On David Gauthier's Theories of Coordination and Cooperation

ROBERT SUGDEN *University of East Anglia*

ABSTRACT: In 1975, Gauthier discussed Schelling's pure coordination games *and Hodgson's* Hi-Lo game*. While developing an original analysis of how rational players coordinate on 'focal points,' Gauthier argued, contrary to Schelling and Hodgson, that successful coordination in these games does not depend on deviations from conventional principles of individually rational choice. I argue that Gauthier's analysis of constrained maximization in* Morals by Agreement*, which famously deviates from conventional game theory, has significant similarities with Schelling's and Hodgson's analyses of coordination. Constrained maximization can be thought of as a pragmatic and contractarian variant of the team-reasoning approach pioneered by Hodgson.*

RÉSUMÉ : En 1975, David Gauthier a discuté la question des jeux de coordination pure *de Schelling et des* jeux Hi-Lo *de Hodgson. Tout en proposant une analyse originale de la façon dont les joueurs rationnels se coordonnent sur des «points focaux», Gauthier a soutenu contre Schelling et Hodgson que dans ces jeux, une coordination réussie ne dépend pas de déviations par rapport aux principes conventionnels du choix rationnel individuel. J'avance que l'analyse de la maximisation contrainte proposée par Gauthier dans* Morals by Agreement*, qui s'éloigne de façon notoire de la théorie des jeux conventionnelle, présente d'importantes similarités avec les analyses de la coordination de Schelling et Hodgson. La maximisation contrainte peut être envisagée comme une variante pragmatique et contractualiste de l'approche du raisonnement par équipe introduite par Hodgson.*

Keywords: Gauthier, coordination, cooperation, salience, focal points

***Dialogue* 55 (2016), 713–737.**

doi:10.1017/S0012217316000494

My topic is a paper of David Gauthier's that is not as well known as it should be—a paper with the plain title "Coordination."[1] In this paper, Gauthier discusses two classes of very simple games—the *pure coordination games* first described by Thomas Schelling, and the *Hi-Lo game*, whose paradoxical features were first noticed by David Hodgson.[2] In these games, the players' problem is to coordinate their choices. Most ordinary people, if confronted with one of these games, find the problem very easy to solve. However, this ability to coordinate is puzzling when viewed in the perspective of classical game theory. Gauthier offers a solution to this puzzle.

I have two reasons for saying that *Coordination* deserves to be better known. The first is that, as far as I know, it is the earliest recognition that there is a link between pure coordination games and Hi-Lo, and that this link might be used to explain how players identify the 'focal points' that allow them to solve coordination problems. This insight is the basis for an explanation of focal points subsequently developed in more detail by Michael Bacharach, André Casajus, Maarten Janssen, and me.[3]

The second reason is that *Coordination* is complementary with Gauthier's greatest work, *Morals by Agreement*.[4] In both the paper and the book, Gauthier discusses the foundations of game theory in relation to games that pose challenges for the conception of practical rationality embedded in the classical version of that theory. In *Coordination*, Gauthier defends the conventional conception of rationality as applied to pure coordination games and Hi-Lo, implicitly rejecting the alternative approach offered by Schelling and explicitly rejecting that offered by Hodgson. In *Morals by Agreement*, however, he rejects the conventional conception of rationality as applied to the Prisoner's Dilemma, proposing the radically different approach of 'constrained maximization.' Each of *Coordination* and *Morals by Agreement* refers to the argument of the other, with the implication that the paper and the book are to be understood as providing different components of a single coherent conception of practical rationality. Thus, *Coordination* can throw light on the logic and scope of Gauthier's argument for constrained maximization.

I will argue that there are tensions between Gauthier's rejection of Schelling's and Hodgson's analyses of coordination and his defence of constrained maximization. Schelling's analysis of focal points is based on a pragmatic understanding of rationality that is quite similar to that used by Gauthier when justifying constrained maximization. Hodgson's key insight is that rational decision-making in games can be construed in terms of the players

---

1 Gauthier, 1975.

2 Schelling, 1960; Hodgson, 1967. Hodgson's game was named 'Hi-Lo' by Michael Bacharach.

3 Bacharach, 1993, 2006; Casajus, 2001; Janssen, 2001; Sugden, 1993, 1995.

4 Gauthier, 1986.

jointly choosing the combination of strategies that is best for them collectively, rather than (as in conventional theories of rational choice) each choosing the strategy that he individually judges to be best, given his expectations about what the others will do. There is a significant parallel between this idea and the logic of constrained maximization. Gauthier's overall position is consistent only because he makes a sharp distinction between the domains of coordination (exemplified by Schelling's games and Hi-Lo) and cooperation (exemplified by the Prisoner's Dilemma), restricting the scope of constrained maximization to the latter. I will argue that Gauthier's fundamental insights would be better represented by erasing this distinction and instead applying the principles of constrained maximization to both domains. Following this approach allows one to see Gauthier's conception of rationality as a distinctive and attractive form of 'team reasoning'—a broad theoretical strategy of which Hodgson is a founding father.

1. Preliminaries

Gauthier's analysis of the Prisoner's Dilemma has sometimes been criticized as inconsistent with the fundamental logic of game theory.[5] I think this criticism is mistaken, but, to ensure that it is disarmed, some care must be taken in defining and interpreting game-theoretic concepts.

In both *Coordination* and *Morals by Agreement*, Gauthier accepts the received theory of rational choice, as applied to *parametric* decision problems—that is, problems that involve just one rational agent, as contrasted with *strategic* problems in which two or more rational agents interact with one another. In *Morals by Agreement*, Gauthier supplements his own arguments in favour of the standard conception of parametric rationality with the claim that, since this conception is "almost universally accepted and employed in the social sciences," the onus of proof is on those who reject it.[6] As a social scientist who has reservations about this conception (and who was expressing those reservations at the time *Morals by Agreement* was written), I do not want to endorse this feature of Gauthier's analysis. For the purposes of this paper, however, I will put these reservations aside and take the standard conception of parametric rationality as a fixed point. That is, I will take it as given that, in parametric decisions, a rational individual acts on preferences that satisfy the axioms of expected utility theory. More specifically, for each rational individual there is a *utility function* that assigns a numerical index to everything that could conceivably be the outcome of a parametric decision for that individual. I shall call the set of such outcomes the individual's 'parametric outcome space.' For each individual considered separately, these utility indices are cardinal (more formally: they are unique up

[5] See, for example, Binmore, 1993.

[6] Gauthier, 1986, p. 8.

https://doi.org/10.1017/S0012217316000494 Published online by Cambridge University Press

to positive linear transformations). Each individual's parametric decisions maximize the mathematical expectation of utility, so defined.

In game theory, it is conventional to define a game 'in normal form' by (i) a set of *players*, (ii) for each player, a set of alternative *strategies*, and (iii) for each profile of strategies (one strategy for each player), a corresponding profile of numerical *payoffs*. The task of game theory is to develop *solution concepts* that specify, for given games, which strategies rational players may or may not choose, and/or what beliefs rational players may or may not hold about one another's choices.

Game theorists differ on how payoffs should be interpreted. For some theorists, payoffs are primitives; for others, they contain information about the choices that the players would make in certain environments, perhaps including the game itself. I believe that conceptual clarity is best maintained by not requiring, *as a matter of definition*, that a player's decisions within a game are constrained by her payoffs. If one takes the position I favour, a solution concept can in principle deem *any* strategy to be 'rational' or 'irrational' for the relevant player, irrespective of payoffs. Of course, someone who proposes a particular solution concept, endorsing its prescriptions about what is rational or irrational, needs to be able to justify those prescriptions, given the payoffs of the game. But normative questions about justification should not be closed off by appeals to definitions.

In this paper, I will interpret 'payoffs' in the following way. For any given game, I will assume that, for each player, each profile of strategies leads to an outcome that is an element in that player's parametric outcome space. So, for each player, each outcome of the game has a utility index that represents that player's preferences *with respect to parametric decisions*. I will interpret the payoffs used in the specification of games as utility indices in this parametric sense. Thus, although payoffs represent players' subjective preferences, as revealed in parametric decisions, they impose no formal constraints on behaviour in a game. There is, therefore, no logical incoherence in claiming (as Gauthier does) that rational players can cooperate in a Prisoner's Dilemma. Whether that claim should be accepted or not, it has a legitimate place on the agenda of game theory.

With this interpretative issue out of the way, I can describe the Hi-Lo and coordination games and explain Hodgson's and Schelling's analyses. This will set the scene for *Coordination*.

2. The Footballers' Problem

In a previous paper, I have presented the following problem about rational play in football (for North American readers, soccer):[7]

[7] The argument presented in this section follows Sugden, 1991, 1993. The quotation is from Sugden, 2003, p. 166.

https://doi.org/10.1017/S0012217316000494 Published online by Cambridge University Press

> A and B are players in the same football team. A has the ball, but an opposing player is converging on him. He can pass the ball to B, who has a chance to shoot. There are two directions in which A can move the ball, *left* and *right*, and correspondingly, two directions in which B can run to intercept the pass. If both choose *left*, there is a 10 per cent chance that a goal will be scored. If both choose *right*, the chance is 11 per cent. Otherwise, the chance is zero. There is no time for communication; the two players must act simultaneously. What should they do?

Given that both players want a goal to be scored, we can define utilities for each player on a scale on which not scoring has a value of zero and scoring has a value of 100. If the players are risk-neutral, we have the game shown in Table 1. This is an example of the Hi-Lo game.

Notice that, in order to arrive at this representation of the game, we do not need to make any particular assumptions about *why* each player wants a goal to be scored. Perhaps both players are self-interested, and know that, if the team wins the match, every member of the team will be paid a bonus. But the payoffs would be exactly the same if each player were interested only in the collective achievements of the team. So, if the players have any difficulty in determining what it is rational for them to do, this cannot be attributed to their being motivated by self-interest rather than by the interests of the team.

Apart from the story (which is my invention), the Footballers' Problem is the game discussed by Hodgson. Hodgson uses it to argue that, contrary to a view that was widely held at the time he was writing, rule-utilitarianism is not a special case of act-utilitarianism. In Hodgson's version of the game, the payoffs are measures of the overall goodness of the outcomes of the game, all things considered, and the players are good utilitarians, motivated only by overall goodness.

To see why this game creates problems for the received theory of rational choice (and thereby for act-utilitarianism), imagine a team meeting the day before a big game. The coach is addressing all the players who will be in the team. He explains that the situation described by the Footballers' Problem might arise during the game and says that, if it does, A and B should both play *right*. This leads to the following exchange:

> A: But why should I play *right*?
> Coach: Because if you and B both play *right*, that maximizes the probability that a goal will be scored. Don't you want that?
> A: Of course I do, but how do I know that my playing *right* will maximize the probability of scoring? *Right* is a good move for me only if B plays *right* too.
> Coach: But B is here too. I'm speaking to both of you.
> A: I know that, but you haven't given him a reason to play *right*, just as you haven't given me one.

https://doi.org/10.1017/S0012217316000494 Published online by Cambridge University Press

Table 1 The Footballers' Problem

		B's strategy	
A's strategy		*left*	*right*
	left	10, 10	0, 0
	right	0, 0	11, 11

The point of the story is that the coach is addressing A and B together, telling them what *they should do jointly* to achieve their common objective of maximizing the probability of scoring. He is assuming that A and B together want an answer to the question 'What should we do?' But A wants an answer to the question 'Given my expectations about what B will do, what should I do?'

Hodgson's diagnosis is that, if A insists on an answer to the latter question, a coach who uses only the argumentative resources provided by the received theory of rational choice cannot give him a decisive reason to play *right*. Correspondingly, act-utilitarianism cannot give individuals decisive reasons to coordinate their actions in ways that maximize the overall good, even if it is common knowledge that maximizing the overall good is the objective of every individual. Rule-utilitarianism is not vulnerable to a similar problem. Consider the Footballers' Problem under the assumption that the overall good is measured by the probability that a goal is scored. Leaving aside problems of imperfect rationality, rule utilitarianism tells each individual to follow the rule that, if followed by all, would maximize overall utility. In the Footballers' Problem this rule prescribes *right* to both players.

3. The Two Trains Problem

In *Coordination*, Gauthier describes the following problem. It is the year 1910. A is travelling from Leicester to London to meet B. A has told B that her train arrives at 12.5,[8] and B has agreed to meet her at the station. When it is too late to communicate, they both realize that there are two trains from Leicester that arrive in London at 12.5—a Midland Railway train arriving at St Pancras station and a Great Central Railway train arriving at Marylebone station. A has to decide which train to take from Leicester; B has to decide which train to meet. It is common knowledge between them that A is indifferent between travelling by Midland and travelling by Great Central, and that both are indifferent between the two stations as meeting places. The game they are playing is a pure coordination game. It is represented in Table 2.

[8] In this example, Gauthier displays his encyclopaedic knowledge of railways. Up to the 1960s, what would now be expressed as 12:05 was written in British railway timetables as '12.5.'

https://doi.org/10.1017/S0012217316000494 Published online by Cambridge University Press

Table 2 The Two Trains Problem

		B's strategy	
A's strategy		*Midland*	*Great Central*
	Midland	5, 5	0, 0
	Great Central	0, 0	5, 5

According to Gauthier, travellers on the Leicester-London route at this time would be likely to know that the Midland service was more frequent than the Great Central service, and so carried more passengers. Thus: "Knowing that this information was available to each of us [i.e., A and B], each of us might have considered the Midland train the salient choice for coordination, although our utilities would have been in no way affected."[9] The point of the story is that, when facing coordination games of this kind, real decision-makers can often coordinate by using features of a game that are excluded from conventional game-theoretic analysis. In Schelling's terminology, A and B recognize that, for them, *Midland* is the *focal point*—the point on which their expectations can converge.

For a conventional analysis of the game shown in Table 2, the 'labels' given to the players and the strategies are completely irrelevant, as are any connotations of these labels; all that matters are the entries in the payoff matrix. Since the salient features of the Midland Railway do not affect the payoff matrix, they cannot affect rational play, as defined by conventional theory. From the perspective of that theory, the ability of real people to use 'irrelevant' properties of labelling to achieve results that are individually and collectively beneficial is puzzling.

Schelling was the first theorist to recognize the existence of this ability and the challenge it poses for game theory. In his book *The Strategy of Conflict*, he describes many coordination games and invites his readers to recognize what he claims are the focal points of these games. These claims are supported by evidence from what he calls an 'unscientific sample' of respondents who have been asked to say what they would choose in these games. (Schelling's 'unscientific' findings have since been confirmed in controlled experiments with financial incentives.[10]) From this evidence, Schelling concludes:

> Most situations—perhaps every situation for people who are practiced at this kind of game—provides some clue for coordinating behavior, some focal point for each person's expectation of what the other expects him to expect to be expected to do.[11]

[9] Gauthier, 1975, p. 208.

[10] See Mehta, Starmer and Sugden, 1994.

[11] Schelling, 1960, p. 57.

https://doi.org/10.1017/S0012217316000494 Published online by Cambridge University Press

If this is right, it presents game theory with three problems. The first is to discover what salience is, and how players recognize it. The second is to understand the mode of reasoning by which real players, having recognized that a particular combination of strategies is salient, conclude that they should play those strategies. The third is to decide whether that mode of reasoning should be incorporated into a theory of *rational* decision-making, or whether it should be treated as a form of imperfect reasoning that happens to lead to desired consequences.

Schelling addresses all three of these problems.[12] His way of dealing with the first problem is to discuss a very wide range of coordination problems, to identify their focal points, and to look for features that they have in common. He presents the discovery of focal points as more of an art than a science, and tells us not to expect to find a general, context-independent theory of focality. Likening discovering a focal point to finding the key to a puzzle, and referring to the coordination problem faced by a husband and wife who have lost one another in a department store, he says:

> Finding the key, or rather finding *a* key—any key that is mutually recognized as the key becomes *the* key—may depend on imagination more than on logic; it may depend on analogy, precedent, aesthetic or geometric configuration, casuistic reasoning, and who the parties are and what they know about one another. Whimsy may send the man and his wife to the 'lost and found'; or logic may lead each to reflect on where they would have agreed to meet if they had a prior agreement to cover the contingency.[13]

In explaining the mode of reasoning that players use when they choose strategies that lead to a focal point, Schelling repeatedly refers to this reasoning as a 'meeting of minds.' He gives various, often metaphorical, accounts of what this involves, but an attentive reader will pick out the core elements. Each of the two players in a coordination game reasons in something like the following way: we (i.e., the other player and I) need to coordinate our expectations of one another. For that, we need a common clue that points to one equilibrium. Almost all coordination games have such a clue, if one looks hard enough. If either of us finds a clue, and finds nothing else that is as obviously clue-like, he or she should use it—and not ask whether the clue itself is 'rational' or not.

As an illustration of this kind of reasoning, consider a variant of the Two Trains Problem in which the payoffs for successful coordination are as in the original version, but each of A and B has a slight preference for St Pancras over Marylebone as the station to be at *in the event of a failure to meet*. (Suppose that the only change from the game in Table 2 is that each player gets a payoff of 1 if she plays *Midland* and the other plays *Great Central*.) Discussing a game with exactly this structure, Schelling attributes to each player the thought:

---

12 The following reading of Schelling is defended by Sugden and Zamarrón, 2006.

13 Schelling, 1960, p. 57.

"Comparing just [the equilibria], my partner and I have no way of concerting our choices. There must be some way, however, so let's look for it." Each player recognizes that the asymmetry in the off-diagonal payoffs seems to "point toward" *Midland* as the more obvious choice. Thus, each player reasons:

> since we need an excuse, if not a reason, for pretending, if not believing, that one of the equilibrium pairs is better, or more distinguished, or more prominent, or more eligible, than the other, and since I find no competing rule or instruction to follow or clue to pursue, we may as well agree to use this rule to reach a meeting of minds.[14]

Clearly, the idea of 'agreement' here is not to be taken literally: the players are *imagining* agreeing to treat the payoff asymmetry as the clue. (Similarly, the 'logical' husband and wife in the department store imagine having agreed on a rule about where to meet in the event of losing one another.) Notice how the clue provides a reason for action, not by virtue of its content, but by virtue of its obviousness *as a clue*. When one is thinking about clues, excuses can be as good as reasons, pretences as good as beliefs.

So (moving on the third problem for game theory), is this mode of reasoning rational? Referring to another coordination game, Schelling insists that it is:

> The basic intellectual premise, or working hypothesis, for rational players in this game seems to be the premise that some rule must be used if success is to exceed coincidence, and that the best rule to be found, whatever its rationalization, is consequently a rational rule.[15]

He rejects the suggestion that the use of apparently irrelevant clues is an imperfect form of rationality:

> The assertion here is *not* that people simply *are* affected by symbolic details but that they *should* be for the purposes of correct play. A normative theory must produce strategies that are at least as good as what people can do without them.[16]

To see the force of this argument, consider the Two Trains Problem. A theory of rational play that requires players to use only the information contained in the payoff matrix must treat *Midland* and *Great Central* symmetrically. So, if it yields a unique recommendation, that recommendation must be that A and B play each strategy with probability 0.5, with the result that the probability of coordination is only 0.5. If, as Gauthier claims is the case in the Two Trains Problem, *Midland* is more 'prominent' than *Great Central* in Schelling's sense

14 Schelling, 1960, p. 298.
15 Schelling, 1960, p. 283.
16 Schelling, 1960, p. 98.

https://doi.org/10.1017/S0012217316000494 Published online by Cambridge University Press

of the term, the supposedly rational theory has produced strategies that are *worse* than what people could do by following their instincts.

It seems that Schelling's conception of rationality is *pragmatic*. On behalf of the mode of reasoning he is recommending, he is making three claims. First, this mode of reasoning works well for each individual, given what other individuals can *in fact* be expected to do: even if you were the only person to read *The Strategy of Conflict*, its recommendations would be good advice to you. (Schelling's experimental evidence is intended to show that people already have a tendency to choose salient strategies in coordination games. Given that others have this tendency, it is in your interest to choose salient strategies.) Second, this mode of reasoning is more successful for each individual, the stronger the tendency for other individuals to use it. Thus, Schelling's advice does not undercut itself when it is made public. Third, all individuals do better if all of them act on the advice than if none of them do. I take it that, for Schelling, that is all that needs to be said.

4. Gauthier's Analysis of Natural Interaction

Hodgson's and Schelling's analyses provide the point of departure for *Coordination*. Setting out the objectives of the paper, Gauthier says: "My main purpose is to show how problems of coordination are to be resolved within the framework of an account of rational action, as rational action is commonly understood." This common understanding is "the conception of rationality commonly accepted in economic and social scientific enquiries, [according to which] a rational person is a utility-maximizer."[17] Gauthier is hinting that neither Hodgson nor Schelling has used the standard account of rational action (as indeed they have not); his contribution will be to show how the problems they have identified can be solved without deviating from that account.

In a footnote to his description of the commonly accepted definition of rationality, Gauthier says that he does not regard the utility-maximizing conception of rationality as "fully satisfactory," but "its inadequacies do not manifest themselves at the fairly simple level of a theory of rational coordination." Gauthier's reasons for thinking this conception inadequate are made clear in *Morals by Agreement*, which distinguishes between two "modes of interaction" for rational individuals—*natural* and *cooperative*.[18] In natural interaction (natural in the sense of 'state of nature'), the relevant conception of rationality is that of standard game theory, in which each individual is a utility-maximizer (or, as Gauthier sometimes says, a 'straightforward maximizer'). The crucial theoretical innovation of *Morals by Agreement* is the idea of constrained maximization, which (according to Gauthier) is the appropriate form of rationality for cooperative interaction. It is central to Gauthier's argumentative strategy that these two ways of conceiving of interaction exist side-by-side, and that fully

[17] Gauthier, 1975, pp. 196-197.

[18] Gauthier, 1986, pp. 113-118.

rational individuals can choose which of them to use in particular decision environments. Roughly speaking, cooperative reasoning is called for when natural reasoning leads to outcomes that are not Pareto-optimal. (To say that an outcome is Pareto-optimal is to say that there is no feasible alternative such that a move to that alternative would make some individual better off, and make no individual worse off, assessed in terms of their respective preferences.)

For Gauthier, natural reasoning serves as a baseline for the analysis of rationality. This default form of rationality is the subject of Chapter III of *Morals by Agreement*. In this chapter, Gauthier briefly restates the analysis of Hi-Lo that is developed in more detail in *Coordination*. The implication is that coordination problems belong to the domain of natural interaction; the mode of rationality that is appropriate to them is, therefore, that of standard game theory.

In *Morals by Agreement*, Gauthier presents the following three necessary but not sufficient "conditions on strategically rational choice":[19]

> A: Each person's choice must be a rational response to the choices she expects the others to make.
>
> B: Each person must expect every other person's choice to satisfy condition A.
>
> C: Each person must believe her choice and expectations to be reflected in the expectations of every other person.

Since Gauthier wants these conditions to apply to both natural and cooperative interaction, he allows some ambiguity in the concept of a 'rational response.' But, for the case of natural interaction, 'rational' is to be read as 'individually utility-maximizing.' This makes a 'rational response' equivalent to the standard game-theoretic concept of a 'best response.'

Given that Gauthier is representing an ideal case of individual rationality, recognizing that the 'full-blown ideal' he is describing is unlikely to occur in reality, Conditions A and B are unexceptionable. But condition C warrants a closer look. Clarifying the concept of 'reflection,' Gauthier restates condition C as "each person views the situation as if her knowledge of the grounds for choice were complete, shared by all, and known by all to be so shared." Explaining why this condition is necessary, he says that, if the specification of a game is common knowledge among the players, "then each person's reasoning from these data to his own expectations and choices must be accessible to every other person."[20]

This argument seems to depend on the hidden premise that there exists some mode of reasoning that constitutes the rationality that is attributed to 'rational' players, and that this mode of reasoning allows each player to deduce a unique choice for himself and a unique set of expectations about the other players' choices. *Given this premise*, we can conclude that each rational player can use

---

19 Gauthier, 1986, p. 61.

20 Gauthier, 1986, pp. 60-61.

https://doi.org/10.1017/S0012217316000494 Published online by Cambridge University Press

the postulated mode of 'rational' reasoning, not only to determine his own choice and expectations, but also to discover the choices and expectations of other rational players. But Gauthier does not try to specify what this mode of reasoning is. In taking this approach, he is following a well-established tradition in game theory, sometimes described in terms of 'the authoritative book.' The idea is to imagine that there is a book which, for every possible game, gives a unique recommendation about what each player should choose (or, in some variants, what each player should believe about other players' choices). The book certifies that all of its recommendations are uniquely required by principles of rationality, but without stating what those principles are or how its recommendations are derived from them. All players have access to the book and accept its authority. Game theorists can then ask: *If* there were such a book, and *if* its claim to authority really were justified, what would it contain?

There is a good argument that, *were the book to exist*, its recommendations would satisfy Gauthier's three conditions. If the book tells some player that a particular choice is rationally required, that choice should be a best response to what the other players, acting on the book's recommendations, will do: that is condition A. Conditions B and C are implications of the idea that all players have access to the book and accept its authority (and that this is common knowledge). One might go further and say that, were the book to exist, truly rational players would not need to read it, because they would already know what recommendations could be derived from principles of rationality.

Gauthier explains that, as is well-known to students of game theory, the three conditions are satisfied if and only if players' choices are in Nash equilibrium (that is, each player's chosen strategy is a best response to the others' chosen strategies), and that (leaving aside some technical issues about finiteness) every game has at least one Nash equilibrium. Thus, *a* book can be written with recommendations that satisfy conditions A, B, and C. Indeed (since many games have more than one equilibrium), many such books can be written, each making a different set of recommendations. But this is not to say that it is possible to write an *authoritative* book—that is, a book whose claims to authority can be justified. Unless we can show that such a book can be written, the idea that a fully rational being would know what it contained is a matter of doctrine and faith, not of science.[21] Gauthier is one of the faithful.

[21] Although a good deal of rationality-based game theory follows the 'authoritative book' approach, there are other traditions of analysis that represent 'rational' reasoning more explicitly and investigate the conclusions that players can reach by using such reasoning. The conclusions of this work suggest that faith in the possibility of an authoritative book is misplaced. See, for example, Bernheim, 1982; Pearce, 1982; Bacharach, 1987; and Cubitt and Sugden, 2014. A case can be made for Lewis as one of the first game theorists to analyze rationality in terms of explicit reasoning: see Lewis, 1969 and Cubitt and Sugden, 2003.

5. Gauthier on Hi-Lo

Recall Hodgson's argument that the principles of act-utilitarianism cannot show the players of a Hi-Lo game that it is uniquely rational for each of them to choose the strategies that lead to the outcome that maximizes the overall good. In *Coordination*, Gauthier reviews this argument. He acknowledges that, if each individual uses act-utilitarianism as his *only* principle of practical reason, Hodgson's conclusion follows. But:

> [Hodgson's] position does not rule out act-utilitarianism, or more generally act-consequentialism, in any genuinely important sense. For he does not show that the act-utilitarian principle is inconsistent, that it requires incompatible actions or rules out all actions in certain circumstances. Nor does he show that the principle can not be supplemented by a further principle or principles, entirely compatible with it, and serving only to determine a particular action in those situations left indeterminate by the act-utilitarian principle itself.[22]

Gauthier claims to have found an additional principle that will allow act-utilitarianism to avoid the problem identified by Hodgson.

Consider any game. If the game has two Nash equilibria, E_1 and E_2, E_1 *payoff-dominates* E_2 if every player's payoff is at least as great in E_1 as in E_2, and if some player's payoff is strictly greater in E_1. The 'Principle of Coordination' stipulates that, if the game has a Nash equilibrium E which payoff-dominates every other Nash equilibrium, then each player should choose the strategy that is consistent with E. Clearly, if all players act on this principle, the Hi-Lo problem is solved. In the Footballers' Problem, for example, the (*right*, *right*) equilibrium payoff-dominates all other equilibria, and so the principle requires each player to choose *right*. In terms of the analysis in *Morals by Agreement*, the Principle of Coordination can be treated as a necessary condition on strategically rational choice, and added to conditions A, B, and C. There is no inconsistency in doing this: every (finite) game has a solution that satisfies the expanded set of conditions. (Conditions A, B, and C narrow down the set of admissible solutions to the set of Nash equilibria, which we know to be non-empty. The Principle of Coordination can come into play only if there are two or more equilibria; it may eliminate some of these but it cannot eliminate them all.)

In this formal sense, the Principle of Coordination is compatible with Gauthier's account of rationality in natural interaction. But is it compatible with the mode of reasoning that defines act-utilitarianism? More generally, is it consistent with the mode of reasoning whose conclusions standard game theory purports to represent? Gauthier argues that it is: it "completes the act-utilitarian principle in a manner consonant with the spirit of act-utilitarianism." For Gauthier,

---

22 Gauthier, 1975, pp. 202-206.

https://doi.org/10.1017/S0012217316000494 Published online by Cambridge University Press

the spirit of act-utilitarianism is contained in two ideas—that each individual "attends to, and only to, the consequences of particular actions" and that each individual "is concerned with the maximization of utility." The Principle of Coordination, he claims, is compatible with both ideas.[23] But this is an inadequate response to Hodgson. It would be reasonable enough to claim that these two ideas express the spirit of *utilitarianism*; but what is at issue is the distinction between *act*- and *rule*-utilitarianism. An act-utilitarian individual attends to, and only to, the consequences of *his* actions, and is concerned that *his* actions should maximize overall utility. In contrast, the Principle of Coordination tells the players (in the plural) how *they* should act if *they* are to maximize utility.

Gauthier might reply that the Principle of Coordination tells each player how he can maximize overall utility through his own actions*, given the expectation that other players will act on the principle too*. But how is that expectation grounded? We are back with the authoritative book. Consider a Hi-Lo game played by two act-utilitarians: each player's payoffs are measures of the overall good. A book of game-theoretic recommendations can tell each player to choose the strategy that leads to the best equilibrium. If that recommendation were genuinely authoritative, each player could expect the other to act on it; thus, each could expect that by acting on the recommendation he would maximize overall utility. But *is* the recommendation authoritative? That is, can it be reached by a mode of act-utilitarian reasoning that can be defended as rational? Gauthier's argument does not answer that question.

I submit that Hodgson's conclusion is correct: the Principle of Coordination cannot be defended as rational by appeal to the principles (or even to the spirit) of act-utilitarianism. Similarly, it cannot be defended by appeal to the standard game-theoretic assumption that players are individually rational and that this is common knowledge. If the Principle of Coordination is a principle of rationality, the rationality must be of a non-standard kind. Specifically, this must be a form of rationality that can respond to the question 'What should *we* do?' One of the distinctive features of rule-utilitarianism is its recognition of the meaningfulness of this question.

6. Gauthier on Pure Coordination Games

In a sentence that I have already quoted from *Coordination*, Gauthier says that his main purpose is to show how problems of coordination are to be resolved within the framework of a standard account of rational action. This sentence is immediately preceded by the remark that Schelling's purpose in discussing the problem of coordination is "somewhat different from mine."[24] The implication is that Schelling's analysis does not use the standard theory of rational choice. Gauthier does not explain what feature of Schelling's analysis is non-standard, but what he seems to have in mind is that Schelling allows his players to use

[23] Gauthier, 1975, p. 206.

[24] Gauthier, 1975, p. 196.

kinds of information that (according to the received theory) rational individuals would ignore.

Recall that, in the Two Trains Problem, *Midland* is salient because the Midland service carries more passengers. However, in the game that is actually being played, it is common knowledge that both players are indifferent between *Midland* and *Great Central*. According to Gauthier:

> Rational maximizers of utility can not use additional information, not incorporated into the utilities of the situation, to generate expectations which converge on one of several best equilibria. So the salience of the Midland, based on mutual knowledge of the greater volume of traffic which it carries, does not in itself give either [player] a reason for selecting one course of action rather than another.[25]

In saying this, Gauthier is taking a position that is shared by many game theorists; but he does not offer any justification for it. It is an *ex cathedra* pronouncement, a statement of doctrine. Schelling's analysis clearly contravenes this doctrine. (Indeed, one of Schelling's purposes is to show why that doctrine should be abandoned.)

Gauthier's objective is to show that rational players can choose a salient equilibrium while using only the information "incorporated into the utilities of the situation." He suggests that each player might reason in the following way:

> Our problem is that we have two equally good meeting places. What we need is a way to restructure our conception of the situation so that we are left with but one. We must restrict the possible actions which we consider, in such a way that we convert our representation of the situation into one with but one best [i.e., payoff-dominant] equilibrium.[26]

Notice that, in Gauthier's account of pure coordination games, just as in Schelling's, the players are recognizing a problem that *they need to solve together* ("our problem," "what we need"): they are trying to find a meeting of minds. By assumption, the players have common knowledge that *Midland* is salient, but their rationality prevents them from treating this as a reason for choosing that strategy. However, they can achieve the same result by finding a new "conception of the situation" in which there is a payoff-based reason for choosing one equilibrium rather than the other. They can do this by "singling out some characteristic of some one equilibrium outcome" and then conceptualizing each player's problem so that there are only two options—to seek an outcome with that characteristic, or to ignore that characteristic.[27]

---

[25] Gauthier, 1975, p. 210.

[26] Gauthier, 1975, p. 210.

[27] Gauthier, 1975, p. 211.

But how do the players, reasoning independently, single out the same characteristic? Gauthier's answer is:

> Each person must be able to single out the restricting characteristic independently of the others, and yet each must expect that all will single out the same characteristic. Coordination on the characteristic is required for coordination in action. Necessarily, therefore, the restricting characteristic is salience. For the salient outcome is, by definition, that which is apprehended as standing out from the others.[28]

Again, Gauthier is imagining the players reasoning about what *they* need to do in order to solve *their* problem. They need to find a characteristic that each can recognize as the 'obvious' one to use. In this respect, Gauthier's argument is essentially the same as Schelling's.

In the Two Trains Problem, the salient characteristic is 'carrying more passengers.' Thus, 'seeking the salient characteristic' is extensionally equivalent to choosing *Midland*, while 'ignoring the salient characteristic' is extensionally equivalent to picking one of *Midland* and *Great Central* at random. If the problem is reconceptualized in terms of seeking or ignoring salience, we arrive at the game shown in Table 3, in which payoffs are expected utilities. In *this* game, one equilibrium—that in which each player chooses *seek salience*—payoff-dominates the others. So the Principle of Coordination tells each player to choose *seek salience*—that is, to head for St Pancras station.

Gauthier's analysis has reached the same conclusion as Schelling's, but by a more circuitous route. In both analyses, the players achieve a meeting of minds, recognizing the salience of the characteristic 'carrying more passengers' as a means of coordinating on *Midland*. Schelling's players then choose *Midland* because it has this salient characteristic. Gauthier's players realize that, if they reconceptualize the game in terms of this characteristic, they will be able to invoke the Principle of Coordination, which will tell them that it is rational to seek salience and hence to coordinate on *Midland*; since they want to coordinate, they do this. Gauthier can claim to have remained faithful to the doctrine that rational players use only the information that is contained in the utilities of a situation, but he has not offered any counter to Schelling's argument that that doctrine should be rejected.

Considering only the text of *Coordination*, one might reasonably conclude that Gauthier's analysis of pure coordination games adds very little to Schelling's. However, Gauthier's analysis introduces two ideas that have proved to be valuable in explaining focal points.

The first is that an analysis of rational play in games must use *the players' own descriptions* of possible actions: if an individual is to choose an action, that action must have some description that is accessible to the individual herself. The descriptions that players use need not be the same as those that game theorists find most natural. This is not just a matter of using different words to

[28] Gauthier, 1975, p. 211.

Table 3 The Two Trains Problem Reconceived

		B's strategy	
A's strategy		*seek salience*	*ignore salience*
	seek salience	5, 5	2.5, 2.5
	ignore salience	2.5, 2.5	2.5, 2.5

describe the same object: there need not be a one-to-one mapping from one set of descriptions to the other. Gauthier's analysis of the Two Trains Problem illustrates this point. Thinking about the story behind the Two Trains Problem, most game theorists would immediately conceptualize it as the game in Table 2, in which each player chooses from the set of strategies {*Midland*, *Great Central*}. But the same story can also be represented by the game in Table 3, in which each player chooses from {*seek salience*, *ignore salience*}. These alternative representations have very different formal structures. In the game in Table 2, it is possible for the players to coordinate intentionally on *Great Central*. In the game in Table 3, the extensionally equivalent outcome cannot be reached by deliberate choice (it occurs with probability 0.25 if both players choose *ignore salience*).

The second idea is that an interaction that is a pure coordination game under one set of descriptions may be a Hi-Lo game under another, and that salience in the former might correspond with payoff-dominance in the latter. In other words, the question 'Why is it rational to choose the salient equilibrium?' might be transformed into 'Why is it rational to choose the payoff-dominant equilibrium?' Intuitively, the latter question seems less puzzling, whatever problems it may pose for the received theory of rational individual choice.

For example, consider the following pure coordination game, which belongs to the class of *blockmarking* games analyzed by Bacharach.[29] Each of two players sees the diagram shown in Figure 1; each knows that the other is seeing an identical diagram. Without communicating with the other, each player must draw a cross on one of the objects in her diagram. Each will be rewarded if and only if they both mark the same object. Intuitively, it seems obvious that each should mark the circle. Bacharach offers a rationale for this intuition in terms of the *act-descriptions* under which a player might mark an object. His idea is that players may perceive the individual triangles as *nondescript*—as lacking accessible descriptions that single them out. For a player who sees the problem in this way, the set of possible act-descriptions is *not* the set of 10 objects. It is the set {*pick an object*, *pick a triangle*, *choose the circle*}. Assuming that picking is random, the players' probability of coordination is 1/10 if they both choose the first option in this set, 1/9 if they both choose the second, and 1 if they both choose the third. This is a Hi-Lo game in which the 'choose the circle' equilibrium is payoff-dominant.

[29] Bacharach, 1993.

https://doi.org/10.1017/S0012217316000494 Published online by Cambridge University Press

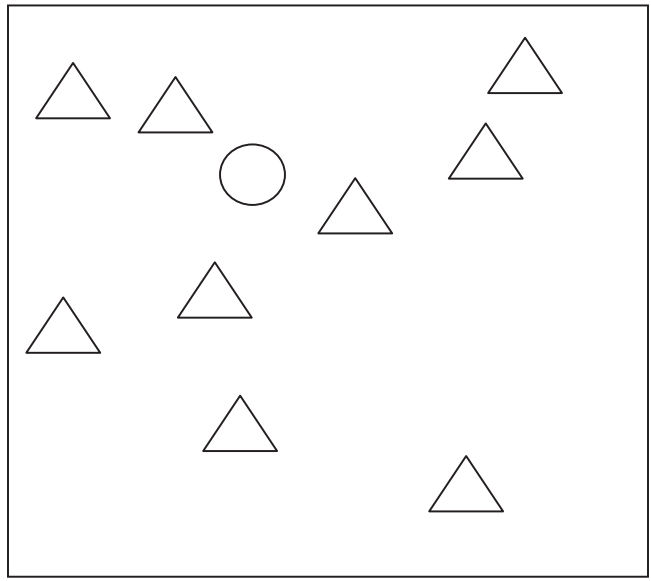

Figure 1 A Blockmarking Game

Bacharach's approach is not quite the same as Gauthier's, in that Bacharach treats the set of available act-descriptions as given to each player, while Gauthier allows players to choose how to conceptualize the game. Bacharach's approach is better suited to an empirical theory of focal points, since it makes a sharper distinction between what is to be explained (players' actual decisions, described in the language of an observing social scientist) and what is to do the explaining (the subjective act-descriptions that are available to each player). Nevertheless, Gauthier should be credited with some of the key insights of what is now a well-developed theory of focal points.

7. Constrained Maximization

Gauthier treats coordination problems, as exemplified by Hi-Lo and Schelling's pure coordination games, as cases of natural interaction. He argues that, by acting on the principles of individual rationality used in standard game theory, the players of these games are able to coordinate successfully. He thereby disagrees with Hodgson and Schelling, who claim that successful coordination depends on non-standard principles of rationality. I have sided with Hodgson and Schelling. I now consider whether Gauthier's own version of non-standard rationality, constrained maximization, might have implications for coordination problems.

In *Morals by Agreement*, Gauthier presents constrained maximization as a form of rationality that is appropriate for cooperative interaction. He introduces

https://doi.org/10.1017/S0012217316000494 Published online by Cambridge University Press

the concept of 'cooperative interaction' in relation to problems of market failure. Under conditions of perfect competition, Pareto-optimal outcomes can be achieved when each individual acts as a parametric utility-maximizer. Gauthier interprets this well-known result as showing that a perfectly competitive market would be "a morally free zone, a zone within which the constraints of morality would have no place."[30] But if there are externalities, this result breaks down. What is then required is a different mode of interaction: cooperative interaction.

Gauthier defines a *joint strategy* as a combination of strategies, one for each player in the relevant game. This concept is fundamental to Gauthier's account of cooperative interaction: "A person co-operates with his fellows only if he bases his actions on a joint strategy; to agree to co-operate is to agree to employ a joint rather than an individual strategy." A joint strategy can be put into practice only by the separate actions of the individual players, each choosing his component of it. But each player, in performing this individual action, construes it as *his part* of the joint strategy:

> An individual is not able to ensure that he acts on a joint strategy, since whether he does depends, not only on what he intends, but on what those with whom he interacts intend. But we may say that an individual bases his action on a joint strategy in so far as he intentionally chooses what the strategy requires of him.[31]

Of course, cooperation requires more than the implementation of a joint strategy: the joint strategy must be one that can be interpreted as *cooperative*. Constrained maximization is based on a contractarian conception of cooperation as the implementation of actual or hypothetical agreements:

> In order to take effective account of externalities, each person must choose her strategy to bring about a particular outcome determined by prior agreement among those interacting. This agreement, if rational, will ensure optimality. It may of course be implicit rather than explicit, an understanding or convention rather than a contract. But it is not a mere fiction, since it gives rise to a new mode of interaction, which we identify as co-operation.[32]

If the players of a game are unable to communicate before choosing their strategies, any 'agreement' must be hypothetical. In this case, Gauthier defines the cooperative joint strategy as the one the players *would have agreed* to implement, had they had the power to negotiate contracts binding them to specific individual choices, and had they negotiated in a fully rational way.

---

30 Gauthier, 1986, p. 84.

31 Gauthier, 1986, p. 166.

32 Gauthier, 1986, p. 117.

To avoid unhelpful digressions, I shall assume that there is a satisfactory theory of rational bargaining that, when applied to the relevant game, identifies a unique joint strategy.[33] Given this assumption, it is legitimate to speak of *the* cooperative joint strategy of a given game. A player who acts according to the principle of constrained maximization chooses his component of the cooperative joint strategy, provided he has assurance that the other players will choose theirs.[34]

Now consider how Gauthier's account of constrained maximization relates to Hodgson's analysis of Hi-Lo. For Gauthier, the implementation of a joint strategy is an essential component of cooperative interaction. And this is precisely what, according to Hodgson, is required in Hi-Lo. In the case of the Footballers' Problem, there is a joint strategy of which each player's component is *right*. In the story of the team meeting, player A fails to understand the coach's recommendation that A and B should each play *right*. The player and the coach are at cross-purposes because A is thinking about 'A plays *right*' as an individual strategy, while the coach is thinking about it as a component of a joint strategy. The coach's recommendation is that A should base his action on a joint strategy for A and B together. In support of this recommendation, he tells each player that, since it and the reasoning that leads to it are common knowledge, each can expect the other to act on it too. The core of Hodgson's argument is the claim that the coach's mode of reasoning is valid, and that rule-utilitarianism is superior to act-utilitarianism in recognizing this validity.

It is puzzling that Gauthier is so reluctant to accept Hodgson's analysis. On Gauthier's behalf, it can be said that his own analysis of Hi-Lo is intended to show that individually rational players will coordinate on the payoff-dominant equilibrium; if this is so, there is a sense in which joint reasoning is redundant. But if, as Hodgson argues and as I think is correct, individual rationality does *not* guarantee that this equilibrium will be reached, the Hi-Lo problem is analogous with the cases of market failure that Gauthier sees as calling for

[33] Gauthier proposes a theory of rational bargaining that applies to cases of actual rather than hypothetical bargaining. For all bargaining games in a very general class, this theory generates a unique solution in terms of players' *utilities*. However, this solution might be compatible with more than one joint strategy, and might be attainable only if the players coordinated on one of those. This coordination problem is trivial if the players can communicate freely, as is typical of actual bargaining; but it is not at all trivial when players are separately thinking about hypothetical agreements.

[34] This is constrained maximization in its purest form. The definition used by Gauthier allows some slippage from the theoretical ideal. Roughly: a constrained maximizer chooses her component of a joint strategy whose outcomes are sufficiently close to those of the rational bargaining solution, provided that the probability that other players will act on that joint strategy is sufficiently high (see Gauthier 1986, p. 167).

cooperative interaction. And, if this problem is treated as belonging to the domain of cooperative interaction, the cooperative joint strategy is clearly the one that leads to the payoff-dominant equilibrium. (This is an implication of Gauthier's preferred theory of rational bargaining, but would be implied by any such theory that was remotely plausible.)

However, there is still a significant difference between Hodgson's analysis and constrained maximization. This concerns the criterion by which the players of a game select which joint strategy to implement. In Hodgson's theory, the criterion is the maximization of a single collective objective—overall utility. Hodgson's insights have since been developed by Bacharach and me in related theories of team reasoning.[35] These theories treat the players of a game as constituting a 'team' and attribute some collective objective to that team. Team reasoning then involves the players implementing the joint strategy (if there is one) that uniquely maximizes that objective. In contrast, Gauthier's approach is contractarian: a joint strategy is selected as the outcome of hypothetical rational bargaining among the players, each of whom is seeking to maximize her own individual objective. In this respect, Gauthier's analysis of constrained maximization can be thought of as a distinctive theory of team reasoning.[36] Interestingly, Schelling's discussions of focal points sometimes hint at a similar kind of contractarianism: recall the husband and wife in the department store who solve their coordination problem by separately imagining a hypothetical agreement about where to meet.

Gauthier presents constrained maximization as a mode of reasoning that exists in parallel with straightforward maximization. Each of these modes of reasoning is based on its own internally consistent conception of rational action. However, Gauthier also works at a higher conceptual level, at which it is possible to ask which mode of reasoning it would be more rational for an individual to be disposed to use in given circumstances. Having restricted his analysis to the case of the Prisoner's Dilemma, Gauthier claims to have "defended the rationality of constrained maximization as a disposition to choose by showing that it would be rationally chosen."[37]

Describing the method of analysis by which this conclusion is reached, he says:

> We consider what a rational individual would choose, given the alternatives of adopting straightforward maximization, and of adopting constrained maximization, as his disposition for strategic behaviour. … Taking others' dispositions as fixed, the individual reasons parametrically to his own best disposition.[38]

---

35 Sugden, 1993; Bacharch, 2006.

36 I find this approach very attractive. I am currently working on formulating a contractarian theory of team reasoning. For a first sketch, see Sugden, 2015.

37 Gauthier, 1986, p. 183.

38 Gauthier, 1986, pp. 170-171.

https://doi.org/10.1017/S0012217316000494 Published online by Cambridge University Press

For this decision problem, the criterion of rationality is the maximization of expected utility. Gauthier sees it as important for his defence of constrained maximization that the argument "does not appeal to any weakness or imperfection in the reasoning of the actor"; his aim is to draw a lesson "about the dispositions and choices of the perfect actor."[39] However, the idea of rational choice among dispositions is not to be taken literally: it is to be understood as "a heuristic device to express the underlying requirement, that a rational disposition to choose be utility-maximizing."[40]

Despite what might be read as a claim to the contrary, Gauthier does not prove that constrained maximization is *unconditionally* better than straightforward maximization for the players of the Prisoners' Dilemma. What he shows is that constrained maximization is the better choice if players' dispositions are sufficiently "translucent"—that is, if players can identify one another's dispositions with sufficient accuracy—and if the proportion of constrained maximizers in the population is sufficiently high. He supplements this theoretical analysis with informal speculations about how translucent "we may reasonably consider ourselves to be" and about the relative frequency of constrained maximizers in "the conditions in which we find ourselves."[41]

The argument shows some signs of stress at this point, because Gauthier is combining a theoretical analysis of perfect actors playing an abstract formal game with empirical propositions about real people involved in real interactions. Gauthier's practical conclusions are best summarized in the following passage:

> In a world of Fooles [i.e., straightforward maximizers] it would not pay to be a constrained maximizer … But if we find ourselves in the company of reasonably just persons, then we too have reason to dispose ourselves to justice. A community in which most individuals are disposed to comply with fair and optimal agreements and practices, and so to base their actions on joint co-operative strategies, will be self-sustaining. And such a world offers benefits to all which the Fooles can never enjoy.[42]

I take Gauthier to be making three claims about constrained maximization, as compared with straightforward maximization. First: in the world as it currently is, it is normally in each person's interest to be disposed to constrained rather than straightforward maximization. Second: holding other features of the world constant, the stronger the tendency for others to be disposed to constrained rather than straightforward maximization, the greater the gain to each individual from being disposed in that way. Third: it is in everyone's interest that everyone is disposed to constrained rather than straightforward maximization. Thus, the

39 Gauthier, 1986, p. 186.

40 Gauthier, 1986, p. 182.

41 Gauthier, 1986, p. 174.

42 Gauthier, 1986, pp. 181-182.

recommendation "Dispose yourself to be a constrained maximizer" is good advice to each individual, given the world as it actually is; it would remain good advice to each individual, were others to act on it too; and it is good advice to people collectively.

But this is exactly the kind of pragmatic recommendation that Schelling gives for choosing strategies that lead to salient equilibria in coordination games. Gauthier can add that the argument for constrained maximization does not appeal to any weakness or imperfection in anyone's reasoning; but Schelling can (and does) say exactly the same about the argument for choosing according to salience. Gauthier might object that an individual who acts on Schelling's pragmatically rational advice violates a doctrinal principle of game theory, namely that rational players cannot use information that is not contained in the payoffs of the game; but the idea of choosing among dispositions rather than strategies is contrary to received doctrine too. Gauthier's argument, like Schelling's, ultimately depends on an appeal to the principle that a normative theory must produce recommendations that are at least as good as what people can do without them. Neither Gauthier nor Schelling has any need to apologize for this.

8. Conclusion

Gauthier's analysis of constrained maximization in the Prisoner's Dilemma has significant similarities with Schelling's analysis of pure coordination games and with Hodgson's analysis of Hi-Lo. These similarities suggest that Gauthier may have been too ready to distance himself from Schelling and Hodgson. One can appreciate the originality and insight of Gauthier's contribution to the theory of focal points without seeing it as being in opposition to Schelling's approach. And the theory of rational action developed in *Morals by Agreement* can be thought of as a pragmatic and contractarian variant of the team-reasoning approach of which Hodgson was a pioneer.

Acknowledgements: An earlier version of this paper was presented at the conference "Contractarian Moral Theory: The 25th Anniversary of *Morals by Agreement*," held at York University, Ontario in May 2011. I thank conference participants for their comments.

References

Bacharach, Michael
1987 "A Theory of Rational Decision in Games," *Erkenntnis* 27, 17–55.

Bacharach, Michael
1993 "Variable Universe Games," in Ken Binmore, Alan Kirman and Piero Tani (eds.), *Frontiers of Game Theory*, Cambridge, MA: MIT Press, 255–276.

Bacharach, Michael
2006 *Beyond Individual Choice: Teams and Frames in Game Theory*, Princeton, NJ: Princeton University Press.

Bernheim, B. Douglas
1984 "Rationalizable Strategic Behavior," *Econometrica* 52 (4), 1007–1028.
Binmore, Ken
1993 "Bargaining and Morality," in David Gauthier and Robert Sugden (eds.), *Rationality, Justice and the Social Contract*, Hemel Hempstead: Harvester Wheatsheaf, 131–156.
Casajus, André
2001 *Focal Points in Framed Games: Breaking the Symmetry*, Berlin: Springer-Verlag.
Cubitt, Robin and Robert Sugden
2003 "Common Knowledge, Salience and Convention: A Reconstruction of David Lewis's Game Theory," *Economics and Philosophy* 19 (2), 175–210.
Cubitt, Robin and Robert Sugden
2014 "Common Reasoning in Games: A Lewisian Analysis of Common Knowledge of Rationality," *Economics and Philosophy* 30 (3), 285–329.
Gauthier, David
1975 "Coordination," *Dialogue* 14 (2), 195–221.
Gauthier, David
1986 *Morals by Agreement*, Oxford: Clarendon Press.
Hodgson, David H.
1967 *Consequences of Utilitarianism*, Oxford: Clarendon Press.
Janssen, Maarten
2001 "Rationalising Focal Points," *Theory and Decision* 50 (2), 119–148.
Lewis, David
1969 *Convention: A Philosophical Study*, Cambridge, MA: Harvard University Press.
Mehta, Judith, Chris Starmer and Robert Sugden
1994 "The Nature of Salience: An Experimental Investigation of Pure Coordination Games," *American Economic Review* 84, 658–673.
Pearce, David
1984 "Rationalizable Strategic Behavior and the Problem of Perfection," *Econometrica* 52 (4), 1029–1050.
Schelling, Thomas
1960 *The Strategy of Conflict*, Cambridge, MA: Harvard University Press.
Sugden, Robert
1991 "Rational Choice: A Survey of Contributions from Economics and Philosophy," *Economic Journal* 101 (407), 751–785.
Sugden, Robert
1993 "Thinking as a Team: Towards an Explanation of Non-selfish Behavior," *Social Philosophy & Policy* 10 (1), 69–89.
Sugden, Robert
1995 "A Theory of Focal Points," *Economic Journal* 105 (430), 533–550.

Sugden, Robert
2003 "The Logic of Team Reasoning," *Philosophical Explorations* 6 (3), 165–181.

Sugden, Robert
2015 "Team Reasoning and Intentional Cooperation for Mutual Benefit," *Journal of Social Ontology* 1 (1), 143–166.

Sugden, Robert and Ignacio Zamarrón
2006 "Finding the Key: The Riddle of Focal Points," *Journal of Economic Psychology* 27 (5), 609–621.