Salience and focusing in pure coordination games — **Source link** ↗

Andrew M. Colman

Related papers:

- Learning how to cooperate : optimal play in repeated coordination games

- The Strategy of Conflict

- Convention: A Philosophical Study

- Agreeing to Disagree

- A Theory of Focal Points

Running head: SALIENCE AND FOCUSING

SALIENCE AND FOCUSING IN PURE COORDINATION GAMES

Andrew M. Colman

University of Leicester

Address for correspondence:

Dr Andrew M. Colman

Department of Psychology

University of Leicester

Leicester LE1 7RH

E-mail: amc@leicester.ac.uk

ABSTRACT

This article is devoted to explaining and justifying the choice of salient equilibria or focal points in pure coordination games -- games in which players have identical preferences over the set of possible outcomes. Focal points, even when they arise as framing effects based on the labelling of options, are intuitively obvious choices, and experimental evidence shows that people often coordinate successfully by choosing them. In response to arguments that focusing is not rationally justified, a qualified justification and psychological explanation is offered in terms of a form of reasoning called the Stackelberg heuristic, which has been used to explain the selection of payoff-dominant (Pareto-optimal) equilibria in common-interest games. Pure coordination games, if appropriately modelled, are shown to be reducible to common-interest games with payoff-dominant equilibria, and it is argued that focusing can therefore be explained by the Stackelberg heuristic.

Keywords: coordination games, focal points, matching games, payoff dominance, salience, Stackelberg heuristic

SALIENCE AND FOCUSING IN PURE COORDINATION GAMES

1. INTRODUCTION

The aim of this article is to suggest a new explanation for the phenomenon of focal point selection, or what I shall call <u>focusing</u>, in pure coordination games. The defining property of a pure coordination game is complete agreement among players' utility functions. In such a game, the players have identical preferences over the set of possible outcomes, so that their interests are not in conflict: they are motivated solely to coordinate their strategies in order to obtain an outcome that is best for both (or all) of them. The simplest (2×2) example, involving the minimum of two players, each with the minimum of two strategies, is the following game, called Heads or Tails. Two people are invited to choose heads or tails simultaneously or (what amounts to the same thing) independently, knowing that if they both choose heads or both tails, then each will receive a payoff of one unit of utility, otherwise their payoffs will be zero (see Figure 1). It is in each player's interest to try to anticipate the choices of the other player in order to obtain a mutually beneficial outcome, and each player knows that the other is similarly motivated. If communication between the players is impossible, as it often is in naturally occurring pure coordination games, then subtle and interesting psychological problems arise, and the outcome may depend on the level of strategic intuition of the players.

Figure 1 about here

From an abstract, mathematical point of view, there seems to be nothing to choose between heads and tails, and a rational player may as well choose arbitrarily, with a probability of successful coordination of 1/2. But suppose that both players perceive heads to be more prominent, conspicuous, noticeable, or in some sense 'obvious' than tails, and

suppose further that each believes that the other is also likely to have noticed this asymmetry in the game. Surely in that case rational players will unhesitatingly choose heads, and the probability of successful coordination will be greater than 1/2? If one combination of strategies or outcome in a pure coordination game is uniquely salient, then it is said to constitute a focal point. Two question then arise. First, is it rational for players to choose the focal point? Second, why do players in fact tend to choose it? In this article I offer a new psychological explanation of focusing and a qualified rational justification of it based on recent developments in game theory.

It turns out to be no simple matter to justify focusing on rational grounds, and some of those who have examined the question have concluded that focusing is not something that is to be expected of rational players. The difficulty of providing a rational justification for focusing in pure coordination games is called the problem of coordination (Gauthier 1975). Some problems are difficult to understand because of their complexity; this problem is difficult to understand partly because of its very simplicity. Intuitively, a focal point is often such a natural and obvious choice that even a child can see it; and for this reason it seems surprising that focusing should raise difficulties for rational choice theory and game theory. The problem is both real and difficult, however, and it has attracted increasing attention from game theorists and experimental researchers in recent years (e.g., Bacharach 1993; Colman 1995: 33-40; Crawford and Haller 1990; Farrell 1987, 1988; Farrell and Gibbons 1989; Gauthier 1975; Gilbert 1989, 1990; Heal 1978; Mehta et al. 1994a, 1994b; Sugden 1991, 1993, 1995a, 1995b; Young 1993).

The discussion that follows begins in section 2 with some preliminary comments and definitions of pure coordination games in general and matching games in particular. Next, in section 3, I summarize the experimental evidence regarding focusing and discuss it in the light of the literature on framing effects. In section 4 I outline the standard game-theoretic assumptions regarding players' rationality and information states during the game. In section

5 I discuss the problem of explaining or justifying salience as a criterion of choice in games with and without payoff-dominant equilibria, and in section 6 I argue that if pure coordination games are modelled appropriately, then the problem of salience reduces to the problem of payoff dominance. In section 7 I outline a solution to the problem of payoff dominance in terms of the Stackelberg heuristic, and finally in section 8 I draw the threads together and return to the problem of coordination.

2. PURE COORDINATION AND MATCHING GAMES

It will be useful at this point to draw attention to a potential source of terminological confusion. In recent years, a number of authors (e.g., Cooper et al. 1990, 1992a, 1992b; Van Huyck et al. 1990) have begun to use the term coordination game to denote any game exhibiting multiple Nash equilibria that are Pareto-rankable, including games in which the players do not have identical preferences over the set of outcomes. In this article, the term pure coordination game will be reserved for any game in which the players have identical preferences and thus receive identical payoffs in every outcome of the game.

Pure coordination games often have more than one Nash equilibrium, but although the players' payoffs in each particular outcome are identical, they need not necessarily be invariant across equilibria, as they happen to be in the case of Heads and Tails (Figure 1). In other words, in some pure coordination games, the players jointly and severally prefer some Nash equilibria to some others, and the equilibria are therefore Pareto-rankable. In the terminology of this article, therefore, pure coordination games form a proper subset of the set of coordination games studied by Cooper et al. (1990, 1992a, 1992b) and Van Huyck et al. (1990).

Heads and Tails is an example of an important class of pure coordination games called matching games (Bacharach 1993; Bacharach and Bernasconi 1994; Colman and Bacharach 1996; Sugden 1995a) or unanimity games (Kalai and Samet 1985), defined by the following

properties: (a) there are two players with identical strategy sets and preferences over outcomes, and (b) the payoffs are positive along the main diagonal of the payoff matrix, that is, where the players choose the same strategy, and zero elsewhere.

Formally, a finite two-person matching game is a square matrix $\underline{A}$ with $\underline{m}$ rows and $\underline{m}$ columns, the matrix elements being ordered pairs ($\underline{v}$, $\underline{w}$), such that, for any particular row $\underline{i}$ and column $\underline{j}$,

$$\underline{i} = \underline{j} \Rightarrow \underline{v} > 0 \text{ and } \underline{w} > 0,$$

$$\underline{i} \neq \underline{j} \Rightarrow \underline{v} = \underline{w} = 0.$$

The matrix $\underline{A}$ is what I shall call a raw payoff matrix -- the reason for the qualifier 'raw' will become clear in section 6. The $\underline{m}$ rows of $\underline{A}$ correspond to elements of a set $\underline{M}$ from which Player I chooses a single element $\underline{i}$, the $\underline{m}$ columns correspond to the same set from which Player II chooses a single element $\underline{j}$, and the ordered pairs ($\underline{v}$, $\underline{w}$) are the payoffs to Player I and Player II respectively. I have defined matching games for the two-person case, but the definition generalizes straightforwardly to multi-person games. Matching games may be thought of as games in which both (or all) players are rewarded for performing some operation on the same element of a specified set $\underline{M}$ of elements. These games expose the problem of coordination in its clearest and simplest form.

Before Schelling's (1960) 'reorientation of game theory', pure coordination games were generally considered to fall outside the ambit of the theory, or at least to be uninteresting and unproblematical. Luce and Raiffa (1957), for example, claimed in their influential monograph on Games and Decisions that any group of decision makers 'which can be thought of as having a unitary interest motivating its decisions can be treated as an individual in the theory' (p. 13); that if the players have the same preferences over the outcomes 'then everything is trivial' (p. 59); and that 'certainly in the extreme case where there is perfect agreement the analysis is trivial' (p. 88). In spite of these and similar discouraging remarks from early commentators, pure coordination games have attracted increasing attention from

game theorists and experimental researchers in recent years (e.g., Anderlini 1995; Bacharach 1993; Bacharach and Bernasconi 1994; Bicchieri 1993; Colman 1995: 33-40; Colman and Bacharach 1996; Crawford and Haller 1990; Gauthier 1975; Gilbert 1989, 1990; Heal 1978; Janssen 1993; Mehta et al. 1994a, 1994b; Sugden 1991, 1995a; Young 1993).

3. FOCUSING AND FRAMING

In an informal empirical investigation of choices in the game of Heads and Tails, Schelling (1960, chap. 3) found that 86 per cent of his American subjects chose heads. Mehta et al. (1994a, 1994b) replicated this finding in England, where an almost identical 87 per cent of subjects chose heads. According to Schelling, what enables a majority of people to focus on heads is a culturally established 'conventional priority, similar to the convention that dictates A, B, C, though not nearly so strong' (p. 64) of heads over tails. Three further pure coordination games investigated by these researchers are worth summarizing briefly.

Schelling (1960, chap. 3) invited members of small groups to write down any amount of money, imagining that if everyone wrote down the same amount, then they would each receive that amount as a prize, otherwise no one would receive anything. This is a multi-person matching game with a countably infinite strategy set available to each player. Schelling found that 30 per cent of his subjects succeeded in focusing on one million dollars (the modal choice), and 93 per cent chose a 'round sum' that was a power of 10. Mehta et al. (1994a, 1994b) investigated a slightly different game in which English subjects were asked simply to 'write down any positive number' with the aim of choosing the same number as their partners, without the stipulation that the numbers represented prize values. The modal choice was 1 (chosen by 29 per cent), and the second favourite was 10 (chosen by 19 per cent).

In another informal experiment, Schelling (1960, chap. 3) presented subjects with the following hypothetical dilemma: a pair of strangers are given the opportunity to meet in New

York at a particular time, but no specific meeting place is designated. They both want to keep the appointment, but they cannot get in touch with each other before the appointed time, and their problem is to choose a good place where they may succeed in meeting. Schelling's subjects had to recommend a suitable location where the pair might meet up. This is a fuzzy two-person matching game with an indefinitely large strategy set available to each player, but more than 50 per cent of the subjects, who were students in New Haven, Connecticut, managed none the less to focus on the information booth at New York's Grand Central Station. In Mehta et al.'s (1994a, 1994b) British replication, based on London rather than New York, the modal choice was Trafalgar Square (38 per cent).

Finally, Schelling (1960, chap. 3) invited subjects who had already attempted the previous game to imagine that the pair had been told the place and date but not the time of the meeting. The subjects were asked to choose a suitable time to within a minute. This is another two-person matching game, each player's strategy set consisting of exactly 1440 strategies (because there are 60 minutes in an hour and 24 hours in a day, and $60 \times 24 = 1440$), yet virtually all of the subjects chose 12 noon. Mehta et al. (1994a, 1994b), using a slightly different pure coordination game, also found the modal choice to be 12 noon, followed by 12 midnight.

What is striking about these findings is the apparent efficiency with which human decision makers manage to focus on salient Nash equilibria, sometimes against enormous mathematical odds. The fact that 86 or 87 per cent of subjects playing Heads and Tails chose heads, and hence that about three-quarters ($.86 \times .86 = .74$) effectively coordinated, is remarkable; the fact that a substantial proportion of subjects succeeded in focusing on a time and place to meet in a large city, despite the dauntingly vast number of options, is almost breathtaking and certainly in need of explanation.

Successful coordination in the examples discussed above depends, of course, on the fact that the decision makers share not only basic cognitive structures and dispositions but also a

common body of background knowledge and information that leads them to interpret the problems in similar ways. This insight can be traced back at least two and a half centuries to David Hume's (1739-40, Bk. III, Part II, Sect. III) discussion of a mouth-watering coordination problem involving a German, a Frenchman, and a Spaniard who come across three bottles of wine, namely Rhenish, Burgundy, and port, and face the problem of how to share the bottles fairly among themselves. Of all the 27 possible ways of assigning three bottles to three people, or the six possible permutations assuming that each person gets exactly one bottle, Hume pointed out that the salient and unique solution to avoid a quarrel would be to 'give every one the product of his own country' (p. 510n). This focal-point solution emerges from the interpretation (or description) of the game, or what has been called the common language shared by the players (Crawford and Haller 1990). If the problem is considered in the abstract -- how to choose a permutation of three objects -- there is clearly no unique solution. Crawford and Haller pointed out that removal of the players' common language, including the culturally determined labels of the strategies, would have the effect of filtering out salient Nash equilibria and would make focusing impossible.

When confronted with any problem of choice, a human decision maker inevitably interprets it in a particular way, and the nature of the interpretation can, in some circumstances, play a part in determining the ensuing choice. The effects of such variations in the interpretations or descriptions of problems are called <u>framing effects</u>, because they are mediated by the decision maker's interpretive frame of reference (Bell <u>et al.</u> 1988; Colman 1995: 94-6, 154-60, 224-5; Fischhoff 1983; Kahneman and Tversky 1982a, 1984; Slovic <u>et al.</u> 1982; Tversky and Kahneman 1981, 1982, 1988). In a pure coordination game, the salience of a Nash equilibrium may arise either from a mathematical property of the game -- for example, one of the Nash equilibria may yield higher payoffs than any other -- or from the way the players frame the game. If the salience arises as a framing effect, as in Hume's problem of the three wines or most of the experimental games discussed in this section, then

it is a property of the game that is lost in translation when the game is formulated as a mathematical model in the conventional way; in fact, games that differ only in the labelling of their strategies are generally regarded as isomorphic in conventional game theory (Harsanyi and Selten 1988: 70-4). In variable-frame theory, however, salience is formally modelled even when it is a framing effect (see section 6 below).

The concept of a <u>focal point</u> was introduced by Schelling (1960: 57), who defined it as a strategy combination or outcome with properties of 'prominence or conspicuousness'. Lewis (1969: 35) introduced the term <u>salience</u>, which he defined, first, as the property of an outcome of 'standing out from the rest by its uniqueness in some conspicuous respect' and, second, as 'being unique in some way everyone will notice, expect the others to notice, and so on'. The first of Lewis's definitions implies that there can be only one salient outcome of a game, whereas his second definition is significantly weaker inasmuch as it allows the possibility of multiple salient outcomes. In this article I shall adopt the weak definition and assume that a salient outcome need not necessarily be unique. Indeed, in the case of a 2×2 matching game, a unique salient outcome is a logical impossibility, because if one strategy has a property that makes it salient, then the other must <u>ipso facto</u> be salient by virtue of lacking that property. I shall follow Bacharach's (1993) variable-universe or variable-frame theory (see also Bacharach and Bernasconi 1994) in considering the salience of a strategy combination or outcome to be a variable, the degree of salience depending on how rare and how noticeable its attributes are in the player's interpretive frame.

4. RATIONALITY AND INFORMATION ASSUMPTIONS

Because I am seeking to provide an explanation for coordination, I must be clear about what assumptions are to be made about the nature of the players' rationality and knowledge when thinking about the game. The standard game-theoretic rationality and knowledge assumptions are called <u>common knowledge of rationality</u> or CKR. They are based on the more general concept of <u>common knowledge</u>, which was introduced into the literature of game theory by Lewis (1969: 52-68) and subsequently formalized by Aumann (1976). A proposition is common knowledge among a set of players if every player knows it to be true, knows that the other players know it to be true, knows that the other players know that the other players know it to be true, and so on. On this basis, CKR consists of the following pair of assumptions (see, e.g., Sugden 1991: 765):

CKR1. The specification of the game, including the rules, the players' strategy sets, and the payoff functions, and every proposition that can be proved about the game by logical reasoning, are common knowledge in the game.

CKR2. The players are rational in the sense of expected utility (SEU) theory, that is, they always seek to maximize their own individual expected utilities, and this is common knowledge in the game.

An immediate logical implication of these assumptions is the <u>transparency of reason</u> (Bacharach 1987), an important characteristic of game-theoretic reasoning according to which, if one of the players has deduced something about the game on the basis of CKR, then the fact that the player has deduced it, and the deduction itself, are common knowledge in the game.

The methodology of game theory generally rests on the postulates of SEU, in spite of the recent emergence of alternative non-SEU models of decision making under uncertainty, notably the theory of probabilistic sophistication (Machina 1982a, 1982b) and regret theory (Bell, 1982; Loomes and Sugden 1982, 1986). As pointed out by Machina (1991), these non-

SEU theories are unlikely to be widely adopted unless they are shown to provide a better fit than the SEU postulates to existing empirical data and to be capable of yielding theoretical models approximating the elegance and power of SEU-based models. In the field of game theory, it is clear that neither of these goals has been achieved, and the standard SEU postulates are therefore generally preferred.

In the light of CKR, it is clear from von Neumann and Morgenstern's (1944, section 17.3.3) indirect argument that, in a pure coordination game (or any other game), if there is a unique solution, then it must be a Nash equilibrium. For a two-person game, the gist of the argument, informally expressed, is as follows (and it generalizes straightforwardly to multi-person games). A Nash equilibrium is a strategy combination in which each strategy is a payoff-maximizing best reply to the other. Rational players can discover each other's thinking about the game through the transparency of reason, and they are rational in the sense of expected utility theory. Therefore, if CKR leads Player I to choose Strategy $\underline{X}$ and Player II Strategy $\underline{Y}$, and if $\underline{X}$ and $\underline{Y}$ are uniquely mandated by rationality, then $\underline{X}$ and $\underline{Y}$ must be best replies to each other, because each player knows the other's intended strategy and, being a utility-maximizer, necessarily chooses the best reply to it. Therefore, because $\underline{X}$ and $\underline{Y}$ are best replies to each other, they constitute a Nash equilibrium by definition.

In a pure coordination game, the indirect argument could not on its own provide a method of finding the unique solution, even if there were one, because there are in general multiple Nash equilibria, and if the players chose strategies corresponding to different equilibria, then the outcome might not be an equilibrium at all. It is clear from the indirect argument, however, that rational players will never deliberately choose strategies that are out of equilibrium, so we may at least confine our attention to the equilibria. As Schelling (1960) was the first to show, informal analysis of the equilibria in a pure coordination game often reveals one or more salient outcomes or focal points, and empirical observations have confirmed that human decision makers often manage to focus on a salient Nash equilibrium

even if, as in Heads and Tails and some of the other games that were discussed in section 3, it does not offer a better payoff to the players than any other.

5. IS FOCUSING A RATIONAL PROCESS?

In spite of empirical evidence discussed in section 3, which shows that people often succeed in solving pure coordination games by focusing on salient Nash equilibria even when the mathematical odds seem unfavourable, Gilbert (1989, 1990) developed an argument that apparently showed, to her own surprise, that successful coordination cannot be the result of rational decisions by the players: 'If human beings are -- happily -- guided by salience, it appears that this is not a consequence of their rationality' (p. 61). She argued specifically that

> mere salience is <u>not</u> enough to provide rational agents with a reason for action (though it would obviously be nice, from the point of view of rational agency, if it did). We may grant that salience is the only thing that <u>could</u> provide a reason here. We must conclude, however, that it fails to do so. (p. 69, Gilbert's emphasis)

Gilbert's argument is disarmingly simple. In the game of Heads or Tails, suppose that both players recognize the outcome <u>HH</u> to be the unique focal point of the game, and both players know that their co-player recognizes it also, perhaps (to make the argument fully transparent) because they discussed the game before playing it and have agreed that <u>HH</u> is uniquely salient. It turns out that CKR provides the players with no reason to choose strategies corresponding to this salient outcome. According to CKR2, the players will seek to maximize their own individual expected utilities, but choosing heads maximizes expected utility only if the other player also chooses it. It would certainly be rational for Player I to choose heads if there were a reason to expect Player II to choose heads, but according to CKR2 Player II would have a reason to choose it only if there were a reason to expect Player

I to choose it, and so both players are drawn into an infinite regress that offers neither of them any valid basis for rational choice.

What is worse, in this game and all other pure coordination games, rationality appears to be self-defeating, to borrow a term from Parfit's (1979, 1984, chap. 1) discussion of a different though related family of games. Rationality is self-defeating because if Player I knew that Player II had a propensity to choose salient Nash equilibria even in the absence of any rational justification, which would mean that Player II was irrational according to the CKR assumptions, then this fact in conjunction with CKR2 would provide Player I with a rational justification for also choosing the salient Nash equilibrium, and coordination would be achieved without difficulty, to the mutual benefit of the players. The major conclusion that focusing in pure coordination games cannot be justified in terms of the standard rationality assumptions is widely accepted in the literature of game theory (e.g., Bicchieri 1993: 67-72; Lewis 1969: 35-7; Schiffer 1972: 145-7; Sugden 1991: 774-8; Ullman-Margalit 1977: 112).

Figure 2 about here

In the game of Heads and Tails, the salience of one of the outcomes arises as a framing effect and not as an inherent mathematical property of the game. In pure coordination games in which salience is intrinsic to the game's mathematical structure, the argument that it provides no rational justification for focusing apparently holds good and seems even more paradoxical. This applies to games involving payoff dominance, such as the simple 2×2 game shown in Figure 2, which is a modified form of Heads and Tails in which the players both prefer the HH Nash equilibrium, which yields payoffs of (2, 2), to the other TT equilibrium, which yields payoffs of (1, 1). There is also a mixed-strategy equilibrium, with each player assigning a probability of 1/3 to heads and 2/3 to tails, but it yields even lower payoffs of (2/3, 2/3). The HH equilibrium is said to payoff-dominate the TT equilibrium,

because each player receives a higher payoff if the outcome is HH than if it is TT; in other words, HH is the Pareto-dominant Nash equilibrium. If e and f are any two Nash equilibria in a game, then e (strongly) payoff-dominates f if it yields a better payoff than f to both (or every) player. The payoff dominance principle is the assumption that, in any game, if one equilibrium point e payoff-dominates all others, then rational players will play their parts in e. Harsanyi and Selten's (1988) general theory of equilibrium selection in games is based on this assumption, and most game theorists have accepted its intuitive force (e.g., Bacharach 1993; Cooper et al. 1990, 1992a, 1992b; Crawford and Haller 1990; Farrell 1987, 1988; Gauthier 1975; Lewis 1969; Sugden 1995a). But, perhaps surprisingly, a modified version of Gilbert's (1989, 1990) argument appears to show that the payoff dominance principle, though intuitively compelling, is without rational justification.

   In the game shown in Figure 2, for example, it seems obvious that rational players will choose the HH equilibrium. But Player I has no reason to choose heads in the absence of any ground for believing that Player II will choose heads, because if Player II were to choose tails, then Player I would do better by also choosing tails. Is there any ground for believing that Player II will choose heads? Apparently not; Player II is in exactly the same position as Player I, lacking a reason to choose heads in the absence of any ground for believing that Player I will choose heads, and we are back in the same infinite regress as before, without any reason for choosing the payoff-dominant HH Nash equilibrium in spite of its intuitive appeal. The payoff dominance principle is clearly in need of some form of justification, and this problem will be addressed in section 7. First, though, I must show an important logical connection between the problem of coordination and the problem of rationalizing the payoff dominance principle.

6. REDUCING PURE COORDINATION TO PAYOFF DOMINANCE

The following Blue Marble game (based on a generalization of a game in Bacharach and Bernasconi 1994) models any matching game with a single focal point. It will provide a vivid and unambiguous prototype of the problem of coordination. Two players are shown $m + 1$ marbles, m of which are white and one blue. Each player chooses one of the marbles without knowledge of the other player's choice. If both choose the same marble, then each receives a payoff of one unit; if they choose different marbles, then each receives a payoff of zero. If both players choose a white marble, it is assumed that there is some way of deciding whether they have chosen the same marble, but the white marbles are indistinguishable to the players themselves when the game is being played, and m is large.

Intuition suggests strongly that rational agents playing the Blue Marble game would focus on the blue marble, although it is not immediately obvious why this is so. Before explaining in section 7 what I believe to be the reason, I shall first show that the problem of coordination in the Blue Marble game, and therefore in pure coordination games in general, can be boiled down to the problem of explaining and justifying the payoff dominance principle.

The raw payoff matrix of the Blue Marble game is an $(m + 1) \times (m + 1)$ square matrix $\underline{A}$, in which each row and column corresponds to one of the marbles that may be selected. But this cannot be a correct specification of the strategic form of the game, because the players do not have the capacity to choose from among $m + 1$ strategies. The white marbles are assumed to be indistinguishable to the players, so there are really only three ways of acting that are discriminable by them:

$\underline{B}$: choose the blue marble;

$\underline{W}_r$: pick a white marble at random;

$\underline{M}_r$: pick a marble at random.

(Whether or not a human decision maker is capable of picking at random without the aid of a randomizing device could be debated; I shall assume here that a player can choose arbitrarily and that arbitrary choices are at least quasi-random.) The strategic payoff matrix of this game is shown in Figure 3.

Figure 3 about here

There are three pure-strategy Nash equilibria in this game -- BB, $W_r W_r$, and $M_r M_r$ -- of which BB yields the highest expected payoff to both players and $W_r W_r$ the next highest. If both players choose B, then they coordinate with certainty; if they both choose one of the other pure strategies, then they coordinate with probability $0 < p < 1$. Both players achieve a higher expected payoff by choosing BB than any other combination of strategies, pure or mixed. It is clear in the strategic payoff matrix shown in Figure 3 that BB is the unique payoff-dominant Nash equilibrium of the Blue Marble game. If there is a rational justification for choosing the payoff-dominant Nash equilibrium in a game of this type, then it must also be rational justification for focusing on the salient strategy combination. I shall now show that the Blue Marble game may be viewed as a template for matching games in general.

Consider a modification of the Blue Marble game, differing from it only in that all $m + 1$ marbles are white and none is blue; I shall call this the White Marble game. Player I and Player II once again try to coordinate by choosing the same marble, but now they do not have an obvious focal point determined by colour to guide them. If the $m + 1$ marbles are truly indistinguishable, then there is nothing for it but for each player to pick a marble at random, and each player's (invariant) expected payoff is $1/(m + 1)$. But suppose we modify the description of the game by supposing that one of the marbles has some unspecified mark or imperfection that the players may or may not notice. If one of the players notices the salient marble, then that player may then assign a subjective probability to the event of the other

player also noticing it. If both players are certain that their co-player has noticed the salient marble, then both subjective probabilities equate to unity, and if the players focus on the salient Nash equilibrium they achieve successful focal-point coordination. In that case, the White Marble game is isomorphic with the Blue Marble game, and a rational justification for focusing would follow as a corollary of any rational justification of the payoff dominance principle.

In variable-frame theory (Bacharach 1993, Bacharach and Bernasconi 1994), a player's awareness of a salient attribute is a stochastic variable and, having noticed it, a player assigns a subjective probability to the event of the other player noticing it also. If both players in the White Marble game notice the salient marble -- if the mark or imperfection belongs to both players' interpretive frames -- but each player assigns a probability of $0 < \underline{p} < 1$ to the event of the other player noticing it, then a focal point exists and the players' expected utilities are higher if they play their parts in it than if they do not, and the focal point is therefore still a payoff-dominant Nash equilibrium in the strategic payoff matrix, properly specified, which implies that its selection would be rationally justified by any justification of the payoff dominance principle. Finally, if there are two or more salient options, and if both players notice them and judge the probabilities that their co-players have noticed them to be nonzero, then if both players choose strategies that involve picking one of the salient options at random the outcome is once again a payoff-dominant Nash equilibrium and essentially the same argument applies.

7. THE STACKELBERG HEURISTIC

It is unexpectedly difficult to provide a convincing rational justification for the payoff dominance principle. What accounts for the powerful intuition that rational players are bound to choose the payoff-dominant HH Nash equilibrium in the maximally simple matching game shown in Figure 2? Suggestions have relied either on repeated plays of the game (e.g., Aumann and Sorin 1989; Crawford and Haller 1990); 'cheap talk' in the form of costless pre-play communication between the players (e.g., Anderlini 1995; Farrell 1987, 1988; Farrell and Gibbons 1989; Kalai and Samet 1985; Matthews 1989; Matthews and Postlewaite 1989); or some form of team thinking or collective rationality that relies on something other than the standard CKR assumptions of game theory (Hurley 1991; Janssen 1993; Sugden 1993, 1995a). All of these suggestions involve radical departures from the standard rationality/information assumptions of game theory or the rules of the game. Colman and Bacharach (1996) have recently put forward an explanation in terms of what they call the Stackelberg heuristic, which relies on a less drastic departure from the standard game-theoretic assumptions. It provides a psychological explanation, by which I mean an elucidation of the intuitive appeal of payoff-dominant Nash equilibria and an answer to the question of why human decision makers choose in practice. In addition, it provides a qualified rational justification for the principle of payoff dominance in pure coordination games and other games of common interests (Aumann and Sorin 1989), but it is not always applicable in other classes of games.

The Stackelberg heuristic is an extension of an idea first suggested by von Neumann and Morgenstern (1944, section 14.4.1: 100-104) to justify the selection of maximin strategies in strictly competitive (two-person zero-sum) games. The key assumption is that players reason by using a type of simulation heuristic (Kahneman and Tversky 1982b). Player i uses the heuristic by choosing a utility-maximizing strategy on the assumption that any strategy choice will be met by a counterstrategy that maximizes Player II's utility, as if Player j were

choosing second in a perfect-information game with knowledge of Player $\underline{i}$'s choice. Player $\underline{i}$ assumes that Player $\underline{j}$ can anticipate Player $\underline{i}$'s thinking about the game as if choosing after Player $\underline{i}$ rather than simultaneously with Player $\underline{i}$. This implies that for every strategy $\underline{s}_i$ in Player $\underline{i}$'s strategy set $\underline{S}$, Player $\underline{j}$ will respond with a strategy $\underline{f}$ such that $\underline{f}(\underline{s}_i)$ is a best reply of Player $\underline{j}$ to $\underline{s}_i$, so that

$$\underline{H}_j(\underline{s}_i, \underline{f}(\underline{s}_i)) \geq \underline{H}_j(\underline{s}_i, \underline{s}_j) \quad \forall \, \underline{s} \in \underline{S}.$$

If we accept a slightly strengthened assumption based on the transparency of reason, namely that <u>any</u> conclusion reached by Player $\underline{i}$ about what strategy to choose will be anticipated by Player $\underline{j}$, then Player $\underline{i}$ expects this to happen and chooses a utility-maximizing strategy based on the assumption that $\underline{j}$'s response will invariably be a best reply $\underline{f}(\underline{s}_i)$. In other words, Player $\underline{i}$ chooses a counterstrategy for which $\max_i \underline{H}_i(\underline{s}_i, f(\underline{s}_i))$ is attained. The strengthening of the transparency of reason involves its extension to embrace forms of reasoning that do not flow automatically from standard rationality assumptions. The modified assumption is that, if Player $\underline{i}$ reaches a conclusion about a strategy choice through a form of reasoning such as the Stackelberg heuristic, then the conclusion will be known to Player $\underline{j}$ and will in fact be common knowledge in the game. It can then be proved that players always focus on payoff-dominant Nash equilibria in pure coordination games, and indeed in all games of common interests (Colman and Bacharach 1996).

According to the Stackelberg heuristic, in the matching game shown in Figure 2 Player I will run the following simulation. Imagine that Player I were to choose first, and Player II were to choose second with knowledge of Player I's choice. If the rules of the game allowed such sequential choosing, then Player II, being rational, would choose a best reply, namely heads if Player I chose heads and tails if Player I chose tails. Player I would receive a higher payoff in the first case than the second. Therefore, back in the real game, Player I acts as if playing the sequential version and chooses what would be the utility-maximizing strategy if Player II were to respond to any choice with a best reply, so Player I chooses heads. Player

I's thinking is transparent to Player II, so Player II anticipates Player I's choice of heads and also chooses heads, which is the best reply. The players therefore focus successfully on the payoff-dominant <u>HH</u> Nash equilibrium. The same conclusion is reached if the mirror-image simulation is run with Player II choosing first. Colman and Bacharach (1996) proved that this form of reasoning leads inexorably to the payoff-dominant outcome in any pure coordination game that possesses one.

The Stackelberg heuristic seems a common-sense way of choosing a strategy in a pure coordination game. It is not unreasonable for a player to predict what strategies the other player is likely to choose, and then to choose the optimal strategy in the light of these predictions and the assumption that the other player will think along similar lines so that the conclusion will be common knowledge. This type of heuristic thinking seems especially justifiable when there is no other obvious basis on which to choose. In matching games and other pure coordination games, it is not clear that there is a better method of choice than the Stackelberg heuristic. Salient Nash equilibria or focal points are intuitively compelling in such games, and the Stackelberg heuristic provides an explanation for this otherwise mysterious psychological phenomenon. In addition to this, there is, in fact, some experimental evidence showing a strong bias towards the strategies prescribed by the Stackelberg heuristic in a variety of two-person games. In an experiment in which 100 subjects made one-off choices in all 12 strategically nonequivalent symmetric 2×2 games, nine of which are Stackelberg-soluble and three non-Stackelberg-soluble, Colman and Stirk (1996) found an overwhelming preponderance (between 78 and 98 percent) of Stackelberg strategies in the Stackelberg-soluble games and variable strategy choices in the other games. This suggests that human decision makers do tend to use the Stackelberg heuristic in choosing their strategies.

8. DISCUSSION AND CONCLUSIONS

Any pure coordination game, and in particular any nondegenerate matching game, provided that it is modelled appropriately with what I have called its strategic rather than its raw payoff matrix, is essentially a game with a payoff-dominant Nash equilibrium. I have shown that this interpretation is valid not only for games with payoff-dominant focal points in their raw payoff matrices, but also to games whose focal points are invisible in their basic mathematical structures but emerge only from culturally determined framing effects. It follows that the Stackelberg heuristic provides a psychological explanation for focusing in pure coordination games, because it has been shown to explain the payoff dominance principle -- the selection of payoff-dominant Nash equilibria in games of common interests, including pure coordination games. In view of the generally acknowledged fact that human rationality is bounded, a psychological explanation can potentially contribute to our understanding of economic behaviour, by explaining how decision makers think and behave, whether or not the form of reasoning that underlies it conforms strictly to the conventional postulates of rationality. From a methodological point of view, a rigid adherence to formal rationality is a priori unlikely to provide a realistic insight into human decision making.

As a psychological explanation, the Stackelberg heuristic elucidates the intuitive appeal of focal points in pure coordination games and their selection by experimental subjects, but whether or not it provides a rational justification for focal point selection and thereby solves the problem of coordination is a separate issue. If it solves this problem, then the solution is purchased at a price. The concession that is required is a strengthening of one of the assumptions derived from the common knowledge of rationality (CKR) assumptions about the players' states of information during the game. A standard and uncontroversial logical inference from the CKR assumptions is the transparency of reason -- the inference that, if one of the players deduces something about the game on the basis of CKR, then the fact that it has been deduced and the deduction itself must be common knowledge in the game. To rationalize the payoff dominance principle, a slightly stronger assumption is required, namely

that when a player chooses a strategy, whether or not the choice is derived logically from the CKR assumptions, the other player(s) will anticipate the choice, and the choice will be common knowledge in the game. This assumption is based on the same basic hypothesis that underpins the transparency of reason, namely that human decision makers tend to be like-minded, and know that they are. It is, in effect, a psychological counterpart of the purely logical transparency of reason.

The Stackelberg heuristic depends on a form of evidentialist reasoning that has been used to rationalize strategy choices in different types of games (e.g., Lewis 1979), but the validity of evidentialist reasoning is controversial. On the basis of evidentialist reasoning, Player i reaches a strategy choice from evidence that the choice itself provides of Player j's likely response. Assuming Player j's best-reply rationality to be transparent, it is indisputable that a strategy choice by Player i, arrived at by any reasoning process that is transparent to Player j, is evidence that Player j will choose a best reply to it. But whether it is rational for Player i to use evidence of that type to choose a strategy is the pivotal issue that divides 'evidentialists' and 'causalists' in decision theory and lies at the heart of the debate over Newcomb's problem, for example. Colman and Bacharach (1996) discussed this issue but remained neutral as to whether evidentialist reasoning is rationally justified, but they argued that the Stackelberg heuristic is justified provided that evidentialist reasoning is justified.

The Stackelberg heuristic may therefore be described as conditionally justified, and this article may therefore be said to provide a qualified solution to the problem of coordination, contingent on the rational justifiability of evidentialist reasoning. If evidentialist reasoning and therefore the Stackelberg heuristic are not rationally justifiable, then this article provides no justification for focusing in pure coordination games and no solution to the problem of coordination; but it none the less offers a psychological explanation for the Stackelberg heuristic in general and for focusing in pure coordination games in particular. For it is quite clear that salient Nash equilibria are intuitively obvious choices in such games and that

human decision makers do, in practice, choose them, and both of these phenomena need to be explained.

ACKNOWLEDGEMENTS

REFERENCES

Anderlini, Luca (1995) 'Communication, computability and common interest games', Economic Theory Discussion Paper 159, Department of Applied Economics, University of Cambridge.

Aumann, Robert J. (1976) 'Agreeing to disagree', Annals of Statistics 4: 1236-39.

Aumann, Robert J. and Sorin, Sylvain (1989) 'Cooperation and bounded recall', Games and Economic Behavior 1: 5-39.

Bacharach, Michael (1987) 'A theory of rational decision in games', Erkenntnis 27: 17-55.

Bacharach, Michael (1993) 'Variable universe games', in K. Binmore, A. Kirman and P. Tani (eds) Frontiers of Game Theory, Cambridge, MA: MIT Press, pp. 255-75.

Bacharach, Michael and Bernasconi, Michele (1994) 'An experimental study of the variable frame theory of focal points', working paper, Institute of Economics and Statistics, University of Oxford.

Bell, David E. (1982) 'Regret in decision making under uncertainty' Operations Research 30: 961-81.

Bell, David E., Raiffa, Howard and Tversky, Amos (1988) 'Descriptive, normative, and prescriptive interactions in decision making', in David E. Bell, Howard Raiffa and Amos Tversky (eds) Decision Making: Descriptive, Normative, and Prescriptive Interactions, Cambridge: Cambridge University Press, pp. 9-30.

Bicchieri, Christina (1993) Rationality and Coordination, Cambridge: Cambridge University Press.

Colman, Andrew M. (1995) Game Theory and its Applications in the Social and Biological Sciences (2nd ed.), Oxford: Butterworth-Heinemann.

Colman, Andrew M. and Bacharach, Michael (1996) 'Payoff dominance and the Stackelberg heuristic', Theory and Decision in press.

Colman, A. M. and Stirk, J. A. (1996) 'Stackelberg reasoning in mixed-motive games: An

experimental investigation', paper read at the International Association for Research in Economic Psychology Conference, Paris, September.

Cooper, Russell W., DeJong, Douglas V., Forsythe, Robert and Ross, Thomas W. (1990) 'Selection criteria in coordination games: some experimental results', <u>American Economic Review</u> 80: 218-33.

Cooper, Russell W., DeJong, Douglas V., Forsythe, Robert and Ross, Thomas W. (1992a) 'Communication in coordination games', <u>Quarterly Journal of Economics</u> 107: 739-71.

Cooper, Russell W., DeJong, Douglas V., Forsythe, Robert and Ross, Thomas W. (1992b) 'Forward induction in coordination games', <u>Economics Letters</u> 40: 167-72.

Crawford, Vincent P. and Haller, Hans (1990) 'Learning how to cooperate: optimal play in repeated coordination games', <u>Econometrica</u> 58: 571-95.

Farrell, Joseph (1987) 'Cheap talk, coordination, and entry', <u>Rand Journal of Economics</u> 18: 34-9.

Farrell, Joseph (1988) 'Communication, coordination and Nash equilibrium', <u>Economics Letters</u> 27: 209-14.

Farrell, Joseph and Gibbons, Robert (1989) 'Cheap talk can matter in bargaining', <u>Journal of Economic Theory</u> 48: 221-37.

Fischhoff, Baruch (1983), 'Predicting frames', <u>Journal of Experimental Psychology: Learning, Memory and Cognition</u> 9: 103-16.

Gauthier, David (1975) 'Coordination', <u>Dialogue</u> 14: 195-221.

Gilbert, Margaret (1989) 'Rationality and salience', <u>Philosophical Studies</u> 57: 61-77.

Gilbert, Margaret (1990) 'Rationality, coordination and convention', <u>Synthese</u> 84: 1-21.

Harsanyi, John C. and Selten, Reinhard (1988) <u>A General Theory of Equilibrium Selection in Games</u>, Cambridge, MA: MIT Press.

Heal, Jane (1978) 'Common knowledge', <u>Philosophical Quarterly</u> 28: 116-31.

Hume, David (1739-1740) <u>A Treatise of Human Nature: Being an Attempt to Introduce the</u>

Experimental Method of Reasoning Into Moral Subjects, London: Thomas Longman. [Edited by L. A. Selby-Bigge, Oxford: Clarendon Press, 1888, 2nd ed. 1978]

Hurley, Susan L. (1991) 'Newcomb's problem, Prisoner's Dilemma, and collective action', Synthese 86: 173-96.

Janssen, Maarten C. W. (1993) 'Rationalizing focal points', discussion paper, Erasmus University Rotterdam.

Kahneman, Daniel and Tversky, Amos (1982a), 'The psychology of preferences', Scientific American 246(1): 136-42.

Kahneman, Daniel and Tversky, Amos (1982b), 'The simulation heuristic', in Daniel Kahneman, Paul Slovic and Amos Tversky (eds) Judgment Under Uncertainty: Heuristics and Biases, Cambridge: Cambridge University Press, pp. 201-8.

Kahneman, Daniel and Tversky, Amos (1984), 'Choices, values, and frames', American Psychologist 39: 341-50.

Kalai, Ehud and Samet, Dov (1985) 'Unanimity games and Pareto optimality', International Journal of Game Theory 14: 41-50.

Lewis, David K. (1969) Convention: A Philosophical Study, Cambridge, MA: Harvard University Press.

Lewis, David K. (1979) 'Prisoner's Dilemma is a Newcomb Problem', Philosophy and Public Affairs 8: 235-40.

Loomes, Graham and Sugden, Robert (1982) 'Regret theory: An alternative theory of rational choice under uncertainty' Economic Journal 92: 805-24.

Loomes, Graham and Sugden, Robert (1986) 'Disappointment and dynamic consistency in choice under uncertainty' Review of Economic Studies 53: 271-82.

Luce, R. Duncan and Raiffa, Howard (1957) Games and Decisions: Introduction and Critical Survey, New York: Wiley.

Machina, Mark J. (1982a) '"Expected utility" analysis without the Independence Axion',

Econometrica 50: 277-323.

Machina, Mark J. (1982b) 'A stronger characterization of declining risk aversion',
Econometrica 50: 1069-79.

Machina, Mark J. (1991) 'Dynamic consistency and non-expected utility', in M. Bacharach
and S. Hurley (eds) Foundations of Decision Theory: Issues and Advances, Oxford:
Blackwell, pp. 39-91.

Matthews, Steven (1989) 'Veto threats: rhetoric in bargaining model', Quarterly Journal of
Economics 104: 347-69.

Matthews, Steven and Postlewaite, Andrew (1989) 'Pre-play communication in two-person
sealed-bid double auctions', Journal of Economic Theory 48: 238-63.

Mehta, Judith, Starmer, Chris and Sugden, Robert (1994a) 'Focal points in pure coordination
games: an experimental investigation', Theory and Decision 36: 163-85.

Mehta, Judith, Starmer, Chris and Sugden, Robert (1994b) 'The nature of salience: an
experimental investigation of pure coordination games', American Economic Review 84:
658-73.

Neumann, John von and Morgenstern, Oskar (1944) Theory of Games and Economic
Behavior, Princeton, NJ: Princeton University Press. [2nd ed., 1947; 3rd ed., 1953]

Parfit, Derek (1979) 'Is common-sense morality self-defeating?', Journal of Philosophy 76:
533-45.

Parfit, Derek (1984) Reasons and Persons, Oxford: Clarendon Press.

Schelling, Thomas C. (1960) The Strategy of Conflict, Cambridge, MA: Harvard University
Press.

Schiffer, Steven (1972) Meaning, Oxford: Oxford University Press.

Slovic, Paul, Fischhoff, Baruch and Lichtenstein, Sarah (1982) 'Response mode, framing,
and information processing effects in risk assessment', in R. M. Hogarth (ed.) New
Directions for Methodology of Social and Behavioral Science: the Framing of Questions

and the Consistency of Responses, San Francisco: Jossey-Bass, pp. 21-36.

Sugden, Robert (1991) 'Rational choice: a survey of contributions from economics and
philosophy', Economic Journal 101: 751-85.

Sugden, Robert (1993) 'Thinking as a team: towards an explanation of nonselfish behavior',
Social Philosophy and Policy 10: 69-89.

Sugden, Robert (1995a) 'A theory of focal points', Economic Journal 105: 533-50.

Sugden, Robert (1995b) 'The coexistence of conventions', Journal of Economic Behavior
and Organization 28: 241-56.

Tversky, Amos and Kahneman, Daniel (1981) 'The framing of decisions and the psychology
of choice', Science 211: 453-8.

Tversky, Amos and Kahneman, Daniel (1982) 'Rational choice and the framing of decisions',
in R. M. Hogarth (ed.) New Directions for Methodology of Social and Behavioral
Science: The Framing of Questions and the Consistency of Responses, San Francisco:
Jossey-Bass.

Tversky, Amos and Kahneman, Daniel (1988) 'Rational choice and the framing of decisions',
in David E. Bell, Howard Raiffa and Amos Tversky (eds) Decision Making: Descriptive,
Normative, and Prescriptive Interactions, Cambridge: Cambridge University Press, pp.
167-192.

Ullman-Margalit, Edna (1977) The Emergence of Norms, Oxford: Oxford University Press.

Van Huyck, John B., Battalio, Raymond C. and Beil, Richard O. (1990) 'Tacit coordination
games, strategic uncertainty, and coordination failure', American Economic Review 80:
234-48.

Young, H. P. (1993) 'The evolution of conventions', Econometrica 61: 57-84.

Figure 1

<u>Heads or Tails</u>

II

	<u>H</u>	<u>T</u>
<u>H</u>	1, 1	0, 0
<u>T</u>	0, 0	1, 1

I (to the left of the table)

Figure 2
Payoff Dominance

	II	
	H	T
H	2, 2	0, 0
T	0, 0	1, 1

I

Figure 3
Strategic Payoff Matrix for the Blue Marble Game

II

		$\underline{B}$	$\underline{W}_r$	$\underline{M}_r$
	$\underline{B}$	1, 1	0, 0	$1/(\underline{m} + 1)$, $1/(\underline{m} + 1)$
I	$\underline{W}_r$	0, 0	$1/\underline{m}$, $1/\underline{m}$	$1/(\underline{m} + 1)$, $1/(\underline{m} + 1)$
	$\underline{M}_r$	$1/(\underline{m} + 1)$, $1/(\underline{m} + 1)$	$1/(\underline{m} + 1)$, $1/(\underline{m} + 1)$	$1/(\underline{m} + 1)$, $1/(\underline{m} + 1)$