November 20, 2020

The Samaritan's Curse
Moral Individuals and Immoral Groups

Kaushik Basu
Department of Economics
Uris Hall
Cornell University
Ithaca, New York 14853
Email: kb40@cornell.edu

Abstract

In this paper, I revisit the question of how and in what sense can individuals comprising a group be held responsible for morally reprehensible behavior by that group. The question is tackled by posing a counterfactual: what would happen if selfish individuals became moral creatures? A game called the Samaritan's Curse is developed, which sheds light on the dilemma of group moral responsibility, and raises new questions concerning 'conferred morality' and self-fulfilling morals, and also forces us to question some implicit assumptions of game-theory.

Key words: Samaritan's Curse, moral responsibility, conferred morality, Nash equilibrium

JEL Numbers: D62, D64, K13

Acknowledgement This paper began as an informal talk in the Philosophy Department of Cornell University, and developed as versions of it were later presented at a game theory conference in IIT Bombay, a conference in honor of Jorgen Weibull at the Stockholm School of Economics, and the Moral Psychology Brown Bag at Cornell. I also benefited from my seminar and discussions as part of the Hamburg Lecures in Law and Economics. I would like to thank Richard Ashley, Abhijit Banerjee, Chris Barrett, Karna Basu, Larry Blume, Matthew Braham, Tad Brennan, Utteeyo Dasgupta, Frederik Van De Putte, Avinash Dixit, Robert Duval, James Foster, Meir Friedenberg, Nicole Hassoun, Karla Hoff, Rachana Kamtekar, Peter Katzenstein, Theodore Korzukhin, Rajlakshmi Mallik, Ajit Mishra, Diganta Mukherjee, Derk Pereboom, Ariel Rubinstein, Sudipta Sarangi, Eric Van Damme and Joseph Zeira.

1. Moral Attribution: Scholarly and Quotidien

Listening to the common person discuss group morality can be a disturbing experience. Listening to scholars discuss group morality, with an attentive ear, can be equally disturbing. The former is conducted with a parsimony of reason. The latter is conducted with complex reasoning but, on closer examination, leaves one as puzzled. One reason for this is that, when we see something good or bad happen, it is a natural human instinct—to the laity, the philosopher, and the economist, alike—to place the responsibility at the doorstep of some agent. It leaves us uncomfortable not to be able to attribute agency to anybody.

This is not an easy problem because to err on either side is likely to have grave consequences. To wantonly hold individuals responsible for bad group behavior can be dangerous. It can lead to the wrong person being punished, and promote facetious group labeling, and even racism. On the other hand, to consistently hold no one responsible for bad group behavior is to risk encouraging bad behavior. It can result in the 'many hands problem,' where a group of individuals drive society to a disastrous outcome, in a manner where no one person's unilateral change of behavior can make any difference (Braham and Holler, 2009; Braham and van Hees, 2012).

This is a philosophical challenge with enormous implications for real-life outcomes, ranging from preventing large corporations from hiding behind the excuse that no *one* in the corporation is responsible for the damage it inflicts on society, to drafting laws for preventing climate change disaster. Not surprisingly, there is a substantial literature in moral philosophy, law, and economics, dissecting the link between group behavior and the moral intention of the members of the group[1].

This paper follows the strategy of bringing concepts and arguments from game theory to investigate this problem. This is not as common as one would expect, even though as a method, especially if we include bargaining games as part of game theory, this has a long history and a growing contemporary literature (e.g., Bacharach, 1999; Braham and van Hees, 2012; Chocker and Halpern, 2004; Friedenberg and Halpern, 2019)[2].

What is attempted here is however different from what happens in much of that literature. What this paper does is to study a counterfactual. Instead of asking what would happen if the individuals *chose differently*, I ask an antecedent question: What would happen if the individuals *were different*, becoming, for instance, moral creatures? In other words, I consider the fall out of

[1] See, for instance, Feinberg (1968), Held (1970), Bernheim and Whinston (1986), Rubinstein (1998), Sartorio (2004), Tannsjo (2007), Petersson (2008), Hakli, Miller and Tuomela (2010), List and Pettit (2011), Hess (2014), Arruda (2017), Sen (2018), Dughera and Marciano (2020).

[2] A more methodological analysis of the role of game theory in dealing with agency problems occurs in Nguyen (2019). One of the earliest instances of using game theory to shed light on moral questions is the paper by Runciman and Sen (1965). More generally, it is worth recalling that game theory began with the analysis of actual games, such as Chess and Bridge. Its importance, even in economics, happened only after classical game theory gave way to an "instrumental view" of the subject (Samuelson, 2016). In that sense the present exercise is an extention of earlier initiatives.

a change in people's motivation. The problem is brought into sharp focus by creating a new game, the Samaritan's Curse, which illustrates the paradoxical result where all individuals becoming moral makes the group's behavior more immoral. By constructing this example, the paper shows that not only is it difficult to allocate to individual members of a group moral responsibility for some bad action by the group as a whole, as already pointed out in a large literature[3], but that there are contexts where individual morality is in fact the source of the group's immoral behavior. The game is then used to analyze some of the problems of moral attribution of group behavior and raise new questions concerning morality and game-theoretic reasoning, and to develop the concept of "conferred morality".

The problem is important because the attribution of moral responsibility to groups occurs all the time, on television, in newspaper opeds, and in our quotidien discourse. Here are Nancy Altman and Linda Benesch in the **Huffington Post**[4] on the immorality of "Republican leaders" in the US: "The Republican elite's immorality goes well beyond Donald Trump … The immorality and disdain of today's Republican elites shine through in the policies that they embrace." And, to ward off any charge of partisanship, let me quote Dave King[5] on the "immoral Democrat Party": "Here is a short list of the damage being done to our country by the immoral Democrat party." I may add that his list is not short.

Such indictments are not exclusive to America. We talk about the lack of morals of big corporations around the world; we use collective moral epithets to describe the Houthis, Tutsis, and protest groups as in the Arab Spring. Such loose group moral attributions do a lot of harm because they prevent us from devising sensible policies to improve our collective behavior. In fact, when it comes to selfish behavior we no longer make such mistakes. The long history of games such as the Prisoner's Dilemma has sensitized us to this. No longer do we say that the depletion of our common resources is occurring because we are not acting in our individual self-interest. It is now understood that our common resources are being depleted because individuals so often act solely in their self-interest. It is this recognition that has helped us make some progress towards managing climate change, and having global contracts to save our environment.

We need to develop a similar awareness about the possible paradoxes of moral behavior. The Samaritan's Curse tries to bring this into sharp focus. In the process the game also forces us to grapple with questions concerning the place of consequentialism and deontological ethics. This paper takes a broad position of normative methodological individualism. It argues that this is the only reasonable way to approach groups and collectives. It argues that we have been too hasty in treating a group or a collective as "a *moral* agent in its own right," with its own "rational point

[3] For recent research, see Braham and van Hees, 2012; Bjornesson, 2014; Chant, 2015; and Arruda, 2017.

[4] See https://www.huffingtonpost.com/entry/the-deeply-immoral-values-of-todays-republican-leaders_us_5a2eb9f7e4b04cb297c2aee5

[5] See https://www.conservativedailynews.com/2015/11/the-immoral-democrat-party-list-of-shame/

of view" (see Hess, 2014, p. 203) and with its "collective intention"[6]. The critique that arises from the game, The Samaritan's Curse, leads to the concept of 'conferred morality,' which helps us think of practical problems, such as environmental degradation and dealing with the problem of corporate responsibility. How should we allocate responsibility for bad corporate behavior among the individual decision-makers within the corporation[7]? The existing literature distinguishes group moral responsibility that can and cannot be distributed among the individuals of the group. When it cannot be distributed, "groups sometimes might be regarded as independent agents… Hence, it seems possible that groups as such can be morally responsible for the effects of their acts, in a sense that cannot be reduced to judgments about individual members' acts." (Petersson, 2008, p. 244; see also Pettit, 2007).

I contest this, arguing that in such cases we have no innate reason to hold the group morally responsible, but we may, nevertheless, create penalties and rewards, which give rise to a 'conferred morality' that can lead society to a better outcome. The need is to draw a distinction between different kinds of responsibility, of which *moral* responsibility is one (Tannsjo, 2007) and within that there may be *a priori* or natural morality, and conferred morality.

The other use of the Samaritan's Curse is that it draws attention to implicit assumptions that underlie conventional game theory. The explicit assumptions of game theory are well-known, as outlined by Braham and van Hees (2012) in the context of group morality. However, like all theory, game theory has explicit assumptions but also many hidden assumptions—assumptions in the woodwork. The challenge of group morality compels us to recognize and contest some of these assumptions in the woodwork.

2. The Samaritan's Curse

Consider a game with two players, 1 and 2, with each player having three actions to choose from: A, B and C. In other words, person 1 can choose from the set of alternative possibilities {A, B, C}. And likewise for person 2. I should clarify that person 1's A, B and C are not the same as 2's A, B and C. Thus 1 may be choosing whether *1 should be* a farmer, lawyer or philosopher and 2 may be choosing whether *2 should be* a farmer, lawyer or philosopher. I could have labelled 2's alternatives as A', B', C', but I prefer to economize on the primes.

The set of actions to choose from represents the alternative possibilities from which each person can freely choose and those are the only actions from which she can choose. The 'outcomes' (meaning all possible eventualities) of the game are pairs of choices, one choice by each player. Hence, this game has nine possible outcomes: (A, A), (A, B), (A, C), (B, A), and so on, where an outcome such as (B, A) refers to the case where 1 chooses *her* B and 2 chooses *his* A.

[6] We can of course talk of shared preferences, as exemplified by expressions like "we prefer" (Sugden, 2000), or the use of what is referred to as the "we mode" approach (Tuomela, 2006), but it is arguable that this is distinct from group agency.

[7] For analyses of moral responsibility of group entities, such as corporations and universities, see Copp (2006), Pettit (2007), List and Pettit (2011), Chiao (2014).

To describe a game fully we must specify what each player earns once an outcome is reached. This is summarized in the 'payoff matrix', labelled The Basic Game in Table 1. Following standard convention, the rows represent player 1's choice and columns represent player 2's choice. Thus, if player 1 chooses B and 2 chooses C, the outcome is (B, C), and from the Basic Game matrix we can see 1 earns 80 and 2 earns 102. In each box, the first number is player 1's payoff and the second number is 2's payoff. Without loss of generality, I shall refer to the payoffs as dollar earnings of the individuals.

Table 1. A Society

The Basic Game

	A	B	C
A	102, 102	80, 120	108, 108
B	120, 80	104, 104	80, 102
C	108, 108	102, 80	106, 106

Bystander's Earnings Matrix

	A	B	C
A	20	4	0
B	4	6	10
C	0	10	4

While there are minor differences in terminology with Braham and van Hees (2012), in essence, the description of a game here is the same as in their paper, which, in turn, is the same as the description in any standard game theory text. In their paper, Braham and van Hees consider different solution concepts or predictions of outcomes, whereas I shall focus here on one solution—the Nash equilibrium. This is for the good reason that I consider a specific game where there is a unique Nash equilibrium which happens to be strict. When that happens it seems natural to expect that the Nash equilibrium is where the players will settle. In fact, in such games virtually all equilibrium concepts of game theory coincide with the Nash equilibrium.

For completeness, a 'Nash equilibrium' is defined as an outcome such that no individual can do better by unilaterally deviating to some other alternative. The Nash equilibrium is called 'strict', if every unilateral deviation actually leaves the deviator *worse* off.

To see why the Nash equilibrium is the natural concept to use here, assume that the players are selfish, and playing the Basic game. It is reasonable to see that this society will get to (B, B). This is the only outcome from which no player has an interest in deviating unilaterally. For any other outcome pair, either player 1 or player 2 will prefer to deviate unilaterally. Hence, (B, B) is the only Nash equilibrium and it is the only place where this society would settle down. Since the Nash equilibrium is the only solution concept used here, I henceforth drop the adjective Nash.

Thus the game I just described, shown in the left-hand panel of Table 1, has a unique equilibrium at (B, B). Note that there are other outcomes where both would be better off, such as at (A, C) and (C, C) but society can never settle there because someone will, in each case, unilaterally deviate.

Note that in this interaction neither player paid any attention to the fallout of their behavior on bystanders (or the other player, for that matter). Bystanders are marginalized people whose well-being depends on the choices made by the players. As of now I have not even described how the bystanders fare. In a society with selfish players, such as what I described thus far, that is of no material consequence.

Let me now describe how bystanders fare. Assume, for simplicity, that there is only one bystander, and the payoff to this person is as shown in the 'Bystander's Earnings Matrix' in the right-hand panel of Table 1. The way to read this matrix is to first find out what the outcome of the game is (that is what the players do) and then read off how much the bystander earns. Thus if the players choose (B, C), the bystander earns 10. I shall henceforth do what I have already done, refer to the full description of the two players plus the bystander as a 'society'. Table 1, with the two matrices describes the full society[8].

With selfish players, since the equilibrium will be (B, B), the bystander will get a payoff of 6 in equilibrium. In this society, when the players make their decision, they do not pay any attention to the effect their actions have on the bystander.

Now suppose a good Samaritan comes to town. She is dismayed by the moral degeneracy of the players, and the fact that they pay no attention to the poor bystander's well-being. Here they are, rich individuals, earning 104 dollars each, whereas the poor bystander gets only 6. Observing this, the Samaritan gets down to teaching the players some basic morals. Morals can be of many kinds[9]. For simplicity, I shall suppose that the Samaritan focuses on a consequentialist ethic--a kind of utilitarianism[10], with some attention to equity, which seems like a natural moral in this context. Thus the Samaritan tells them: Your choice of (B, B) gives you (104, 104), but do you not see this leaves the bystander with a miserable payoff of 6? If you, the super-rich, opted for (A, A), you would each lose only 2 and the bystander would get 20. Surely you should be prepared to sacrifice $2 for an additional $14 for the poor bystander? Ignore the other player, if you wish, since she is super-rich like you, but be mindful of what happens to the poor when you choose. With that the Samaritan vanishes.

Suppose the players now become moral creatures, who value the earnings of the bystander (or anyone who is below the poverty line of less than, say, $25) as much as his or her own payoff. If the outcome is, for instance, (A, B), player 1 gets a consolidated payoff of 84, which consists of 80 for herself and 4 for the bystander, and player 2 gets a consolidated payoff of 124, consisting of 120 for himself and 4 for the bystander. The consolidated payoffs for both players, for each of the possible outcomes, gives us a new game, The Samaritan's Curse. This is shown in Table 2. It

[8] From a formal game-theoretic point of view, we could actually think of the bystander to also be a player, but a player who has no choice. She has a strategy set consisting of one action. By this description this is a standard 3-player game, the two players earlier introduced plus the bystander.

[9] There is a growing literature in economics on this (Sen, 2018; Alger and Weibull, 2019).

[10] For a description of consequentialist and non-consequentialist public persona, see Gintis (2015).

is easy to see that the equilibrium of the new game is (C, C). This is in fact the only equilibrium. This is an outcome from which each *moral* player will not have reason to deviate.

We have seen games like this in standard game theory, where a better collective outcome, such as (A, A) eludes the players. What is special about this game is what lies behind it—the fact that it emerges from another game by virtue of players *becoming moral*. What is interesting is that it is starting from another game and the players turning into good Samaritans, concerned about the bystander, that worsens the bystander's welfare. The bystander was earlier getting 6, and now gets only 4. What we have is a kind of Prisoner's Dilemma, or related games, such as the Traveler's Dilemma (Basu, 1994) or the Gingerbread game (Hollis, 1994), but in a moral domain. The Samaritan's Curse illustrates how *transforming* players into good Samaritans can lead to a new equilibrium that hurts rather than helps the poor. In short, *individually* moral behavior makes the *group's* behavior more immoral.

Table 2. The Samaritan's Curse

	A	B	C
A	122, 122	84, 124	108, 108
B	124, 84	110, 110	90, 112
C	108, 108	112, 90	110, 110

Hence, the Samaritan's game warns us that when we look at a group's immoral behavior, and implicitly think of the individuals constituting the group as immoral, we may be wrong. Further, more interestingly, it is the fact of people becoming more moral that makes the group behave morally worse.

Let me pause here to connect with the existing literature on moral responsibility. I am taking the standard view that if a bad outcome occurs and there was an action that was available to a person such that by taking that action the outcome would have been changed to a good one, then we would say that this agent is morally responsible for the bad outcome. In the above example, suppose we take the view that an outcome is morally good if and only if the poor bystander gets more than $15. All other outcomes are bad. Suppose that society happens to be settled in the equilibrium at (C, C). Clearly, no individual can take the society from this bad outcome to a good outcome. It needs coordinated action by all individuals, which is in nobody's feasible set[11].

[11] I am not here getting into another foundational question about the meaning of each individual's feasible set, in this case {A, B, C}. There is plenty of debate about whether a person choosing from among a set of alternatives can be held responsible for his or her choice (Frankfurt, 1969; Haji and McKenna, 2004; Pereboom, 2008). Would the fact that people may be fully determined render them free of responsibility? I have elsewhere taken the line that free will and determinism are fully compatible (Basu, 2000). Let us stay away from this debate here and treat the feasible set as the collection of all possible alternatives that a person can choose from and be held responsible for.

In short, every individual can be genuinely moral, try to do whatever he or she can do to enhance the modified utilitarian objective but nevertheless find that they have collectively taken society to an outcome that all of them consider morally inferior[12]. This is the central problem highlighted by the Samaritan's Curse. It illustrates the helplessness of individually moral people. All agree that there is a morally better state but the players are hindered, ironically, by their own morality, from taking society there.

Some may argue that in the Samaritan's Curse, the players described as moral are not truly moral. If they were, they should analyze the consequences of their behavior and see that their morality is doing harm. This should prompt them to change their behavior.—For instance, be selfish in order to achieve the moral outcome[13]. The problem with this is that a moral person would always be tempted to cheat and play morally after having duped the other person into thinking that the person is selfish. But, knowing this, the other person will not be duped.

This same problem arises in standard game theory with selfish players. When a certain behavior does harm to all individuals, why don't individuals change their own behavior? This has been discussed in the context of individual rationality and there is no easy resolution of the problem (see Basu, 1994; Hollis, 1994).

In similar situations, there has been discussion of group agency and group responsibility that is a collective responsibility, one of the most ingenious being the work of List and Pettit (2011). If we think of the group, namely, players 1 and 2 together—call it group G—as having agency, then we could say that the group has a feasible set consisting of (A, A), (A, B), (A, C), (B, A) and so on—basically all the 9 outcomes of this game, from which to choose one. In that case, of course, G is morally responsible for the bad outcome (C, C), because G is assumed to have agency and G chose (C, C), when G could have chosen (A, A). The problem with this approach is that it would *seem to* violate methodological individualism[14]. Further, even if we agreed that the group is morally responsible, this would be a semantic attribution and of little consequence if we could not apportion that responsibility to the individuals who constitute the group. In case we tried to do this by arguing that there was evidence that this group was consciously formed by players 1 and

[12] In situations where you fall into a trap that no one wishes to be in, such as in the Prisoner's Dilemma or the Traveler's Dilemma, one way out is the celebrated suggestion of Gauthier (1986), whereby individuals agree to place restrictions on their own choices in order to collectively achieve what is ultimately in their own self-interest. At first sight, it may look as though this argument can be used to solve the problem here. But there is a hurdle in trying to apply Gauthier's approach which is founded on the assumption of "mutual unconcern." The use of this assumption for solving moral problems has been contested (see Vallentyne, 1989). But even if we ignore that, note that, in my model, the problem arises *because of* a new concern for the other—the poor bystander. So mutual unconcern tends to undo the very basis of the problem. Of course we may mechanically apply Gauthier's approach by treating the payoffs in the Samaritan's Curse game (in Table 2) as selfish payoffs and then applying his idea of "constrained maximization" on this. This, however, is not persuasive because it tends to reduce the approach to a tautological one of being able to solve every problem of sub-optimality by redefinition.

[13] It is in fact arguable that being selfish to enhance a moral objective is not really selfish, since in the end it is the intention that matters.

[14] It is important to point out that List and Pettit (2011) are clear that their approach is one rooted in methodological individualism, or what I here refer to as normative methodological individualism.

2 in order to play this game, we could hold 1 and 2 individually responsible. But then, we would be talking of a different game than the one we have described. The formation of the group would have to be a part of the game. I shall pick up on both these ideas later in the paper, namely, how morality may be conferred in retrospect and what might it mean to step beyond the game of life.

Essentially, what the the Samaritan's Curse does is to remind us that, when it comes to policymaking, we need to do for moral behavior what we have done for selfish rationality. A major contribution of game theory was to demonstrate that individually rational behavior may not lead to rational outcomes for the group, as a lot of the orthodoxy in economics had come to believe. This led to thinking about how we as a collectivity can introduce laws, taxes and rewards to bring individual behavior into alignment with our group interest. There are interesting examples of this kind of theorizing being used to design regulation for preserving the commons and global agreements on climate change (Barrett, 1994).

The Samaritan's Curse prompts us to think in similar terms to bring individual *morality i*nto alignment with group *morality*. I do not have a full resolution of the problem. One of the objectives of the paper is to bring this paradox to the attention of scholars and policymakers, and to open up discussion for its resolution. As a tentative first step, I argue below that the paradox may be reason to consider certain kinds of deontological ethics to guide our behavior, recognizing that these deontological principles may be derived from distant consequential aims which are not within the reach of any individual, somewhat in the spirit of team reasoning (Bacharach, 1999). There is also the possibility that we may have sets of actions, such that each action has no consequence but the set of actions together has a significant consequence, which is a formalization of the sorites paradox or the paradox of the heap, the original formulation of which goes back to Eubulides of Miletus (Parfit, 1984; Basu, 2000; Voorneveld, 2010). In such contexts, to justify taking each action we have to appeal to some rule-based behavior. Indeed, it is arguable that morally optimal outcomes cannot be assured by creating incentive compatible structures. In the end we may have to rely on players giving up opportunism, even moral opportunism, and altering their behavior types, such as nurturing trusts (Francois and Zaboljnik, 2005).

3. Conferred Morality

The Samaritan's Curse is one specific construction where the paradox of morality arises but the game is not a pure figment of imagination. It is possible in reality, if for no other reason, because we could simply create a board game with exactly the features of the Samaritan's Curse. How ubiquitous the problem is, is open to debate. The Samaritan's Curse warns us that just being moral may not always yield the moral outcome.

Is it solvable in the sense of there being a way to alter the disconnect between individual intention and group outcome? While I do not have a full answer, one particular approach that I want to explore here to suggest a way out involves the idea of 'conferred morality'. I should point out, even at the outset, that this is merely a pointer towards a resolution, rather than a resolution.

Braham and van Hees (2012) use a game-theoretic analysis, where, unlike in this paper, individual motivations remain unchanged. Each player wants to maximize his or her own payoff. They conclude that all individuals who deplete a resource, as in the commons problem, cannot each claim no moral responsibility on the ground that "my contribution to the depletion of resources would not have any noticeable effect." They construct an argument in which, even if individuals are miniscule parts of the game, each can be held morally responsible[15].

The Samaritan's Curse, by considering the case of individual motivation undergoing a change, from selfish to moral, blocks this path to resolve the problem. In the Samaritan's Curse we cannot hold the individuals morally responsible for the bad group outcome because it is, in fact, their morality that makes the problem worse. However, this paradox, along with the argument used by Braham and van Hees, suggests a way of resolving the problem of moral responsibility that is different from theirs but reaches the same ultimate goal. To see this, we need to recognize that there are situations where the law confers a morality that may not have been there prior to the announcement of the law. This does not happen with all laws but there are contexts where, after some behavior is declared illegal, such behavior is deemed immoral by us. Once this happens, individuals may not need the law to uphold the moral outcome. Our acquired morality does the job of monitoring our behavior. I shall refer to such morals as 'conferred morality'.

Consider the case of income tax. Most societies consider it a moral lapse not to pay one's income tax. The same money not paid to the government would not be a moral lapse if the nation did not have a law requiring individuals to pay income tax. Hence, the same action can become moral or immoral by virtue of the existence of a law. That is the idea of morality conferred by the law. I shall explain later that this is true not just for some laws but can also happen with some proclaimed or publicly-declared morality.

But before that, I want to present a stark and interesting example of conferred morality where, of two otherwise identical actions, one becomes a moral act. Consider an unusual commons problem. There is a village with 100 people and 2 ponds, X and Y. Garbage in this strange village must be thrown into a pond, any pond. Suppose that every pond in this strange world has a large fish stock that survives if and only if less than two people throw garbage in it. If two or more persons throw garbage in a pond, all the fish in that pond perish.

The game that these villagers play is the following. Each villager chooses a pond to throw garbage in. As long as the fish stock is unaffected by their choice of pond, they are indifferent about which pond they throw garbage in, and in that case, by force of habit, people throw garbage in the pond closest to their residence. But if by unilaterally choosing to throw garbage in a particular pond the fish stock in the village becomes higher, the person would choose that pond.

[15] This is distinct from arguments based on an explicit rejection of the dictum "ought implies can," such as developed by Tessman (2015). She takes a line where if you do not save a drowning child because there was no way for you to do so, even then you may have to take some moral responsibility, which seems to me to be a recipe for a life of guilt.

Assume that half the population lives close to pond X and the other half close to Y. It is possible that, to start with, half the population will be throwing garbage in X and the other half in Y. In this society, there will be no fish but there is nothing any individual can do to change this. In short, this is a Nash equilibrium.

Now suppose a law is enacted which prohibits throwing garbage in one pond, which is arbitrarily chosen, say pond X. It is arguable that, once this law becomes effective, and people stop throwing garbage in X, this acquires a moral status. Once people begin to abide by the law, it is in each individual's self-interest to abide by the law. Drawing on Schelling's (1960) powerful idea of the focal point, it is arguable that the the power of the law derives from its ability to create a focal point (Sunstein, 1996; McAdams, 2000; Posner, 2000; Basu, 2018). Starting from the fishless equilibrium in which half the village population throws garbage in each pond, what the law does is to nudge society to another equilibrium, where everybody throws garbage in pond Y and everybody is better off because there is quite a lot of fish available.

It is, further, arguable that this law, which leaves pond X clean and reserves the other one, namely Y, as the garbage dump, once in place, will acquire a moral status. For anybody to throw garbage in X will be considered not just an illegal act but an immoral act (like not paying one's taxes to a legitimate government). Since ponds X and Y are *a priori* identical it is the law that *confers* this morality. In the original equilibrium, where garbage was freely thrown in both ponds, the act of throwing garbage in pond X (or for that matter in Y) was not an immoral act.

There are real life examples of this. The law that one should not smoke in an auditorium with other people is now so much a part of our morality that it is arguable that even if the law is revoked today, people will not smoke in an auditorium. A more immediate example, unfolding in front of our eyes, concerns rules of behavior pertaining to controlling the spread of coronavirus. Most governments are beginning to develop rules and on some occasions actual laws about how many persons can congregate in one place, in what kinds of situations people are supposed to wear masks, and how much social distance one should maintain from other people. There is an arbitrariness concerning the choice of rules, somewhat akin to the above example of ponds and pollution. But once those rules are chosen and people begin to follow them, they acquire the added force of morals and even social sanctions to enforce them.

What the above example does is to open up the question about what laws should be enacted with the explicit purpose of conferring morality on certain actions? I would argue that in the standard commons problem where the environment is damaged by a large group of well-intentioned individuals, each of whom is too small to be of any consequence, we cannot describe each of their actions as immoral. In other words, I am expressing a disagreement with Braham and van Hees (2012) and, in fact, with a commonly-held view. However, their argument serves an important role. It provides grist for selecting cases where we should *confer* morality. Their argument suggests that we should enact laws to prevent the over-exploitation of the commons. Once such a law is in place, then for someone to exploit the commons is to be immoral, even if such unilateral action is inconsequential. It is not a naturally immoral act (on the part of any single

individual, since a single person's action does not have any effect) but an act of conferred immorality.

There is a subtle but important distinction between the traditional view of legitimacy and natural law, and what is being suggested here. There is a long tradition of writings on legitimacy which take the view that the legitimacy of the law is based on our prior, natural sense of morality (for recent discussions see Singer, 2006; Tyler, 2006; Akerlof, 2016). The line being taken here is that moral legitimacy is important but, at times, the legitimacy is a consequence of the law. The idea of conferred morality stresses that there are situations where, instead of nature providing us with *a priori* morals, the need is to design and enact laws, such that they subsequently acquire a moral status, because of the law. It is not that every behavior rule can acquire a moral status but a subset of rules has this potential, and to promote good group behavior we need to choose from this subset, make that into a law, then have morality conferred on it.[16]

The argument I just presented pertains not just to laws but also to the proclamation of morality. In brief, there are some actions which are not *a priori* immoral, but once they are publicly declared to be immoral, they trigger behavior such that after a while they become immoral. The immorality (and by virtue of that, morality) of certain actions gets conferred. In other words, morality can, in certain situations, be self-fulfilling. Therefore, the argument here is that, faced with the commons problem, it is not that individuals who are part of this large collective tragedy are morally responsible. But we may nevertheless want to publicly declare them as morally responsible. This can trigger behavior change and the preservation of the commons, so that the few who may violate the norm and over exploit the commons are then commiting an immoral act. A morality that was not there gets conferred by the proclamation of a moral judgement. There is a small ethical problem with the proclamation because *when* that is made we know that is not true but are aware that it will eventually be true. It is like a central bank forecasting a certain inflation rate, knowing that it is the central banks's announcement that will cause the inflation to be that.

In the context of the Samaritan's Curse, having analyzed the context, we may decide to enact a law that declares it illegal to choose actions B and C. Once this law is implemented and society stabilizes at (A, A), this can acquire a moral connotation. Anyone who deviates to B or C (and hurts the poor bystander) will be viewed as being immoral. This would then be a conferred morality, a deontological ethic, which, once it becomes ingrained, allows society to get to the morally good outcome without third party enforcement and purely by our own sense of morality.

The discussion of conferred morality in the context of the Samaritan's Curse explains the above caveat, namely that this is merely a step *towards* a resolution of the problem. Note that conferred morality entails an element of psychology. Not every proclamation or law would acquire moral force. When and how that happens is an open matter. It is possible to argue that if the

[16] This is also to be distinguished from the interesting idea that the *mode* of law enforcement (Is it done by the police, who earn a fixed salary, or is it by private agents who profit from the law's enforcement?) can help confer legitimacy (for an experimental study, see Dickson, Gordon and Huber, 2015).

proclamation is too *ad hoc*, it will not acquire moral force. Therein lies the problem. In life we are likely to encounter many gaming situations. If the law devised is not from any general principle but an *ad hoc* response to the specific game that has suddenly cropped up, it is unlikely to acquire the force of conferred morality. Ideally, we should be able to proclaim some rules of behavior, *before* we confront the game, have those rules be ingrained in us and then we can hope that laws derived from these will be reinforced with moral force. Whether we can always do this and, so to that extent, whether problems such as the Samaritan's Curse (and there can be many variants to it) are always solvable remains for now an open question.

What I have done here is to give formal shape to an argument that has been hinted at in the literature, such as when Pereboom (2017) argues, "On one type of revisionary account, our practice of holding agents morally responsible in a desert sense should be retained, not because we are in fact morally responsible in this sense, but because doing so would have the best consequences relative to alternative practices." What I argued is that in such cases we should create laws or norms that declare certain behaviors as wrong since that creates a conferred morality, which acts as a deterrence against such behavior.

4. Morality behind the Veil of Ignorance

Though I believe that faced with moral paradoxes of the above kind we either have to live with them as unresolvable problems of life or, in some cases, as just shown, resort to some kind of deontological ethic via a conferred morality to avert the problem, there is one more avenue of inquiry that has not been explored thus far.

Note that the paradoxical result that the Samaritan's Curse game draws attention to, involves people becoming moral in somewhat different ways. Each of them values their own incomes plus that of the poor but not of other rich people. This results in moral preferences whereby one moral person may prefer outcome p over q, while the other prefers q over p.

Some readers may argue that truly moral preferences cannot be different. If each of us views society as an impartial spectator or from behind the veil of ignorance, we will reach the same conclusion. Whether or not this is a valid observation concerning moral preference, let me go along with it here[17]. Without loss of generality, assume that when people become moral they all rank outcomes the same way. For simplicity, in the above example, assume that when they become moral they are utilitarian in a Benthamite sense. Their aim is to maximize total well-

[17] It may be worth pointing out that even if we take this view of morality, the Samaritan's Curse game does remain relevant for real-life problems. This is because, though I have been talking about morality, preference switches can take place for many reasons, as recognized in the large literature on behavioral economics. The same person can have different preference at different points of time. People can learn to develop altruism and kindness. Kindness can often take the form of the preference described in section 2 after the arrival of the good Samaritan. And kindness can have many sources. People may enjoy being kind, they may not enjoy being kind but consider it their duty to be kind, or they may like it when others view them as kind (see, for instance, Dufwenberg and Kirchsteiger, 2019; Rabin, 1993), all resulting in the same manifest behavior. Interpreted this way, what the Samaritan's Curse shows is that, in strategic environments, all individuals becoming kind may not result in greater kindness.

being in society (in this case equated with total dollar incomes earned by society). In this case, of course, if one person considers outcome p morally superior to q, so will the other. In other words, after people become moral, what we have is a 'unanimity game,' namely, a game in which all players are unanimous about what they seek to maximize (Basu, 2010). They all have the same payoff function. Thus, the question being asked in this section is whether the Samaritan paradox can occur in a unanimity game?

This has an interesting answer. The paradox cannot arise in the same way and in the same sense as above but a different problem can emerge. It cannot arise in the same way because if all players have the same preference, for instance when everybody becomes a Benthamite utilitarian, the morally-best outcome in society will always be a Nash equilibrium. This is for the simple reason that there is no way to do morally better by a unilateral deviation because nothing morally better exists. So becoming moral cannot lead society to an immoral outcome in the way that this happened in the Samaritan's Curse. But it can, in a different way. So the short answer is the same problem cannot occur when everyone is utilitarian but a related one can. The rest of this section illustrates this.

Consider a society with ten rich players and one poor bystander. Each player has to choose an integer from 1 to 10. If all of them do not choose the same integer each gets a payoff of 0 dollars. If all choose the same integer they are paid as follows. If the integer chosen is 2, they all get 99 dollars each. Otherwise they get 100 dollars each. In this game, there are exactly 10 equilibria, where the players earn a positive amount of money, since any integer chosen by all players is an equilibrium, and in such an equilibrium they all earn either $99 or $100. Let me call each of these equilibria a 'good equilibrium'. Of no consequence here but it is worth noting that the game has other equilibria. Any choice of integers such that less than 9 players choose the same integer is also a Nash equilibrium. In such a situation all players earn 0 and there is nothing anyone can do individually to earn more.

Clearly, what the players will want is to converge on a good equilibrium—that is, has all of them choose the same integer. Of course, $100 would be ideal but $99 is not bad, and both much better than 0. But with so many good Nash equilibria which one will actually happen? In fact, there is a question as to whether a good equilibrium will happen at all, because players may fail to coordinate.

An intriguing answer can be constructed using an argument by Thomas Schelling in his seminal 1960 book. Schelling realized that while there is no 'hard' answer to questions of this kind involving multiple equilibria, there is an intuitive answer, based on human psychology, which is quite compelling. This is best understood by staring at the payoff matrix of the game just described. A truncated version of the game is shown in Table 3. This is a 2-player version of the above 10-player game. If both choose 2, they get $99; if both choose some other number but the same they get $100. If they choose different numbers they get 0.

All players want to choose the same integer, since any mismatch will mean they will all get 0. But how will they coordinate (no conversation is allowed)? One way to do so is to look for something

'salient' in one of those many equilibrium points. If there is a salient outcome and you have a strong hunch that the other players will also consider this salient, you would go for the salient outcome and hope that so will all others. That is the idea behind the concept of 'focal point'.

Table 3. A Coordination Game

	1	**2**	**3**	**…**	**8**	**9**	**10**
1	100	0	0	…	0	0	0
2	0	99	0	…	0	0	0
3	0	0	100	…	0	0	0
…	…	…	…	…	…	…	…
8	0	0	0	…	100	0	0
9	0	0	0	…	0	100	0
10	0	0	0	…	0	0	100

It is a powerful, if mysterious, idea rooted in humanity's common psychology. In the game just described, one choice by all players is special—the choice of 2. This is the only outcome where you are paid differently. You get 99. Everywhere else you get 100. It is likely that all players will realize this. So this is what you will go for and hope that so will the others.

It may be pointed out that this is a rather unusual focal point. It is the *lower* payoff that makes it salient. But most players will feel that a one dollar loss in order to be able to coordinate is well worth it. In other words, if this game were to be played in a laboratory, I feel confident, that virtually all players will choose 2.

Now suppose, with the waving of a magic wand, all the players become utilitarian. So they aggregate everybody's payoff and try to maximize that. Let us also suppose (a matter that was nobody's concern earlier in the selfish society) that this society has a poor bystander, whose payoff is as follows. If all ten players choose 2, he gets 11; if all choose 9, he gets 0; and for all other choices by the players he gets 1.

It is now easy to see that a new unanimity game emerges. The reader may wish to write up a payoff matrix, such as in Table 3, for this new game. If all players choose the same integer and the integer is different from 9, the payoff (the aggregate dollar earnings or aggregate utility) is 1001. If all players choose 9, the payoff is 1000. If all players do not choose the same integer the payoff is 1.

What will be the outcome in this game of utilitarians? By creating a payoff matrix like in Table 3 it becomes obvious that one strategy—every player choosing 9—is salient. This is the only Nash equilibrium with a different payoff. It is safe to predict that all players will choose 9, since, as before, not being able to coordinate leads to a disastrous outcome. In other words, while as selfish players they were choosing 2 where aggregate welfare was 1001, now, after they turn utilitarian and want to maximize aggregate welfare, they choose 9 and aggregate welfare drops to 1000.

The byproduct of this argument is that the bystander earns the lowest conceivable utility, namely, 1 dollar, by virtue of the players turning moral. This is a very different argument from the one used in Section 2 but, in essense, it is the same. It shows that players turning moral can prompt group immorality.

The importance of focal points in real-life is ubiquitous. We use it all the time. Consider a group of friends, who want to meet on a Sunday. Each has to choose a time to show up in the park. All they want is for all of them to be there at the same time. This is a game full of Nash equilibria. Any time, chosen by all, is a Nash equilibrium. We usually solve such problems by someone saying, “Since we are indifferent about when we meet, I suggest we meet at 6 pm.” All players nod, and leave. This interaction (the announcement and the nods) does not change the game and hence the Nash equilibria of the game. What it does is it creates a focal point. Everyone knows that everyone else will come to the park at 6 pm and everyone shows up at 6 pm. Indeed, a major purpose of language is to create focal points in order to solve coordination problems.

Of course, the idea of focal point has its fragilities, and we should be aware of it. In the above game, the focal point emerges by looking at the final payoffs (the aggregation of everybody’s utility). But all players can also see the components of this aggregate payoff. And this may make different players believe in different ‘focal points,’ which basically means there is no focal point. In short, there is no surefire way of knowing what is a focal point. We may need laboratory experiments to get a sense of what players do focus on and what kinds of outcomes are likely to be focal.

Further, it has been shown, for instance, that when the salience of the focal point derives from the labels used for the strategies, its definition can be fragile with respect to minor changes in payoff descriptions (Crawford, Gneezy and Rottenstreich, 2008).

5. The Meaning of a Game

Bringing game theory to help understand the challenge of group morality has the advantage of imposing a certain discipline on the discourse. The point is made strongly by Nguyen (2019). Game theory compels us to a precision that we may otherwise lack. Nguyen makes this out to be an advantage and, indeed, that is the spirit in which this paper has been written. I want, however, to close with a word of warning. All disciplines and, in fact, all discourse use explicitly stated assumptions or axioms, but they cannot avoid using implicit, unstated assumptions. What makes this a challenge is that those engaged in the discourse are often not even aware of them.

But before getting into that, let me ask what is the upshot of the discussion thus far (based on the previous four sections) concerning group moral responsibility? I believe we should recognize the fact that there are some bad group behavioral outcomes for which no moral responsibility can be attributed to anyone. Such a conclusion hurts us because most of us have an instinct whereby we like to apportion blame. But it would be folly to deny that there are collectively-committed bads for which we have no option but to treat them like earthquakes or meteor hits. We can lament the outcome but cannot apportion blame on anybody. In brief, some collective human behaviors are morally akin to natural calamities where no one has a hand. If the Samaritan's Curse was truly a 'game of life' (Binmore, 1998), which describes the ultimate game with no restrictions on individual choice beyond what arises from the laws of physics (Mailath, Morris and Postlewaite, 2017), then the two players and the bystander would be the only three persons on the planet, playing exactly the game described, with no other actions available to anyone. And in that case the outcome described in section 2 would be a foregone conclusion.

However, it will be equally wrong if we reacted to all bad group outcomes with resignation, treating them like meteor hits. Afterall, even with the Samaritan's Curse we know that the three persons in that society are not the only ones involved. The writer of this paper and the readers engaged in discussing the problem and trying to device solutions are a part of a larger game. This is in some ways contradictory. To describe a game of life and then talk about some characters who are not a part of the game and can come out with a solution amounts to describing the undescribable.

What I would argue is that this is an unavoidable problem of all theory. Some of our problems and paradoxes arise from assumptions that are hidden in the woodwork. There are everyday examples of this. Consider the case where your dinner host asks you if there is something you do not eat, and you say, "Just crab." Then at the party your host serves you a block of wood and you realize that the implicit universal sets in your heads, when you said that crab is the only thing you do not eat, were different. While to some this may appear to be a point of logical nitpicking that we cannot imagine happening in reality, if we replace 'block of wood' with 'grasshoppers' or 'dog meat' we can visualize actual scenarios where this happens.

A stark example of this comes from geometry. Euclid developed geometry laying out the axioms carefully. But there was at least one assumption that he was, in all likelihood, not even aware of. The assumption was that the entire exercise was being done on a 2-dimensional plane. His axioms (at least some of them) would not work if the same exercise was being done on a sphere, like the earth. It was the realization of this assumption in the woodwork in the 18th century that triggered the development of non-Euclidean geometry.

In all of economics, from general equilibrium theory to game theory, we start by specifying a feasible set of actions open to individuals. Thus in neoclassical equilibrium models we assume that, given the prevailing prices of goods and services, consumers can buy whatever their incomes allow. The set of all the things they can buy with their income is their 'feasible set'. Some conservative economists point out that as long as individuals are left free to choose from this set

without anybody else, such as the government, placing restrictions on their choice, society will achieve optimality. What they do not realize is that when we talk of all the things that consumers can do in neoclassical economics, we have already slipped in many assumptions without stating them and often without being aware of. Thus we do not start our analysis by assuming that the feasible set of actions open to consumers includes punching another consumer on the nose and running out of the store without paying. Indeed, if all these options were there to choose from, freedom to choose anything would typically not lead to an optimal outcome. The central result of neoclassical economics would then cease to be valid. This deep and unwitting use of norms and customary restrictions on individual choice is only now beginning to be recognized in economics (see Richter and Rubinstein, 2020). This is a method that opens up a large research agenda.

Let us return to the Samaritan's Curse. Here each agent is assumed to choose from among A, B, and C. In reality, an agent can also get up and say, "I propose that, since we are both moral creatures, we make an agreement to play A and not veer from that." Or, one player can make it clear that if the other player does not play A, she will punch the other player on the nose. Whether or not this will have any effect on actual play is an open question, but the point is, in standard game theory, we do not allow such actions, which lie outside the feasible set. There are also those outside of the game analyzing it, who could say this is how we should hold the players responsible, so that they change their behavior, which makes these outsiders a part of a larger game.

Opening up these avenues of interaction can open up the possibility of achieving a morally-good outcome. That, in turn, creates novel ways to hold individuals morally responsible for immoral outcomes. When we see the bad outcome, (C, C), occur in the Samaritan's Curse, some of us argue that the players are responsible for this brutish behavior. Why could neither of them show a little leadership and say that, since they both want to help the poor bystander, they should choose A and violators will be punished?

Economists and philosophers proposing such a resolution are often unaware that they are stepping beyond the game they began with, since they are proposing that the players should do what was never a part of the game. This tension of describing the game of life, specifying everything a player can do, and then saying that there is something else the player ought to do is a contradiction. But this may be the only way out of the problem of moral attribution in games such as the Samaritan's Curse. This is a seemingly contradictory suggestion, at least at this stage of our knowledge. What we need to do is to try and unearth some of the hidden assumptions of game theory, which would allow us to *legitimately* deal with more open-ended alternative possibilities which we as individuals have in reality, and which can solve the problem of moral creatures leading society to immoral outcomes.

References

Akerlof, R. (2016), 'The Importance of Legitimacy,' **World Bank Economic Review**, vol. 30.

Alger, I. and Weibull, J. (2013), '*Homo Moralis*—Preference Evolution under Incomplete Information and Assortative Matching,' **Econometrica**, vol. 81.

Arruda, C. (2017), 'How I learned to worry about the Spaghetti Western: Collective Responsibility and Collective Agency,' **Analysis**, vol. 77, No. 2.

Bacharach, M. (1999), 'Interactive Team Reasoning: A Contribution to the Theory of Co-operation,' **Research in Economics**, vol. 53.

Barrett, S. (1994), 'Self-enforcing International Environmental Agreements,' **Oxford Economic Papers**, vol. 44.

Basu, K. (1994), 'The Traveler's Dilemma: Paradoxes of Rationality in Game Theory,' **American Economic Review: Papers and Proceedings**, vol. 71.

Basu, K. (2000), **Prelude to Political Economy**, Oxford University Press, Oxford.

Basu, K. (2010), 'The Moral Basis of Prosperity and Oppression: Altruism, Other-regarding Behavior, and Identity,' **Economics and Philosophy**, vol. 26.

Basu, K. (2018), **The Republic of Beliefs: A New Approach to Law and Economics**, Princeton University Press, Princeton.

Bernheim, B. D. and Whinston, M. D. (1986), 'Common Agency,' **Econometrica**, vol. 54.

Binmore, K. (1998), **Just Playing: Game Theory and the Social Contract II**, MIT Press, Cambridge, MA.

Bjornsson, G. (2014), 'Essentially Shared Obligations,' **Midwest Studies in Philosophy,** vol. 38.

Braham, M. and Holler, M. J. (2009), 'Distributing Causal Responsibility in Collectivities,' in R. Gekker and T. Boylan (eds.), **Economics, Rational Choice and Normative Philosophy**, Routledge, New York.

Braham, M. and van Hees, M. (2012), 'An Anatomy of Moral Responsibility,' **Mind**, vol. 121.

Chant, S. R. (2015), 'Collective Responsibility in a Hollywood Standoff,' **Thought: A Journal of Philosophy**, vol. 4.

Chiao, V. (2014), 'List and Pettit on Group Agency and Group Responsibility,' **University of Toronto Law Journal**, vol. 64.

Chocker, H. and Halpern, J. Y. (2004), 'Responsibility and Blame: A Structural-Model Approach,' **Journal of Artificial Intelligence Research**, vol. 22.

Copp, D. (2006), 'On the Agency of Certain Collective Entities: An Argument from 'Normative Autonomy',' **Midwest Studies in Philosophy,** vol. 30.

Crawford, V., Gneezy, U. and Rottenstreich, Y. (2008), 'The Power of Focal Points is Limited: Even Minute Payoff Asymmetries may yield Large Coordination Failures,' **American Economic Review**, vol. 98.

Dickson, E. S., Gordon, S. C. and Huber, G.A. (2015), 'Institutional Sources of Legitimate Authority: An Experimental Investigation,' **American Journal of Political Science**, vol. 59.

Dufwenberg, M. and Kirchsteiger, G. (2019), 'Modelling Kindness,' **Journal of Economic Behavior and Organization**, vol. 167.

Dughera, S. and Marciano, A. (2020), 'Self-Governance, Non-reciprocal Altruism and Social Dilemmas,' mimeo: University of Paris Nanterre.

Feinberg, J. (1968), 'Collective Responsibility,' **Journal of Philosophy,** vol. 65.

Francois, P. and Zabojnik, J. (2005), 'Trust, Social Capital and Economic Development,' **Journal of the European Economic Association**, vol. 3.

Frankfurt, H. G. (1969), 'Alternative Possibilities and Moral Responsibility,' **Journal of Philosophy**, vol. 23.

Friedenberg, M. and Halpern, J. Y. (2019), 'Blameworthiness in Multi-Agent Settings,' Association for the Advancement of Artificial Intelligence: https://www.cs.cornell.edu/home/halpern/papers/group_blame.pdf

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

Gintis, H. (2015), 'A Typology of Human Morality,' mimeo: Santa Fe Institute.

Haji, I, and M. McKenna (2004), "Dialectical Delicacies in the Debate about Freedom and Alternative Possibilities," **Journal of Philosophy**, vol. 101.

Hakli, R., Miller, K. and Tuomela, R. (2010), 'Two Kinds of We-Reasoning,' **Economics and Philosophy**, vol. 26.

Held, V. (1970), 'Can a Random Collection of Individuals be Morally Responsible?' **Journal of Philosophy**, vol. 67.

Hess, K. (2014), 'Because They Can: The Basis for Moral Obligations for (Certain) Collectives,' **Midwest Studies in Philosophy**, vol. 38.

Hollis, M. (1994), 'The Gingerbread Game,' **Analysis**, vol. 54.

List, C. and Pettit, P. (2011), **Group Agency: The Possibility, Design, and the Status of Corporate Agents**, Oxford University Press, Oxford.

Mailath, G., Morris, S. and Postlewaite, A. (2017), 'Laws and Authority,' **Research in Economics**, vol. 71.

McAdams, R. (2000), 'A Focal Point Theory of Expressive Law,' **Virginia Law Review**, vol. 86.

Nguyen, C. (2019), 'Games and the Art of Agency,' **The Philosophical Review**, vol. 148, No. 4.

Parfit, D. (1984), **Reasons and Persons**, Clarendon Press, Oxford.

Pereboom, D. (2008), 'Defending Hard Incompatibilism Again,' in **Essays on Free Will and Moral Responsibility**, eds. Trakakis and Cohen, D. (eds.), Cambridge Scholars Press, Newcastle.

Pereboom, D. (2017), 'Responsibility, Regret and Protest,' **Oxford Studies in Agency and Responsibility 4** (D. Shoemaker, ed), Oxford University Press, Oxford.

Petersson, B. (2008), 'Collective Omissions and Responsibilities,' **Philosophical Papers**, vol. 37, No. 2.

Pettit, P. (2007), 'Responsibility Incorporated,' **Ethics**, vol. 117.

Posner, E. (2000), **Law and Social Norms**, Harvard University Press, Cambridge, MA.

Rabin, M. (1993), 'Incorporating Fairness into Game Theory and Economics,' **American Economic Review**, vol. 83.

Richter, M. and Rubinstein, A. (2020), 'The Permissible and the Forbidden,' mimeo. http://arielrubinstein.tau.ac.il/papers/The_Permissible_and_the_Forbidden.pdf

Rubinstein, A. (1998), **Modeling Bounded Rationality**, MIT Press, Cambridge, MA.

Runciman, W. and Sen, A. (1965), 'Games, Justice and the General Will,' **Mind**, vol. 74.

Samuelson, L. (2016), 'Game Theory in Economics and Beyond,' **Journal of Economic Perspectives**, vol. 30.

Sartorio, C. (2004), 'How to be Responsible for Something without Causing It' **Philosophical Perspectives**, vol. 18.

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

Sen, A. (2018), **Collective Choice and Social Welfare**, Expanded edition, Oxford University Press, Oxford.

Singer, M. (2006), 'Legitimacy Criteria for Legal Systems,' **Kings College Law Journal**, vol. 17.

Sugden, R. (2000), 'Team Preferences,' **Economics and Philosophy**, vol. 16.

Sunstein, C. (1996), 'On the Expressive Function of Law,' **University of Pennsylvania Law Review**, vol. 144.

Tannsjo, T. (2007), 'The Myth of Innocence: On Collective Responsibility and Collective Punishment', **Philosophical Papers**, vol. 36.

Tessman, L. (2015), **Moral Failure: On the Impossible Demands of Morality**, Oxford University Press, Oxford.

Tuomela, R. (2006), 'Joint Intention, We-mode and I-mode,' **Midwest Studies in Philosophy**, vol. 30.

Tyler, T. (2006), **Why People Obey the Law**, Princeton University Press, Princeton.

Vallentyne, P. (1989), 'Contractarianism and the Assumption of Mutual Unconcern,' **Philosophical Studies**, vol. 56.

Voorneveld, M. (2010), 'The Possibility of Impossible Stairways: Tail Events and Countable Player Sets,' **Games and Economic Behavior**, vol. 68.

Zaibert, L. A. (2003), 'Collective Intentions and Collective Intentionality,' **American Journal of Economics and Sociology**, vol. 62.