargainers in his contracting position to accept what he calls the principle of minimax relative concession. By referring to the acceptance of equal proportional sacrifice as a kind of “practical rationality”,¹⁰ he is obviously referring to it more as a common sense of fairness than as a rational choice in egoist sense.

So, finally, what reciprocity requires may be a certain sense of fairness instead of pure rational cooperation. Indeed, reciprocity would be a redundant concept if what it demands is simply

⁷ As opposed to what he calls “reciprocity as requital”, which is the simpler and more direct form of reciprocity. See Barry, “Justice as Reciprocity”, p. 220.

⁸ Ibid., p. 230.

⁹ D. Gauthier, *Morals by Agreement* (Oxford: Clarendon Press, 1986).

¹⁰ Ibid., p. 22.

rational choice. White, for example, endorses this interpretation in formulating his account of fair-dues conception of reciprocity, which requires that “where the institutions governing economic life are otherwise sufficiently just, those who claim the generous share of the social product available to them under these institutions...make a decent productive contribution, suitably proportioned and fitting to ability and circumstances, to the community in return.”¹¹ White regards this notion of reciprocity to be grounded in the ideal of mutual regard¹² and, since he is not adopting the contract approach, he does not think it problematic for the notion to cover people with disabilities.

Rawls, on the other hand, seems to similarly endorse this fairness interpretation. He regards his difference principle, chosen by parties in the Original Position where they are deprived of knowledge about their social positions, personal attributes and preferences, to be a principle of reciprocity grounded in the ideal of equality.¹³ If equality is his ultimate concern in the demand of reciprocity, then he should be able to accommodate people with disabilities within his contract device. In fact, by putting parties in the Original Position behind the Veil of Ignorance, Rawls’s contract almost points naturally toward the possibility of including people with disabilities. If the Veil is introduced to prevent arbitrary distinctions of natural assets and abilities, of intelligence and strength, from affecting the choice of cooperative terms, then it seems whether a person possess normal physical or cognitive capacities is simply one of such arbitrary factors the knowledge of which should be excluded from the Original Position. However, Rawls forestalls such extension by committing himself at the same time to Hume’s

¹¹ S. White, *The Civic Minimum* (Oxford: Oxford University Press, 2003), p. 59.

¹² *Ibid.*

¹³ Rawls, *A Theory of Justice*, p. 88.

account of circumstances of justice.¹⁴ In *A Theory of Justice*, he stated that parties know that their physical and psychological capacities are “within the normal range,” that they “are full and active participants in society and directly or indirectly associate together over the whole span of their life,”¹⁵ thus implying that people who are incapable of full participation have to be excluded from the contracting position at least in the initial stage. Such exclusion is then made explicit in *Political Liberalism* where he said,

By taking this as the fundamental question [of how to specify the terms of social cooperation between citizens regarded as free and equal, and as normal and fully cooperating members of society over a complete life] we do not mean to say, of course, that no one ever suffers from illness and accident; such misfortunes are to be expected in the ordinary course of life, and provision for these contingencies must be made. But, given our aim, I put aside for the time being these temporary disabilities and also permanent disabilities or mental disorders so severe as to prevent people from being cooperating members of society in the usual sense.¹⁶

Some people believe that this tension is a result of Rawls’s attempt of bringing together two potentially contradictory ideas into his contract—the Kantian idea of impartiality and the Humean idea of justice as rational cooperation.¹⁷ According to this view, on the one hand, Rawls constructed the Veil of Ignorance to make the differences in capacities irrelevant for

¹⁴ Some philosophers, such as Kittay and Sen, seem to ascribe the exclusion of people with disabilities to Rawls’s wrong or incomplete list of primary goods. Kittay, for example, suggested that people with disabilities and care workers can be properly included if care is added as a primary good. See E. Kittay, *Love’s Labor: Essays on Women, Equality and Dependency* (London: Routledge, 1999), pp. 100-14. Sen, on the other hand, suggested to replace the list of primary goods with a list of capabilities. See A. Sen, “Equality of What?”, in *Tanner Lectures on Human Values*, ed. S. McMurrin (Salt Lake City: University of Utah Press, 1980). These approaches, however, do not address the fundamental problem of including people with disabilities, which results from the requirement of the circumstances of justice.

¹⁵ Rawls, *A Theory of Justice*, pp. 83-4.

¹⁶ J. Rawls, *Political Liberalism* (New York: Columbia University Press, 1993), p. 20.

¹⁷ See Barry, “Circumstances of Justice and Future Generations”, p. 228; Nussbaum, *Frontiers of Justice*, p. 57.

considerations of justice, thus satisfying the Kantian ideal; on the other hand, he accepted the circumstances of justice, which excludes people based on their lack of capacities to contribute, so that the ideal of rational cooperation is retained. However, it is not clear if Rawls actually wanted to endorse the latter. In fact, he acknowledged that, after the Veil is lifted, what guarantees support for the chosen principles of justice is not rational choice but a sense of justice, with which parties can “rely on each other to understand and to act in accordance with whatever principles are finally agreed to.”¹⁸

Nevertheless, in discussing the basis of the sense of justice, Rawls did appeal to a notion of reciprocity that is different from what has been discussed so far. In distinguishing the three possible requirements of reciprocity, I have always treated it as an inherent demand of just social cooperation. It is also in this sense that Rawls suggested that the difference principle is a principle of reciprocity. But when he related the sense of justice to reciprocity, he was referring to reciprocity as a norm which is deeply rooted in moral psychology. In the next section, I will show that this notion of reciprocity represents a feasibility constraint for Rawls. What causes the tension in his account is not two different interpretations of reciprocity as contradicting demands of justice but two different roles reciprocity plays in the theory. Rawls committed himself to the circumstances of justice not because he endorsed justice as rational cooperation but because he relied on its requirements to guarantee that the reciprocity-related feasibility constraint is not violated, which is the real concern that needs to be addressed in order to include people with disabilities into the social contract. I will argue that, while Rawls’s commitment to equal ability is sufficient in achieving the purpose, it is not necessary. Thus there exists a possibility of including people with disabilities into the contract without violating reciprocity-based concerns.

Reciprocity as a Feasibility Constraint of Justice

¹⁸ Rawls, *A Theory of Justice*, p. 125.

Rawls referred to reciprocity as a norm when he addressed the problem of stability. Rawls emphasised that principles of justice need to be able to “generate its own support”.¹⁹ That is, the chosen principles should be psychologically suited to human inclinations so that stable support can be expected. According to him, “however attractive a conception of justice might be on other grounds, it is seriously defective if the principles of moral psychology are such that it fails to engender in human beings the requisite desire to act upon it.”²⁰ Since he takes human beings’ common psychological inclination to be a reciprocal tendency to “answer in kind”, he requires the demands of his principles of justice to be aligned with such a norm of reciprocation.²¹

The contract device seems to be a natural solution to this concern of stability. Since it is constructed to strike an appropriate balance between altruism and the claims of self, the contract device could naturally incorporate the notion of reciprocity which reflects reconciliation of the two competing tendencies.²² What matters, then, is to make sure that the resulting balance derived from the contract is one that could actually be supported by the prevailing norm of reciprocity. When the demands of justice is aligned with the prevailing norm of reciprocation, stable cooperation can be secured through norm conformity rather than constantly relying on coercion, thus satisfying the requirement of stability.

In this sense, the existing social norm of reciprocity represents a feasibility constraint for Rawls. It is this concern of feasibility that explains his insistence on the circumstances of justice. Circumstances of justice require participants to have similar physical and mental powers to participate in social cooperation.²³ In this way it guarantees the prospect of reciprocation and thus the satisfaction of the feasibility constraint. But this does mean that it is the only solution.

¹⁹ Ibid., p. 399.

²⁰ Ibid., p. 398.

²¹ Ibid., p. 433.

²² Rawls, *A Theory of Justice*, p. 440.

²³ Ibid., pp. 109-10.

To suppose that equal or similar ability as required by circumstances of justice is a necessary condition of reciprocation is to assume that it is impossible for the norm of reciprocity to apply to interactions among people with greatly different cooperating capacities. What is implicit in this assumption, moreover, is a view that this norm of reciprocity is both universal and static. If this is not the case, that is, if the current norm of reciprocity could in fact accommodate social interactions among people with different contributing capacities, or if a more generous norm of reciprocity than what is currently prevailing can be fostered, then the requirement of equal ability may be dropped without violating the feasibility constraint.

Therefore, requiring the demands of justice to be aligned with the norm of reciprocity does not by itself foreclose the possibility of extending the scope of cooperation to non-fully cooperating members. Whether people with disabilities can be included without undermining a cooperative scheme's long-term stability depends on the norm of reciprocity, either actual or potential, in the context of that scheme. Rawls did not specify what he had in mind about this reciprocal norm other than describing it generally as the demand to answer in kind. Put more formally by Becker, reciprocity as a norm is "a matter of making a fitting and proportional return for the good or ill we receive".²⁴ While people may generally agree that this is what is demanded by the norm, there may be different interpretations of what counts as "fitting" and "proportional". According to Becker, these different conceptions may compete to turn the general idea of reciprocity into more determinative social norms and standards.²⁵ Society members may subscribe to a more strict version of reciprocity which requires return to be in kind and benefit to be equal in value. Alternatively they may conform to a more generous notion of reciprocity which accepts return

²⁴ L. Becker, "Reciprocity, Justice, and Disability", *Ethics* 116 (2005), 9-39, p. 18.

²⁵ *Ibid.*, p. 18.

for the good to represent equal marginal sacrifice and return for the bad to be directed toward restoring and sustaining reciprocal relationships.²⁶

The way Rawls appealed to reciprocity as a norm seems to suggest that what he had in mind lies somewhere in between. On the one hand, by acknowledging the difference principle, which grants more benefits to the worst off than their contributions, to be an acceptable principle of reciprocity, his seems to allow deviation from the strict demand of equal benefit. On the other hand, by excluding people with disabilities from the contract, he does not seem to consider the norm to be so generous as being concerned with restoring reciprocal capacity when it is lacking, and what counts as “sacrifice” for him seems to be limited to productive efforts. In fact, it seems that his notion of reciprocity is one that is specifically adapted to the classic context of social cooperation in production and market exchange. When looking beyond this context, as observed by Becker, people may exhibit the willingness to accept a more generous conception of reciprocity in certain occasions which allows the less capable to fulfil its demands.

For example, instead of expecting reciprocation of equal value, parents would happily accept a thank you card from their kids in return for their love. People would also readily accept a kiss, a gift or some not-in-kind return for taking care of their injured family members or friends. Proportionality still matters but can be sufficiently satisfied when people know the return reflects appropriate efforts, equal marginal sacrifice in Becker’s term, from the reciprocator. On the other hand, when a bad is received, if it is a case of negligent wrongdoings, instead of retaliation, people may accept sincere apologies and corrections where possible in return for the harms they receive. If it is because the other party fails to reciprocate their good, they will consider if it is due to their lack of reciprocating capacity and, if so, may support the restoration

²⁶ Ibid., pp. 20-31.

of their capacity before demanding repayment. This scenario applies well, again, to the parent and child relationship, as well as the relationship between the ill or injured and care givers.

Becker suggests that this more generous norm of reciprocity should be extended to a larger scope of social cooperation so that temporarily and permanently disabled people could be included without failing the demand of reciprocity. He believes that this proposal is likely to impress what he calls the “tough crowd”—people coming from different political backgrounds, with different moral ideals and religious beliefs, and so on—because all these bargainers are aware that they themselves are also vulnerable to disease, injury and accidents.²⁷ However, while everyone may indeed agree that it would be desirable to have the prevalence of a social norm which reflects this more generous view of reciprocity, a norm is not something that can just be chosen by contracting parties and thus be realised. In addition to being rules which guide people’s actions, social norms are also rules internalised by people and become, in some sense, their moral preferences and dispositions. It is, in other words, not something that can simply be enforced. It is in this sense that Rawls takes reciprocity to be a feasibility constraint which makes cooperation according to principles of justice sustainable.

In fact, Rawls seems to believe that the existing psychological tendency of reciprocation is a result of human evolution. He wrote that “beings with a different psychology either have never existed or must soon have disappeared in the course of evolution.”²⁸ If one sides with this evolutionary view, it seems unrealistic to expect people to move beyond what is firmly fixed in their mind. Indeed, while human beings do have the kind of more generous reciprocal relationships described by Becker with their family members, lovers and friends, such a norm has never prevailed on a large scale among strangers. If this is the result of human evolution and

²⁷ Ibid., pp. 16-8.

²⁸ Rawls, *A Theory of Justice*, p. 433.

history, then it also seems infeasible to build socioeconomic institutions presupposing this non-existing norm while expecting them to receive stable support.

However, to agree that social norms evolve out of human history and have a psychological root, one does not need to also commit to the view that they are fixed. In fact, norms often change over time along with the changes in political, social and economic conditions. In fact, as suggested by Rothstein, social norms can be shaped by social institutions in which people participate and form an understanding and expectation of others.²⁹ This seems to suggest that we can probably influence a norm, or “realise” a norm, through proper institutional designs. But, on the other hand, if we accept this possibility, does it mean that we can potentially realise any norm we desire through institutional designs? Or does it turn out that some norms are more feasible to realise than others? To address these doubts, we need to look into empirical studies concerning how social norms in the real world have evolved and stabilised.

In fact, this is a widely researched topic in sociology. To answer questions like “What are the structural features of cooperative situations that generate a ‘demand’ for social norms?” “What mechanisms contribute to the enforcement, or ‘effective realisation’ of norms?”, sociologists often appeal to explanations offered by game theory models.³⁰ They find that empirical observations of social norms often correspond well to what is predicted by static or evolutionary game theory. Such models can also successfully explain the results of simulative laboratory experiments. According to game theory models, a social norm could be effectively realised if it represents the Nash Equilibrium strategy in certain repeated games, that is, if it is a strategy no rational player in the game would have an incentive to deviate from once it is commonly adopted. Therefore, if social cooperation can be constructed as a repeated game, sociological

²⁹ B. Rothstein, *Just Institutions Matter: The Moral and Political Logic of the Universal Welfare State* (Cambridge: Cambridge University Press, 1998), p. 139.

³⁰ T. Voss, “Game-Theoretical Perspectives on the Emergence of Social Norms”, in *Social Norms*, ed. M. Hechter and K. App (New York: Russell Sage, 2001), p. 105.

theory may help us to understand whether a more generous norm of reciprocity, such as the version proposed by Becker, could be effectively realised. I will therefore turn to the game theory model and examine how it may inform us about the feasibility of realising a more generous norm of reciprocity.

Reciprocity as a Strategy of Conditional Cooperation in Repeated Games

The provision of public goods can be structured as a game that resembles the typical game of a multi-player prisoners' dilemma. In the public good game, the rational choice for every participant is to defect, though collectively they would be better off if all choose to cooperate. In this one-off public good game, defection is the unique Nash Equilibrium (NE) strategy. That is, universal defection is a steady state in which no player has the rational incentive to unilaterally deviate from the present strategy. One solution to this collective action problem is to introduce third parties to enforce cooperation. The threat of sanction increases the cost of defection and transforms universal cooperation into a Pareto-optimal NE. However, such an external sanction system is not cost-free. The joint provision of the system would itself introduce a new collective action problem that resembles the original dilemma.

An alternative solution is to have the game repeated indefinitely. In a repeated game, while universal defection in all rounds is still a NE strategy, it is no longer the only NE. With multiple rounds of choices, a player can condition his action in a given round on other players' previous actions. For example, he may adopt a conditional cooperation strategy called the grim trigger strategy, which states that he will start the game by choosing cooperation but, once a player defects in a single round, he will then defect on him in every subsequent round. When this strategy is adopted by all other players, a defector will effectively be excluded from all future benefits of the scheme. Players' commitment to this trigger strategy may form a credible threat

of sanction, thereby increasing the cost of defection and transform cooperation into the rational choice for every player. Moreover, for any player, adopting the grim trigger strategy is the rational choice if all other players are committed to the same strategy. Therefore, universal adoption of the strategy is a NE that is Pareto-superior to the NE of universal defection in all rounds.

In the real world, through the iteration of such a game, the NE strategy leading to universal cooperation may be internalised by people as a moral preference and become embedded in their disposition. That is, compliance becomes the social norm. Once it becomes the social norm, deviating from it may yield sanction in the form of shame or guilt, which has the similar effect of the threat of actual retaliation. This norm of conditional cooperation is essentially the norm of strict reciprocity described earlier. It depicts the reciprocal relationship between participants which requires repayment of others' efforts in making contributions and sanctioning attitude toward those who fail to reciprocate.

By looking at the norm of reciprocity as a NE strategy in a repeated game, the question—can we possibly realise a norm through institutional designs?—then becomes—can we structure the game in a way so that a more generous cooperation strategy would become a NE strategy of that game? In the next section, I will suggest that when the possibility of becoming disabled and a mechanism to share the burden of such a risk are incorporated into the public good game, then a more generous norm, which I call the risk-embedded reciprocity, would become a NE strategy. This is a norm that could accommodate people with disabilities without punishing them for their failure to contribute.

The Risk-Embedded Public Good Game

I have so far assumed that all players have an equal ability to contribute. Now, if a player has only limited contributive capacity, so that including him in the cooperative scheme represents a net drain for other players, it seems clear that no rational cooperation could be sustained with him. Such a player could not form an effective threat of sanction toward those attempting to defect on him because they have no net benefits of cooperation to forego by doing so. This result based on formal analysis is consistent with social contract theorists' reason for excluding people with disabilities due to their lack of prospect to reciprocate. However, this analysis makes a crucial assumption that every player's ability to contribute is fixed throughout the game. That is, they are never subject to the risk of losing their capacity to contribute. If the possibility of becoming disabled, either temporarily or permanently, is taken into consideration, it may change the NE strategy for each player in the public good game. I now turn to a special risk-embedded game which I call the risk sharing game.³¹

In this game, n players cooperate to maintain a risk sharing fund by making a fixed contribution in every round. This fund will be used to cover various risk events occurring for all players. In each round i , $p(i)\%$ of the group will find themselves disabled and therefore lack the ability to contribute in that period. This does not need to mean that every player is faced with the same risk of $p(i)\%$ of falling or staying disabled. It suffices for each player to know that his individual risk is non-zero. A player effectively suffers permanent disability if he finds himself in this disabled group in every round of the game. If he only falls into this group in a few rounds, he experiences temporary disability.

Surely, when a player is disabled, he cannot recover without some external resources (including medical and care service), so I further assume that the risk sharing fund can be used to restore

³¹ For a game with a similar structure, see an example called "the mutual-aid game" provided by R. Sugden in *The Economics of Rights, Co-operation and Welfare* (Oxford: Basil Blackwell, 1986), pp. 122-8.

the capacity of the disabled players, provided, of course, that they are not excluded from the game. There is no guarantee that a player, with the necessary resources, will recover in the next round, but assume that at least some will, so that $p(i)$ remains relatively stable throughout the game.³²

Now consider the cooperative strategy of this game. A strict exclusion trigger strategy may state that, if a player fails to contribute in a single round, regardless of whether he intentionally defects or he is unfortunately disabled, he will be excluded from all future rounds of the game. This strategy, however, will not be agreed upon by any rational player as everyone is aware of the possibility that he may one day become disabled and thus be excluded. Similarly, an alternative strategy which permanently excludes those who defect in any round but only temporarily excludes those who are disabled is not viable either because, by denying resources to those who are disabled, they are unlikely to survive, let alone to recover, on their own. So this strategy is effectively equivalent to permanent exclusion of disabled people.

It is clear then that rational players, knowing they always face a non-zero risk of disability, would adopt a conditional cooperation strategy that takes a form similar to the following—players will start the game by choosing cooperation; only when a player is capable of cooperation in any round but chooses to defect will he be permanently excluded from the game in the future. Common employment of this strategy would lead to a NE of universal cooperation in each round of the game. This cooperative strategy is, in fact, very similar to the kind of generous reciprocity Becker is arguing for. It effectively requires building or restoring people's capacity of reciprocation before demanding their reciprocal contribution. I call the kind of

³² If most players cannot recover once they fall disabled, the disabled group will grow larger in each period, leading to the collapse of the cooperative scheme. This may in fact be true in most of human history, which might explain the long history of exclusion of people with disabilities from considerations of justice. But in contemporary societies technology has advanced to an extent that many disabilities can be treated.

reciprocity implied by this NE strategy risk-embedded reciprocity. Risk-embedded reciprocity allows people with disabilities to be included in the cooperative scheme and exempts them from making a contribution when they are not able to.

It may be objected that, while those suffering temporary disability can potentially recover and contribute to the cooperative venture in the future, the permanently disabled is likely to have no chance of reciprocating at all. However, first of all, in any round of the game, no player can tell who would be permanently disabled and thereby exclude them from the game. This may seem different from reality where people know that people suffering certain disabilities are unlikely to recover during their life course. But, even in the real world, with significant medical and technological advancement taking place every day, many disabled people do still have the potential to recover. Moreover, even setting such possibilities aside, from the rational choice perspective, it does not matter even if a player is likely to be permanently disabled. Since every player in the game faces the same risk, however slim, of becoming permanently disabled, they would not rationally adopt a strategy which requires such exclusion.

Another worry is concerned that the disabled players in each round have no power to form a credible threat against the potential defection of an able player. However, the lack of direct reciprocity between the able and the disabled players does not matter here because every player, including those with full capacity in the current round, has an interest in sustaining cooperation in the long term. A player who is currently able knows that one day he may become disabled and would at that time need the fund to meet his own basic needs and to restore his own capacity, so he would be rationally incentivised to uphold this strategy and to make a credible threat of sanction toward all other able players to prevent them from defecting on the disabled group.

This risk sharing game can be part of a larger game of social cooperation while yielding the same implication of the inclusion of people with disabilities. As long as every player is aware of the permanent possibility of suffering disability himself and part of the social production resulting from cooperation is used to meet basic needs and to restore capacity for the disabled group, the equilibrium strategy for the game would always be to share the risk of disability and to never exclude the disabled players from the game.

A different set of objections may be raised not to the game theory model itself but to applying formal analysis to the issue of disability. The first objection is that, by basing the analysis on rational choice theory, it precludes the inclusion of those suffering cognitive impairments or mental disorders from the contract. This would be a mere confusion. The formal analysis is only used to explain how a social norm might emerge and stabilise. What is sufficient to sustain real-world institutions presupposing the more inclusive notion of reciprocity is the prevalence of the corresponding norm, not the actual rational decisions made by every member of the cooperative scheme. By predicting the possibility of achieving this more generous norm, the force of the model lies not in its use of the rational choice theory but in its success in using the theory to explain empirical observations as well as the results of behavioural experiments conducted by sociologists.

The second objection is concerned again with permanent disability, though stated from a slightly different angle from the one raised within the model. The objection states that people with permanent disabilities will turn out to be, as a matter of fact, free riders of the scheme. The anti-free-rider principle or the fairness principle states that “in a just institution, when a number of persons engage in a mutually advantageous cooperative venture according to rules, and thus restrict their liberty in ways necessary to yield advantages for all, those who have submitted to these restrictions have a right to a similar acquiescence on the part of those who have benefited

from their submission.”³³ People with permanent disabilities cannot satisfy this principle because they have, so to speak, no liberty to be submitted. However, first, formally speaking, people with disabilities could still be regarded as having acquiesced to a conditional submission of their liberty if their capabilities can be fully or partially restored, a hypothetical promise that would not be broken by those suffering permanent disabilities. Moreover, what matters, in fact, is not whether they can comply with the principle by contributing their efforts but whether their failure to do so will discourage the able participants from supporting the cooperative scheme. If able people can commit to the more generous norm of reciprocity, as predicted by the formal analysis, they can also endorse a less demanding anti-free-rider principle which regulates only the unwilling, not the unable.

It may sound cold-blooded to base the inclusion of people with disabilities on formal analysis. But the major takeaway from it, that players should be made aware of the permanent risk of becoming disabled, actually reflects an important and humane idea of empathy—the simple yet often neglected thought that “it could be me”. Because the human condition is such that we are all exposed to the risk of disability, which may render us incapable of sustaining ourselves, let alone carrying out our life projects, without external support, we are in the common need of a cooperative scheme to share this risk. Such recognition of our shared vulnerability is what makes it possible for the risk-embedded reciprocity to prevail.

Moreover, formal analysis has also shown that such a sense of empathy can be encouraged by properly structuring the game. When the game is designed so that the risk of disability can be shared among players through the provision of a risk fund, the more inclusive strategy of cooperation, instead of the strict exclusion trigger strategy, becomes the NE strategy and gets commonly adopted by players. This implies that, even when a more compassionate notion of

³³ Rawls, *A Theory of Justice*, p. 96.

reciprocity is currently missing among a group of people, it may be fostered through the establishment of a properly designed risk sharing institution which acknowledges the urgency of restoring capacities for people with disabilities before demanding their contributions.

Social Norm and the Feasibility Constraint

It may be confused that, when applied to the contract approach, the cooperative strategy of a game is something to be chosen in the original contracting position, thus being subject to the same concern of compliance after the Veil is lifted. However, as I have shown, what needs to be chosen in the contracting position is the structure or mechanism of the game. If people are placed in a cooperative structure which takes the form of a risk sharing game, their rational choices may converge to a NE strategy that is similar to what I have just described so that the Pareto-optimal outcome of universal cooperation can be achieved.³⁴ This strategy may, over time, be internalised as the social norm among the group of people interacting under the cooperative scheme.

This conclusion bears on the issue of feasibility of relying on certain norms to sustain social cooperation. Rawls seems to think that the prevailing norm of reciprocity represents a hard feasibility constraint for generating principles of just social cooperation. My analysis shows that it may in fact only be a soft one. Fostering a norm that is not currently present may be feasible if institutions are designed in a way so that the norm is transformed into a NE strategy of cooperation. Hence, in the case of risk sharing, even if strict reciprocity is the prevailing norm in a society, with the establishment of fair risk sharing institutions, risk-embedded reciprocity may emerge as the new norm. While this is not guaranteed to happen, as how a social norm is

³⁴ When there exist multiple NE strategies, which one is commonly employed by members of a community or of a society is often the result of historical contingency. See Voss, “Game-Theoretical Perspectives”, p. 117.

actually realised can be influenced by other historical and social factors, efforts can at least be made to promote its realisation through good institutional designs.

On the other hand, it is not always feasible for an institution to simply demand allegiance to any desirable norm. A norm of unconditional cooperation, for example, can hardly prevail because it is not a NE strategy in an imaginable cooperative scheme in public good provision and its conformity is therefore not self-enforcing. This is not to say that an institution presupposing such a norm cannot employ coercive measures to enforce compliance, but such enforcement would not be sustainable. An institution presupposing universal support of a non-equilibrium strategy could not, in Rawls's words, "generate its own support". It is feasibility in this sense that may justifiably condition a contractarian theory that aims to guide social cooperation.

What I hope I have shown is that the feasibility of moving beyond the more widely accepted norm of strict reciprocity and realising the social norm of risk-embedded reciprocity is categorically different from the feasibility of, say, achieving unconditional cooperation among society members. The latter may be regarded as a hard feasibility constraint while the former is only a soft one. The difference is not simply a matter of degree in terms of demandingness of the goal. Rather, the crucial difference lies in whether the norm to be realised is a self-enforcing one. Admittedly, external enforcement may always be necessary to initiate a change of norms. Once the relevant institution is established, however, it should be able to render norm conformity self-enforcing rather than permanently relying on coercion. It is this requirement of stability that, I think rightly, concerns Rawls as a real feasibility constraint in determining the scope of cooperation. What I have argued is that, because the more generous norm of reciprocity could be a NE strategy in a social cooperative scheme which includes a risk sharing mechanism, presupposing this norm, that is, presupposing that people with disabilities could be accommodated in the social contract, does not violate reciprocity as a feasibility constraint.

To conclude, I have argued that including people with disabilities, even severe and permanent, in the scope of social cooperation does not violate reciprocity as a demand of justice. On the other hand, prevailing norms of reciprocity may indeed represent a feasibility constraint for sustaining cooperation with people with disabilities. But it is merely a soft one and may be overcome with the establishment of fair risk sharing institutions. Therefore, presupposing a more generous norm of reciprocity, that is, presupposing that people with disabilities could be accommodated in the social contract, does not violate reciprocity as a feasibility constraint.