

Gauthier, Equilibrium, and the Emergence of Morality

BRETT MULLINS *Miami University*

ABSTRACT: David Gauthier develops morality in the social contract tradition as an emergent property rationally necessitated by the presence of inefficiency. To demarcate situations in which morality arises from those in which it does not, two principles, Strategic Emergence and Market Emergence, are motivated and assumed by Gauthier to be equivalent. Following the work of Bob Bright, this paper formalizes and expands upon a demonstration of the inconsistency of the two principles. Eliminating each of the emergence conditions is considered to resolve the inconsistency. Additionally, the Kantian equilibrium is examined in place of the Nash equilibrium; however, Gauthier's approach resists such amendments.

RÉSUMÉ : David Gauthier présente la moralité dans la tradition du contrat social en tant que propriété dont l'émergence est rendue rationnellement nécessaire par la présence de l'inefficacité. Pour distinguer les situations dans lesquelles la moralité apparaît de celles où elle n'apparaît pas, Gauthier présente deux principes réputés équivalents, l'émergence stratégique et l'émergence de marché. Cet article formalise et poursuit, à la suite des travaux de Bob Bright, la démonstration du caractère contradictoire de ces deux principes. Je considère que la contradiction peut être résolue en éliminant chacune des conditions d'émergence. De plus, j'examine la possibilité de remplacer l'équilibre de Nash par l'équilibre kantien; l'approche de Gauthier, cependant, ne souffre pas de ce type de modification.

Keywords: formal ethics, David Gauthier, equilibrium, strategy, contractarian

Dialogue 55 (2016), 677–693.

© Canadian Philosophical Association/Association canadienne de philosophie 2017

doi:10.1017/S0012217316000627



In *Morals by Agreement*, David Gauthier aims to provide an account of morality divorced from the language of objective moral values and reduced entirely to the terms of strategic and parametric choice. On Gauthier's view, morality is an emergent property rationally necessitated by the presence of inefficiency, when the employment of individual utility maximizing strategies fails to produce optimal collective outcomes. This account is critiqued by Bob Bright in "The Poverty of Market Contractarianism" insofar as an inconsistency is established between Gauthier's supposedly equivalent bases for the emergence of morality: Strategic Emergence and Market Emergence. In this paper, I will formalize, expand upon, and challenge Bright's criticism to demonstrate an inconsistency within Gauthier's Market Contractarianism. In the first section, the relevant features of Gauthier's approach are introduced formally. In the second, I present a case used by Bright on which Gauthier's program yields contradictory conclusions. In the third, eliminating either of the emergence principles as a potential resolution is critically considered. In the fourth, Kantian equilibrium is evaluated as an alternative solution concept to Nash equilibrium with respect to the Market Contractarian program.

1. Market Contractarianism

Gauthier conceives of morality in the tradition of social contract theory, whereby "moral justification rests on the agreed, impartial, and rational constraint [of an agent's behaviour] for mutual benefit."¹ Market Contractarianism is the position in which moral justification is rationally motivated by purely instrumental considerations.² Consider the case of parametric choice: an agent is confronted with a choice among several outcomes. The equilibrium is the outcome that represents the greatest utility to the agent, i.e., the outcome that is most consistent with the agent's preferences. The outcome of a choice is Pareto optimal if no other possible outcome would make at least one party better off while making no others worse off (henceforth, the use of Pareto is suppressed). The agent's choice is necessarily efficient in the sense that the equilibrium outcome must also be optimal in the case of parametric choice; otherwise, the agent would choose contrary to his or her preferences and would, hence, choose irrationally.

Suppose instead that the choices of one agent affect the choices of other agents; this represents strategic choice. Under this latter condition, the equilibrium outcome need not coincide with the optimal outcome. The Prisoner's Dilemma game is offered as the paradigm instance.³ A standard presentation of the game proceeds: two alleged criminals, P1 and P2, are arrested, and each is offered a plea bargain to testify against the other. If each rejects the bargain,

¹ Voice (2002, 1).

² Gauthier (1991, 16).

³ Gauthier (2013, 606).

then they will both be convicted of minor crimes and serve short sentences. If, instead, both accept, then both will serve harsh punishments. If one criminal accepts but not the other, then the accepting criminal walks free, while the rejecting criminal faces the maximum punishment. This game is presented in strategic or matrix form in **Figure 1** where the outcomes are listed in ordinal utility terms.

		P2	
		Reject	Accept
P1	Reject	(3,3)	(1,4)
	Accept	(4,1)	(2,2)

Figure 1 Prisoner's Dilemma

Not knowing how the other will choose, each criminal assesses the possible outcomes. By choosing to reject, one faces outcomes ranked 3 and 1, depending on whether the other accepts or rejects, respectively, and, by choosing to accept, one faces outcomes ranked 4 and 2. Irrespective of what the other chooses, one is at least as well off in all outcomes and is better off in at least one outcome by accepting the plea bargain; hence, the 'accept' strategy dominates the 'reject' strategy. Generalizing to both criminals yields an equilibrium outcome where both accept. This result is significant insofar as there exists an outcome, *both reject*, where both prisoners are better off when compared to the equilibrium outcome, *both accept*.

In such a case, Gauthier claims that it is rational for agents to adopt a disposition to constrain individual utility maximizing choice behaviour to achieve optimal outcomes. As summarized by Bright, "Gauthier locates room for morality in the divergence between equilibria and optimality."⁴ As his paper in this issue of *Dialogue* confirms, Gauthier continues to think that this divergence, which he now calls 'the interaction problem,' creates the conditions necessary for rational morality to emerge and become established.

Proposition 1 (Strategic Emergence): Morality emerges just in case the equilibrium strategy is not optimal.

Morality, then, is identified with the adoption of a disposition for cooperation in cases in which the maximization of individual utilities produces sub-optimal outcomes.

⁴ Bright (2000, 354).

Under conditions of efficiency, there exists no instrumental reason for an agent to constrain his or her choice behaviour for the sake of cooperation. When the optimal strategy is in equilibrium, either necessarily with parametric choice or potentially with strategic choice, morality does not apply. This is not to say that such cases are immoral; rather, efficient outcomes are amoral. For the purposes of this analysis, Gauthier's distinction between the morality or immorality of a particular act is beside the point. We are concerned here with the demarcation between the domain of morality and amorality.⁵

Theorem 1: Any case of strategic interaction either is in the domain of morality or is amoral, exclusively.

The above theorem requires qualification; it is possible that for some game no equilibrium outcome exists. We must, then, appeal to the well-established result that a game with finitely many players, each with finitely many choices, necessarily has at least one equilibrium strategy.⁶

Theorem 1' (Decidability): Any case of finite strategic interaction either is in the domain of morality or is amoral, exclusively.

This result is adequate for the Market Contractarian, since the finitude of agents is a practical limitation.

Let φ be a function that maps instances of finite strategic interaction to either the domain of morality or amorality. As φ applies to all cases of finite strategic interaction, φ is a total function. Let ψ be a function that maps instances belonging to the domain of morality to either moral or immoral, given the constrained maximization criteria set forth by Gauthier. To speak of the domain of morality or the emergence of morality is to indicate the domain of the function ψ , resulting from the total function φ . Note that, while φ and ψ are distinct functions, representing distinct notions, ψ depends upon φ to determine its domain.

Gauthier introduces the notion of the perfectly competitive market as a limiting case for market interaction.

Proposition 2 (The Perfectly Competitive Market) (PCM): is a situation in which agents have perfect information regarding prices and no agent's consumption or production decision affects the prices of goods. In the PCM, outcomes are necessarily efficient.

In the PCM, aggregate production and consumption are determined independently of any particular agent; agents are referred to as 'price takers' as opposed

⁵ Voice (2002, 15).

⁶ For the sake of simplicity, no distinction is made here between pure and mixed strategies. See Luce and Raiffa (1957, 106).

to 'price makers.' This to say that the choice behaviour of a single agent does not affect the choices of other agents, either producers or consumers, effectively reducing choice in the PCM to parametric. Recall that the equilibrium in parametric choice is necessarily optimal. As a consequence of Strategic Emergence and PCM,

Theorem 2 (Moral Anarchy): Morality does not emerge in the PCM.⁷

Gauthier argues that the PCM "guarantees the coincidence of equilibrium and optimality, and so its structure is the very antithesis of the Prisoner's Dilemma;" hence, it constitutes a "morally free zone ... beyond good and evil."⁸ In other words, the PCM is not in the domain of morality.

Market failure occurs when the competitive equilibrium produces a sub-optimal outcome, thereby diverging from a state outside of the domain of morality. To capture this inefficiency, Gauthier offers the following:

Proposition 3 (Market Emergence): Morality emerges just in case market failure obtains.

For an illustration of the intuition of Market Emergence, Gauthier offers "where the invisible hand fails to direct each person, mindful only of her own gain, to promote the benefit of all, co-operation provides a visible hand."⁹

To conclude this section, two points must be noted. First, Moral Anarchy additionally follows from Market Emergence and the PCM. This is to say that in this instance Strategic Emergence and Market Emergence are in agreement. Second, the notion of equilibrium in Strategic Emergence is distinct from that in Market Emergence, though they need not be contrary, of course.¹⁰ The former is Nash equilibrium, the utility maximizing response given the anticipated choices of others. This notion of equilibrium, named after John Nash, Jr., is the standard solution concept of classical game theory. Pertinent to the present analysis, Nash equilibrium, and the game theoretic framework more generally, may be applied to isolate instances of strategic interaction from a broader context. For instance, in some cases, the strategic behaviour of two competing firms may be analyzed without considering the particular market as a whole or all markets in conjunction.

The latter notion is Walrasian equilibrium or Competitive equilibrium, an aggregation of individual consumption decisions coordinating productive efforts to allocate goods throughout the whole of an economy. This view

⁷ Moral Anarchy is a highly contested position. See Hausman (1989).

⁸ Gauthier (1986, 83-113).

⁹ Gauthier (1986, 113).

¹⁰ Ghosal and Polmarchakis (1997, 31-32).

of equilibrium, named in honour of the French economist Léon Walras, was largely developed by Kenneth Arrow and Gérard Debreu and forms a core of microeconomic theory, alongside game theory. The Walrasian equilibrium represents the consumption bundle of all agents in an economy at some level of prices. An allocation is said to be efficient if no reallocation of goods makes at least one agent better off and no other agents worse off. Note that efficiency does not imply equitability; an equilibrium allocation in which a single agent possesses all goods is efficient since any redistribution would harm that agent. Note that this model is rather restrictive insofar as agents interact only through the price mechanism. In contrast, the language of game theory leaves the terms of interaction for the theorist to specify.

2. Bright's Criticism

Gauthier focuses his analysis almost exclusively on externalities, whereas market failure obtains in a variety of conditions. An externality “arises whenever an act of production or exchange or consumption affects the utility of some person who is not party, or who is unwillingly party, to it.”¹¹ Consider an idealized example in which a manufacturing plant pollutes its surrounding area. The plant produces goods for cost X and receives Y benefits by selling goods such that Y is greater than or equal to X ; otherwise, it would not produce. In manufacturing these goods, the air pollution by-product harms the local environment at Z per good produced. If the total cost of producing a good, $X + Z$, outweighs the cost realized by the plant, X , then an externality, Z , is present. Irrespective of the value of Z , the plant will still produce goods if the benefits received are greater than (or equal to) the realized costs.

The pollution case above exemplifies a negative externality or, in Gauthier's terms, “parasitic behaviour” on the part of the manufacturing plant; the plant is not realizing the full cost of production and, resultantly, overproduces. Alternatively, positive externalities occur with the presence of “free riders.”¹² In this case, one does not realize the full benefit of the equilibrium outcome and will, hence, underproduce or underconsume. In the presence of externalities, efficient market outcomes do not necessarily obtain; in general, the Walrasian equilibrium is sub-optimal.¹³ Outside of the PCM, it may be the case that the utility derived from consumption or production is interrelated among agents. This occurs in a way that is more subtle in comparison to instances such as the Prisoner's Dilemma.

¹¹ Gauthier (1986, 87).

¹² Gauthier (1986, 96-97).

¹³ Kreps (1990, 202-205).

Bright's contribution to this analysis begins with the following case:

Proposition 4 (Housing Externality Case): There exists a case defined in finite game theoretic terms which induces Market Failure but the Nash equilibrium strategy is optimal.

Consider an instance in which a homeowner, H, deliberates which grade of siding to add to his or her house. As the grade increases, the marginal benefit gained diminishes, i.e., increases at a decreasing rate, and the marginal cost expended per unit remains constant. By performing this upgrade, H's neighbour, N, gains utility proportional to H's increase in utility through the increase in the value of a house in close proximity to N's property. If H chooses to upgrade his or her siding to any grade, then H will produce a positive externality realized by N.¹⁴

At the point where the marginal cost equals the realized marginal benefit for H, the particular siding grade G_n is the equilibrium outcome but is potentially sub-optimal. Consider the extensive form game depicted in **Figure 2** below:

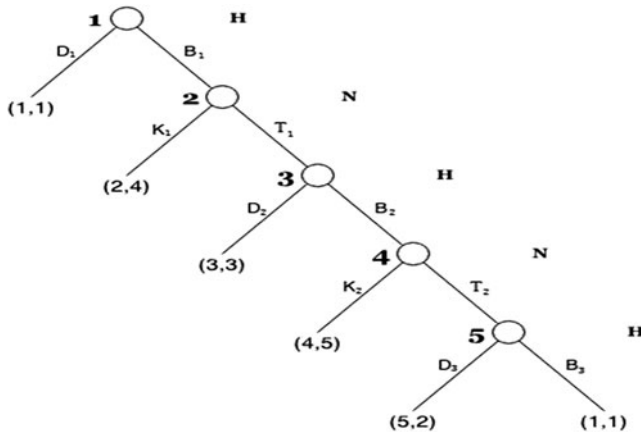


Figure 2 Extensive Form of the Housing Externality Case

Each node in the game tree corresponds to a choice to be made by the player indicated to the right of the node. Outcomes are represented as pairs of respective ordinal utility rankings such that H is the abscissa and N is the ordinate. If a player chooses an outcome represented by utilities, then the game ends; otherwise, the game continues if another node is reached. Node 1 represents H's choice where H either buys (B_1) or does not buy (D_1) grade G_n . At Node 2, N chooses to either transfer (T_1) a share of the positive externality to H in the

¹⁴ Bright (2000, 355-357).

form of a payment or keep (K_1) the whole of the benefit; let us assume that the amount N would transfer is such that G_{n+1} would be the equilibrium grade for H . At Node 3, H chooses whether to keep N 's transfer or to purchase the better grade of siding. Node 4 repeats N 's choice to either transfer to or keep some amount from H ; while the potential size of the transfer is greater than at Node 2, let us assume that the amount required to bring G_{n+2} into equilibrium for H exceeds the amount of the externality. At Node 5, H , again, chooses whether to purchase the next grade of siding.

By backward induction, the equilibrium strategy, $B_1T_1B_2K_2$, yields utilities (4, 5) for H and N , respectively. Intuitively, this reveals that N will only transfer a share of the externality to H in return for H purchasing a higher siding grade. **Appendix A** presents a numerical illustration of the Housing Externality Case. As no other outcome can make one individual better off without making another worse off, the equilibrium strategy is optimal; hence, this strategy is efficient. The following is a consequence of Strategic Emergence and the Housing Externality Case:

Theorem 3: The Housing Externality Case is not in the domain of morality.

Despite Gauthier's claims to the contrary, with the Housing Externality Case, Bright successfully illustrates that the presence of externalities does not necessarily result in the emergence of morality. If this is the case, then Gauthier is unjustifiably concerned with so-called "free riders" and "parasites." This is Bright's conclusion: either Market Contractarianism fails to designate all instances of free riders and parasites as immoral, or these excluded instances require no intervention from morality.¹⁵

This result opposes Gauthier's assessment, as he argues "when markets fail, each person, seeking to maximize her utility given the strategies she expects others to choose, fails to maximize her utility given the utilities those others receive."¹⁶ This is to say that Bright's treatment does not take seriously Gauthier's commitment to the immorality of free riders and parasites. Gauthier's case against externalities and market failure surely relies on Market Emergence. As Walrasian equilibria are not optimal in the presence of externalities, as developed above, Market Emergence and the Housing Externality Case imply the following:

Theorem 4: The Housing Externality Case is in the domain of morality.

Note that the outcome of the Housing Externality Case potentially affects agents other than H and N ; however, by construction, these agents are not in a

¹⁵ Bright (2000, 356).

¹⁶ Gauthier (1986, 116).

position to interact strategically with either H or N. Resultantly, they are excluded from the game theoretic specification, but the externality they receive potentially distorts their consumption decisions nonetheless. Moreover, the effect of the positive externality is non-excludable, i.e., H cannot choose who receives the benefit from the increased siding, so results similar to the Coase Theorem are not relevant.¹⁷

Though Theorems 3 and 4 are *prima facie* inconsistent, it has yet to be demonstrated that they are logically independent. The Housing Externality Case suffices to illustrate the inequality of Strategic Emergence and Market Emergence. The possibility remains that Market Emergence implies Strategic Emergence, i.e., the states of the world in which morality emerges through Strategic Emergence is a subset of Market Emergence. With this being said, Strategic Emergence is a more expansive notion than Market Emergence as it applies to instances of isolated strategic choice, where the notion of Walrasian equilibrium is not relevant.

In illustrating this apparent inconsistency between Theorems 3 and 4, we may now conclude the following:

Theorem 5: There exists a case of finite strategic interaction that both is and is not in the domain of morality.

This result is in contradiction with Decidability, which holds that any instance of finite strategic interaction is either moral or amoral, exclusively. It is this difficulty that is insurmountable for the present form of Gauthier's Market Contractarianism. While the Housing Externality Case may appear contrived, it is sufficient to demonstrate that this system does not satisfactorily handle every case of finite strategic interaction. Moreover, this approach fails to handle this case by issuing inconsistent verdicts, rather than no verdict either way, on the emergence of morality.

3. Potential Resolutions

Taken together, the four propositions that compose Gauthier's approach to demarcating the domain of morality from amorality imply an impossibility; hence, at least one of the propositions must be false. As both the PCM and the Housing Externality Case merely present morally interesting situations, from the Market Contractarian perspective, in the terms of economic analysis, the two emergence principles are resultantly suspect. In an attempt to both resolve the contradiction and preserve Gauthier's program, the prospect of eliminating each Emergence principle will be considered in turn.

¹⁷ The Coase Theorem says that, if property rights are complete and negotiating is costless, then agents will bargain to obtain an efficient solution whenever an externality is present. While we assume that negotiating is costless, notice that property rights are not complete. See Coase (1960).

3.1. *The Exclusion of Market Emergence*

Suppose Market Emergence is excluded from the formal characterization. The sub-optimality of Nash equilibrium is, then, the sole criterion for the emergence of morality. Following Bright, market failure need not imply morality, and, in some cases of externalities, at least with the Housing Externality Case, morality does not arise. This is consistent with Gauthier's primary claim that morality does arise in instances structured similar to the Prisoner's Dilemma, where the Nash equilibria are sub-optimal.

While this approach resolves the contradiction between Theorem 5 and Decidability, as Theorem 5 no longer follows, the amorality of some cases of externalities violates the condition of impartiality, a motivation for Market Emergence. Impartiality is the notion that no individual or group should be differentially affected, commensurate to their contribution, by an outcome, either favourably or unfavourably. For instance, in the context of society as a cooperative venture, this leads Gauthier to assert that "those who make no net contribution, then, are entitled to no net benefits."¹⁸ Gauthier treats impartiality as basic to account for traditional moral intuitions, similar to the approaches of both John Rawls and John Harsanyi.¹⁹ While partiality is sufficient for immorality, impartiality is a feature of both moral and amoral cases, and, hence, impartiality alone is not sufficient to distinguish between the two cases. To exemplify this, both the PCM and the outcome of constrained maximization exhibit impartiality; however, the former is the exemplar case of amorality, while the latter is the instantiation of morality.

Appealing to Theorem 3, we find that with Strategic Emergence the Housing Externality Case is not in the domain of morality. Given the explication in the preceding paragraph, this is a rather unanticipated result, since the presence of an externality implies differential treatment, even if such treatment is to the benefit of another, namely N. There is a sense in which N did not contribute but is receiving benefits; yet, morality does not emerge. This result can be made stronger by supposing that the situation is structured such that N opts not to transfer a partial sum of the benefit to H. If H purchases the siding grade that is optimal, then this outcome is still efficient and, hence, not in the domain of morality. If Strategic Emergence is the single principle that determines the domain of morality, then there exists at least a single case in which partiality is insufficient for immorality.

At this point, one might claim that, in eliminating Market Emergence, there is no reason not to also exclude the notion of impartiality. If this is the case, then the Market Contractarian approach with Strategic Emergence survives intact.²⁰ Gauthier, however, insists that a rejection of impartiality is equivalent to

¹⁸ Gauthier (1998, 124).

¹⁹ Gauthier (1986, 4-6).

²⁰ I am indebted to Jeffrey Dunn for raising this question.

rejecting the Hobbesian approach to the Social Contract insofar as impartiality implies that the involved parties will recognize the bargaining positions of one another equally.²¹ Most notably, without impartiality, instances of coercion need not be ruled out. As an extreme case, suppose one is robbed at gunpoint. Faced with the choice between potential death and parting with one's goods, one can imagine that conceding to the robber's demand is an efficient outcome. By Strategic Emergence, then, this and other coercive instances are not in the domain of morality, a rather counterintuitive and unwelcome result.

From this, it is clear that impartiality is indispensable to Gauthier's approach inasmuch as this theory is intended to accord with traditional moral intuitions, i.e., affirms common sense solutions to moral dilemmas. In the initial formal characterization, it was not necessary to specify impartiality, since the notion is captured by Market Emergence. Any instance of differential treatment, relative to contribution in an agreement, can be described as an instance of an externality and, hence, is within the purview of Market Emergence. The notion of impartiality distinguishes a theory based upon mutual advantage from the one which Gauthier supports that is, instead, based upon rational mutual agreement, i.e., an instance preferred by a theory of rational mutual advantage may not be preferred by a theory of rational mutual agreement, as evidenced above.²² Since Strategic Emergence, alone, violates impartiality, this characterization must be rejected as untenable given the commitments of Gauthier's program.

3.2. *The Exclusion of Strategic Emergence*

Suppose, instead, that Strategic Emergence is excluded from the formal characterization. With this approach, the emergence of morality occurs *only* when the Walrasian equilibria are not optimal. This is to say that, in general, the PCM exclusively constitutes amorality. In the PCM, recall that choice is parametric. If an agent is affected differentially with respect to his or her contribution by the choices of other agents, then the PCM does not obtain and, hence, morality emerges. Since agents are not differentially affected when morality does not apply, the impartiality condition is preserved. Likewise, this approach justifies Gauthier's focus on externalities. As the presence of an externality implies sub-optimality, in general, morality emerges in the Housing Externality Case as the neighbour N is a "free rider."

The difficulty with relying solely on Market Emergence is that the Prisoner's Dilemma is no longer necessarily in the domain of morality. In the absence of Strategic Emergence, the language of game theory loses relevance. This is to say that isolated, individual instances of strategic interaction cannot be evaluated solely with the notion of Walrasian equilibrium. This is problematic for

²¹ Gauthier (1998, 122).

²² Gauthier (1998, 123).

Gauthier, since the Prisoner's Dilemma is presented as the paradigm case of an instance that necessitates the intervention of morality. If the designation of the Prisoner's Dilemma as belonging to the domain of morality is to be preserved, then one must construct it and similar instances in the language of prices, initial endowments of goods, and Walrasian equilibria.

Gauthier denotes the intervention of morality upon instances within the domain of morality as constrained maximization. This approach, too, is defined in terms of game theory and strategic interaction.²³ As a result, this strategic construction of constrained maximization is no longer relevant to Market Contractarianism. The upshot is that Gauthier later abandons this position resulting from advances in bargaining theory, specifically those by Ariel Rubinstein.²⁴ While the apparent loss of the Prisoner's Dilemma and constrained maximization are not conceptual defeaters to this approach, Gauthier's case is clearly weakened.

With Market Emergence determining the domain of morality, in general, only the PCM is amoral. While other efficient Walrasian equilibria potentially exist, they are not generalizable. Since morality acts to correct for inefficiencies, agents constrain their behaviour to obtain efficient allocations. Hereafter, we are concerned with the PCM and deviations from this case for the sake of generality. Let us explore conditions that violate PCM and, hence, potentially lead to inefficient allocations.

In the PCM, it is assumed that all agents have knowledge of current prices and all other relevant information. In the absence of full information, inefficiencies can arise. It is worth noting that the game theoretic framework assumes perfect information as well. Interestingly, relying only on Market Emergence suggests that morality will lead individuals to correct for information asymmetries and other violations of perfect information. This is to say that imperfect information is within the purview of morality. Given Gauthier's commitment to traditional moral judgments and recognizably moral theories as illustrated by the condition of impartiality, this is an unwelcome result. In addition to externalities, as developed in Section 2 above, imperfect markets result in inefficient allocations. Imperfection in markets occurs when an agent's consumption or production affects the price of a good. This is significant, since, in the PCM, choice is assumed to be parametric, i.e., consumption by one agent does not affect the consumption of another. Note that the present violations of the PCM are not meant to be exhaustive.

As should be obvious to the reader, this characterization led by the Market Emergence principle is not the sort of theory that Gauthier defends throughout *Morals by Agreement* and other work. The consequences are counterintuitive, which is of theoretical importance to Gauthier, and the exclusion of Strategic

²³ Gauthier (1986, 67-70).

²⁴ Gauthier (2013, 609-610).

Emergence both limits the strength of the approach and nullifies much of constrained maximization. Resultantly, this alternative formalization of Market Contractarianism ought to be rejected as unacceptable.

4. Kantian Equilibrium

In the context of Market Contractarianism, Nash equilibrium has been supposed to be the appropriate game theoretic solution concept. My purpose in this section is to consider a proposed alternative to Nash equilibrium: Kantian equilibrium. This solution concept, developed by John Roemer, is constructed to solve the inefficiencies of collective action, such as the Prisoner's Dilemma and related multi-agent games. Given that the same sort of instances motivate the relevant work of both Roemer and Gauthier, a natural extension is to consider their ideas in tandem. Following a brief introduction to the Kantian equilibrium, Gauthier's Market Contractarianism is re-examined with respect to Kantian equilibrium as the foundational game theoretic solution concept.

Consider a motivating example similar to the Prisoner's Dilemma: the Tragedy of the Commons.

A set of fishers must expend labor on a lake to catch fish, but there is a congestion problem, so the fish caught per unit of time decreases with the number of total hours expended in fishing by the community. Each fisher has a utility function over fish caught and labor expended. In the Nash equilibrium of the game where each fisher's strategy is a labor choice, there is over-fishing: the equilibrium is Pareto inefficient, and everyone would gain in welfare from a small decrease in labor expenditures. Some kind of cooperation is necessary to solve the problem.²⁵

With the Tragedy of the Commons, the common strategy space is represented by the labour expended. Let L_i be the labour expended by the i th fisherman and let L be the vector containing L_i for all i . Suppose further that α is a real valued scalar. Roemer defines the Kantian equilibrium as an L vector, i.e., a set of strategies, such that there is no α where everyone prefers αL to L . This is intended to capture the universalizability aspects of Kant's Categorical Imperative. Roemer, then, demonstrates existence and efficiency results for Kantian equilibrium.²⁶

Unlike with the Tragedy of the Commons, the strategy space in the Prisoner's Dilemma is binary. To analyze this instance, we will consider mixed strategies, denoted by p , where p is the probability that one rejects the plea bargain. In this case, however, the Kantian equilibrium does not guarantee an optimal outcome, i.e., both reject. The equilibrium outcome depends upon the relative cardinal utility gains from both deceiving the other (accept-reject) and being

²⁵ Roemer (2010, 1).

²⁶ Roemer (2010, 4-5).

deceived (reject-accept).²⁷ Though Kantian equilibrium only potentially provides efficient outcomes, it weakly dominates Nash equilibrium insofar as, in this case, Nash equilibrium never produces efficient outcomes.

With the introduction of Kantian equilibrium, notice that Decidability no longer follows from Strategic Emergence. Recall that, for any game with finitely many players, each with finitely many strategies, there exists a Nash equilibrium strategy. Roemer, however, concedes that, at present, the Kantian equilibrium lacks generality insofar as one must specify the strategy space for each case.²⁸ In instances in which agents do not share a common strategy space, such as the Housing Externality Case, the Kantian equilibrium is not defined. This is to say that Strategic Emergence no longer denotes the Housing Externality Case as amoral, and, hence, it is not the case that every instance of finite strategic interaction either is in the domain of morality or is amoral, exclusively.

The demonstrated inconsistency fails to follow on both fronts. In addition to the lack of Decidability, Theorem 5 no longer follows. With this being said, Market Emergence still denotes the Housing Externality Case as in the domain of morality. Though the contradiction is resolved, notice that the resulting Market Contractarian theory is significantly weakened. Whereas, previously, Gauthier's approach applied both to all isolated instances described in terms of finite strategic interaction and to all instances of markets in terms of Walrasian equilibrium, the adoption of Kantian equilibrium applies to isolated instances only insofar as the players share a common strategy space.

Assuming that no inconsistencies arise between the modified Strategic Emergence and Market Emergence, the resulting approach is only slightly improved from the case present in 3.2. This is to say that, in general, for instances where Kantian equilibrium is not applicable, morality will be required in every instance outside of the PCM such that the moral act is to partake in the joint strategy where the PCM obtains. It follows that this approach ought to be rejected for similar reasons.

A question beyond the scope of the present paper that is worth briefly exploring is by which criteria one should choose between Nash equilibrium, Kantian equilibrium, or alternative notions, more generally, as the foundational solution concept for models of strategic interaction. Roemer offers evidence from anthropology and evolutionary biology regarding the possibility of altruism to explain behaviour.²⁹ It is unclear, however, that such empirical evidence is relevant for a normative, as opposed to a positive or descriptive, game and decision theory. Inasmuch as Kantian equilibrium is to be accepted in place of Nash equilibrium, Kantian equilibrium, in its present construction,

²⁷ Roemer (2010, 11-12).

²⁸ Roemer (2010, 2).

²⁹ Roemer (2014, 1-2).

is theoretically weaker relative to the Market Contractarian approach. This does not rule out possible developments and suggest an avenue for future research.

5. Concluding Remarks

The poverty of Gauthier's Market Contractarianism is that the two bases for the emergence of morality, Strategic and Market, are taken to be equivalent when, in fact, they are inconsistent. As Prisoner's Dilemma type situations necessitate morality through Strategic Emergence; when market conditions diverge from the PCM, it is assumed that inefficient outcomes obtain. This assumption, implicitly developed and critiqued throughout this paper, is a source of the difficulties with Gauthier's approach. As noted earlier, Walrasian equilibria and Nash equilibria are distinct notions and, as illustrated by the Housing Externality Case, yield opposing conclusions in some cases. As a result of this inconsistency, it cannot be the case that the relevant features and the results of Gauthier's approach are jointly preserved. If Market Emergence is excluded, then impartiality is violated; however, if Strategic Emergence is excluded, then every instance outside of the PCM is in the domain of morality. As Market Contractarianism resists the present attempts to resolve the demonstrated inconsistency, Gauthier's approach fails to rationally motivate morality.

Appendix A: Numerical Simulation of Housing Externality Case

Suppose that H's current grade of siding is G_0 and that siding can be upgraded in whole increments whose total cost is given by the function $\$(G_n) = \$10n$. This is to say that the marginal cost of an upgrade in siding is \$10. Let H's willingness to pay for siding be \$99.50, and suppose that the externality received by N is given by the function $E_n = 0.1(\$(G_n))$. **Table 1** presents the dollar values for each possible outcome of the game tree depicted in Figure 2:

Table 1 Expected Monetary Benefit

Outcome	H's Expected Monetary Benefit	N's Expected Monetary Benefit
B_1K_1	\$90	\$9
$B_1T_1D_2$	\$90.50	\$8.50
$B_1T_1B_2K_2$	\$100	\$9.50
$B_1T_1B_2T_2D_3$	\$109	\$0.50

Note that outcomes denoted as (1,1) are excluded from Table 1 since both H and N receive \$0 as H opts not to upgrade from his or her current siding G_0 .

At Node 1, H purchases his or her individual optimal grade of siding G_9 for \$90. N receives a \$9 positive externality. At Node 2, N transfers \$0.50 to H. At Node 3, H upgrades to G_{10} and N receives an addition \$1 positive externality.

No transferred amount of N's accrued benefit will bring G_{11} into equilibrium for H; hence, N chooses to keep the whole of the \$9.50 at Node 4 and the game ends. Note that the particular values used are arbitrary. If the amount appears to be insufficiently large, then the dollar values may be scaled upward by any positive linear transformation to receive the same result. In addition, we may modify the case to accommodate a continuum of siding grades up to the nearest cent. As before, we may scale the amounts upward to approach any situation with arbitrary accuracy.

Acknowledgements: I am indebted to Virginia Baxter for substantive and constructive comments on early drafts of this paper.

References

- Bright, Bob
2000 "The Poverty of Market Contrarianism." *The Journal of Value Inquiry* 34 (2–3): 349–357.
- Coase, Ronald
1960 "The Problem of Social Cost." *Journal of Law and Economics* 3: 1–44.
- Gauthier, David
1986 *Morals by Agreement*. Oxford: Clarendon Press.
- Gauthier, David
1991 "Why Contractarianism?" In *Contractarianism and Rational Choice: Essays on David Gauthier's Morals by Agreement*, edited by Peter Vallentyne, 15–30. New York: Cambridge University Press.
- Gauthier, David
1998 "Mutual Advantage and Impartiality." In *Impartiality, Neutrality and Justice*, edited by Paul Kelly, 120–136. Edinburgh, UK: Edinburgh University Press.
- Gauthier, David
2013 "Twenty-Five On." *Ethics* 123 (4): 601–624.
- Ghosal, Sayanta, and H. Polemarchakis
1997 "Nash-Walras Equilibria." *Research in Economics* 51 (1): 31–40.
- Hausman, Dan
1989 "Are Markets Morally Free Zones?" *Philosophy and Public Affairs* 18 (4): 317–333.
- Kreps, David
1990 *A Course in Microeconomic Theory*. Princeton, NJ: Princeton University Press.
- Luce, R. Duncan, and Howard Raiffa
1957 *Games and Decisions: Introduction and Critical Survey*. New York: Dover Publications, Inc.
- Roemer, John
2010 "Kantian Equilibrium." *The Scandinavian Journal of Economics* 112 (1): 1–24.

Roemer, John

2015 "Kantian optimization: A microfoundation for cooperation." *Journal of Public Economics* 127: 45–57.

Voice, Paul

2002 *Morality and Agreement: A Defense of Moral Contractarianism*.
New York: Peter Lang Publishing, Inc.