MARGARET GILBERT

RATIONALITY AND SALIENCE

(Received 6 June, 1988)

It turns out that sophisticated subjects in an experimental setting can often do very well — much better than chance — at solving novel coordination problems without communicating. They try for a coordination equilibrium that is somehow *salient*: one that stands out from the rest by its uniqueness in some conspicuous respect. It does not have to be uniquely *good*, indeed, it could be uniquely bad. It merely has to be unique in some way the subjects will notice, expect each other to notice, and so on. (Lewis (1969) p. 35; Lewis is reporting experimental findings of Schelling, see Schelling (1960).)

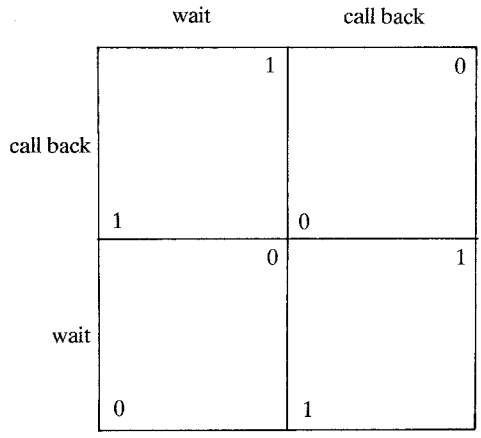
A number of authors have envisaged a model of the generation of action in coordination problems in which salience plays a crucial role. (Apart from Schelling and Lewis, I have in mind later work by Schiffer, Ullman-Margalit, and others.) As the above quotation indicates, appeal to salience accords with experimental results: empirical studies suggest that human subjects are likely to try for the salient combination of actions in situations of the relevant type. This tendency to choose the salient seems all to the good. If everybody chooses the salient, a fortunate result is obtained. There remains an important question in the theoretical understanding of action: how precisely does salience facilitate the successful coordination of actions in a coordination problem? Some have thought that rationality dictates that one do one's part in the salient solution. That this is so may even seem obvious: failing any other reason for action, surely one has reason to do one's part in the salient combination of actions? In this paper I shall argue that one does not. If human beings are — happily — guided by salience, it appears that this is not a consequence of their rationality. It looks as if it may be said: When we do our parts in the salient combination, we do not reason. We act blindly.[1]

I

Some preliminary clarification of terms is in order. A standard example

Philosophical Studies **57**: 61—77, 1989.

of a *coordination problem* is the telephone case: a telephone call is cut off, and each party has to decide what to do, whether to call back or to wait. Both want the call restored. The situation may be represented by the following payoff matrix.



Payoff matrix

(The cells of the matrix represent combinations of the agents' actions. A given cell represents the combination of the action described at the top of the relevant column, together with the action described at the side of the relevant row. The numbers in the lower left hand of each cell represent the payoff to the player whose available choices of action are described at the side of the rows, the numbers in the upper right represent the other player's payoffs, the actions referred to at the top of the columns represent his[2] available choices.)

I shall not attempt to settle on a formal definition of a coordination problem here.[3] For present purposes, the following general features may be assumed. First, the value of the outcome of any one person's action to that person will depend on what the other does. (In the telephone case, if and only if only one waits, and only one calls back, will the call be reconnected.) Second, the agents' rankings of the various possible combinations of agents' actions are identical. Hence, were it possible for the agents to make an agreement as to who would do what,

there would be nothing very problematic about the situation. Third, where it is not possible for the parties to make an agreement about who will do what, or otherwise to communicate their intentions, and there is no special background knowledge, the parties have a genuine problem: Should each make a random choice of action, the chances of achieving the desired coordination of actions are not good.

Our question can be put thus: will agents who are rational in the sense of mathematical game-theory find a reason for action in mere salience? In the context of game-theory the attributes of *rational agents* are implicitly held to include the following: (A) They are perfect reasoners, in particular, given the information available they will make no mistaken inferences. They will, moreover, utilize all relevant information that is available. (B) Each will act on the balance of reasons, if the balance of reasons dictates a particular course of action (which it may not do). (C) Each one is out to do as well as possible according to his/her individual ranking of the outcomes. In what follows, when I refer to 'rational' agents, I should be understood to be referring to agents with the features just mentioned.

In game-theoretic discussions of problems of coordination (and other problems) among rational agents, it has always been assumed that each agent's rationality, and each agent's preferences over the outcomes, are entirely out in the open between them. Thus each can assume that the other will reason appropriately, and that the other will assume that he will reason appropriately, and so on. In what follows I shall take it that, when discussing the situation of rational agents, we are discussing agents whose rationality and preferences are out in the open in this way. Following David Lewis, and, later, Robert Aumann in economics, it is now standard to write of 'common knowledge' in this connection.[4]

There is an issue as to whether and to what extent human beings are game-theoretically rational. I shall not enter into that question here, but what I have to say will have an obvious bearing on it, given the reported experimental data.

Following what seems to be the general understanding, I shall take it that a particular combination of actions is *salient* if and only if it is entirely out in the open among the agents concerned that this combination is "the odd man out" or "stands out from the rest" for all.[5] Given

that a certain combination of actions is salient in a certain situation each agent can assume that the premiss that the combination in question stands out for all will be recoverable from his reasoning by the others. In other words, he can take it for granted in his reasoning, while relying on the others being able to figure out that he does, and to figure out that he relies on this, and so on.

II

Salience in general may be a psychological matter, in the sense that what is salient depends on who is involved. Whether or not this is so, there remains the question whether salience, once it is present, will generate a reason for action for rational agents as such, other things being equal.

In the literature we find partisans of both sides on the question of reason-generation. David Lewis, who appeals to salience in presenting a model of the genesis of conventions, appears to assume that salience is *not* in itself reason-generating.[6] Meanwhile, others assume that it is. Indeed, as far as I can tell this has become a standard assumption among writers influenced by Lewis's account of conventions.[7] Possibly Lewis's use of a broadly game-theoretical framework has misled some people. It is quite easy to fall into the assumption that he has presented a flawless 'rationality-driven' model of a convention's origins.

Jane Heal argues for reason-generation explicitly against Lewis. Many others simply take it for granted. In this paper I shall, in effect, be presenting considerations in favour of Lewis's (unargued) position, and against that of Heal and others.

Is salience sufficient, ceteris paribus, to solve a coordination problem for all rational agents? To provide a focus for discussion, here is an example of a coordination problem containing a combination of actions which would (presumably) be salient at least in a population of Western academicians.[8]

Suppose that Sally Brown and Joe Smith, strangers to each other, have been kidnapped and put in separate cells. They have no means of communicating with each other. The kidnapper has given both of them a board with four coloured squares (red, blue, yellow and purple) and a button by each square. The boards are connected to a central unit

which records who has pressed which button. Neither Sally nor Joe can see the other's coloured board. At noon, the kidnapper makes the following announcement over a loudspeaker that can plainly be heard by both of them. 'At 12. 15 p.m. precisely I want each of you to press a button on your colour board. If, without communication, you both press the button for the same colour, you will both go free. If you are not unanimous, or someone fails to press a button, you will both die. It matters not at all which colour you choose. The only thing that matters is that you choose the same colour.' He pauses, then begins to speak again. 'To afford you some distraction before 12. 15 p.m., my secretary will turn on the radio for five minutes and play it over the loudspeaker. Have a good day!' After a moment the radio comes on. There is some music at first and then the announcer says: 'Today is our President's birthday. His wife is giving a party for him using red as the theme colour. Red is the President's favourite colour. There will be red tablecloths on the tables, the waiters will wear red jackets and ties, the President's wife will wear a red dress, red stockings and red shoes, and the food eaten will all be red: gaspacho soup, sole in a tomato sauce, and a strawberry dessert.' At this point the radio clicks off. Let us take it that when, a few minutes later, Sally and Joe look at their colour boards, the combination of actions constituted by each one's choosing red will 'stand out' for each one, and this will be entirely out in the open. In other words, this will be the salient combination of actions as far as they are concerned.

Now suppose that Sally and Joe are rational agents according to the characterization given above, and suppose that the following facts are entirely out in the open: (1) They are both rational. (2) Each wants to live. (3) Combination of actions C (choosing red) stands out for both. (4) *There are no other relevant facts bearing on the decision as to what is to be done.* Is each one now in a position to figure out what to do? This should be so, if, ceteris paribus, salience is a reason for action in an otherwise recalcitrant coordination problem. In my view, it is not. No reason for acting can be directly inferred at this point. This can be argued quite briefly, as follows.[9]

Given that C is salient, Sally can argue as follows: "C stands out for us both. Clearly, I should do my part in C if Joe does. But will he?" How can Sally figure out what Joe will do? Suppose she tries to look at

things from Joe's point of view, that is, to figure out how Joe will reason. She will by hypothesis see that Joe also knows that C stands out for both. She will also see that he is faced with the question whether to do his part in C. He will see that he should do his part if Sally does hers, and he will ask himself whether he has reason to think she will do her part. But it is clear that he will find no such reason; in particular, an attempt to replicate Sally's reasoning will get him nowhere.

Someone might wonder whether it would help matters if we stipulated that it was common knowledge that Sally and Joe were out to maximize *expected utility*. This would not seem to make any difference. The expected utility of a given action of Sally's will depend on the probabilities associated with the different actions open to Joe. But why should she assign a higher probability to his doing his part in C? The original argument arises again in terms of probability rather than proof.[10] Someone may suggest that Sally could rely on induction, knowing that people like herself and Joe tend to rely on the salient, and thus assign a higher probability to his doing the salient thing than to his doing the non-salient thing. But we are not supposing that there is any such background knowledge. We are considering precisely what Sally and Joe might figure out given only the salience of one combination of actions, plus common knowledge of rationality, as defined, and of the matrix. One reason for so doing is precisely to see whether a 'tendency' of the kind just referred to might just be, in effect, the result of the exercise of reason alone in a pared-down context of the kind in question. My argument is that any such tendency cannot be seen as having such a genesis.

It should now be apparent that if anything is going to give Sally or Joe a reason for action in the presence of salience, it is going to be some new background belief such as the belief that the other person will in fact do his part in the salient combination. It may help to reinforce the idea that salience does not itself provide a reason for action to consider a case that involves the introduction of a belief providing a reason *against* choosing the salient. Assume that, after his statement about the President's birthday, the announcer had presented the following news item: 'Sally Brown, who has been kidnapped along with Joe Smith, is said by her husband to be salience-shy. In other words, in a coordination problem with a salient combination of actions,

Sally may be expected not do her part in the salient combination: she will do her part in some other combination.' Whether or not Sally is indeed salience-shy, Joe now has reason to believe that she is, and Sally knows that he has this reason. In brief, each now has reason to avoid the salient option, and, surely, will be rational to choose some other option. Though we could think this while allowing that salience always provides a prima facie reason to choose the salient, it may well be judged in this case that it is not as if an existing prima facie reason for action has been overriden or outweighed by another reason.[11]

A rather special type of case may suggest this even more clearly. Suppose that Sally and Joe hear a single radio announcement, as follows; 'It is well known that Joe Smith and Sally Brown are phobic with respect to the colour red. Each will avoid choosing something red at any cost.' Presumably when they have heard this the combination of actions consisting in their both choosing red will be the salient combination of actions: it will definitely be the 'odd man out' and 'stand out' more than any of the other action combinations. But here the fact that makes choosing red salient is the fact that red *will not be chosen.* It seems hard to credit the idea that salience so derived could provide even a prima facie reason for action. And, if it does not, then it is false that salience in general provides such a reason.

Whatever one's immediate reactions to these cases, the argument I have given above seems to me conclusive against the idea that salience as such provides a reason for action. Let me now consider a counter-argument, from Jane Heal.

> Both A and B have good reason for choosing the salient simply in virtue of the fact that, clearly, doing so is the only (or the easiest and most reliable) way of coordinating.[12]

This argument is used "*contra* Lewis": "we do not have to suppose as a last resort that A thinks that B has a tendency to choose the salient so that A himself has a reason for choosing the salient". In other words, Heal supposes that nothing other than (valid) reasoning directly from the fact of salience need be 'resorted to', to explain why agents choose the salient.

As stated here, this argument does not seem to work. If both agents were to do their parts in any acceptable combination of actions, surely

this would be an equally good 'way of coordinating'? So each agent has an equally good reason for doing his part in any one of the other acceptable combinations of actions. In other words, no special reason is available to support doing one's part in the salient combination.

Heal's suggestion that choice of the salient is the 'most reliable' way of coordinating may suggest that she implicitly assumes that a significant empirical premiss about human behavioural tendencies will be common knowledge to the participants. I have in mind the premiss that human beings tend to do their parts in a salient solution. The rest of Heal's discussion goes against this interpretation, however, for she explicitly claims that there is no need for the agents to appeal to anyone's 'tendency to choose the salient'.

Prior to the passage quoted above, Heal writes at greater length on the issue. Though I do not think her argumentation here is more successful, it has points of contact with that of other authors (in particular Stephen Schiffer) and is worth examination.

Heal suggests that (1) the agents in a colour matching case "know that they can co-ordinate their choices only if they can single out one colour from the rest, by some feature" (ibid).[13] She goes on: (2) "A and B know that by choosing a colour which does stand out for them, and *only* by doing this, can they hope to coordinate" (ibid). She continues: (3) "This provides a reason for each to make the choice of the outstanding colour, which is reinforced by knowledge that the other has that reason" (ibid).

These remarks suggest the following points with which one can easily agree. First, the situation when there is only common knowledge of rationality, the matrix of preferences, and the actions available to them is truly problematic. Assuming that, when rational agents with a colour-matching problem individually lack a reason to choose a particular colour, they will then choose at random, their chances of coordinating their choices is not good: the more colours that are involved, the worse their chances become. Second, given this initial situation, it will be both necessary and sufficient for the successful coordination of action that for each agent one and the same colour is 'singled out', in the sense that, for each agent, there is a reason to choose that colour. For, if each has such a reason, and, rationally, acts on it, their choices will be coordinated in the desired fashion. What have these indisputable points

to do with salience? Something, indeed, but not enough for Heal's argument to go through. If the situation is otherwise unchanged but one acceptable combination of actions is salient, one and the same colour is now, indeed, 'singled out' for each agent, in the sense that it appears as the 'odd man out' — this is so by the definition of salience. Thus, as Heal suggests, granted that it is common knowledge that no further facts that could aid their reasoning is available, it seems that the agents know this: if anything can help them, it must be salience.

At this point, however, we seem to have gone as far as we legitimately can down the path Heal traces. *Can* salience do the trick? Does it, that is, provide each agent with a reason for choosing the salient? If the argument I have given is correct, it does not. Mere salience is *not* 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 *could* provide a reason here. We must conclude, however, that it fails to do so.

Nothing I have said here is intended to deny that salience may allow agents with instincts or psychological propensities of the right type unfailingly to solve otherwise recalcitrant coordination problems. For agents endowed with the appropriate psychological features, salience may presumably serve as a 'clue' sufficient for the achievement of coordination. What I have argued is that salience gets us nowhere within the pared-down framework of game theory. In order to ensure that a rational agent moves in a particular direction we need to give him knowledge that he *has a reason* to move in that direction rather than another. I have argued that salience as defined here does not provide what is needed.

It looks as if hidden equivocations may sometimes lie behind the sense that salience can conquer all. What is in fact necessary for a solution of the problem is that each agent single out one and the same combination of actions *as the one he personally has reason to do his part in.* Any other sort of 'singling out' is not enough. Each must 'notice' that a given combination is *the one he should do his part in.* Each must see that a given combination is (obviously) *the one he should do his part in.* Unfortunately, participation in common knowledge that a certain combination is "uniquely conspicuous" or (perhaps most ambiguously and treacherously) "obvious" does not lead to the

right kind of 'singling out' or 'noticing'. It will not help matters to redefine salience in a reason-providing manner. For salience is supposed to provide a mechanism by reference to which the question 'How might rational agents discover what they ought to do?' may be answered. Clearly it is pointless to answer: 'By finding that some particular combination of actions is *obviously the one they should do their parts in*'. How are they supposed to find *that* out? [14]

A point about the general notion of salience as defined here may now be noted. Presumably the notion is at least sometimes properly applied to a combination of actions that rationality dictates one should do one's part in quite independently of salience considerations. In such cases the intersection of required strategies could be called 'salient', from the point of view of the general concept of salience. But even though this is so, the fact that this combination is (or may be) salient *will not affect anyone's reasoning*, since independently of the salience, reason already dictates what each is to do, and being rational they will do that thing. A reference to the salience of a particular combination in this context, then, is theoretically pointless, albeit conceptually proper. But it could also be misleading. Since the salient combination is also the one each ought to do his part in (by hypothesis), the important conceptual distinction between combinations of these two types in general could become blurred. That it is an important conceptual distinction has been the burden of my discussion of salience here. To use a phrase of John Mackie's, I have argued that salience in the sense at issue here is not *intrinsically action-guiding*. This is one good reason for not talking about salience in the kind of context in question.[15]

Of course, one might point out, as Thomas Schelling does in effect, that if we were all to follow the maxim 'ceteris paribus, do your part in the uniquely salient solution' we would solve more problems than we would otherwise.[16] Pointing such things out can be helpful. But it is the main burden of this paper that this maxim will not necessarily be followed by all rational agents as such.

A similar response is appropriate, I suggest, to an idea of David Gauthier's. Gauthier argues that salience "does enable persons (in a coordination problem) to coordinate their actions" (1975 p. 210). Though this is his overall view, his position is complex. He first argues that, insofar as salience as such does not alter the payoff structure, it

cannot be taken as in itself providing a reason for acting by 'rational maximizers of utility'. Hence 'salience does not in general allow rational persons to generate mutually supporting expectations which converge on the salient equilibrium'. Nonetheless he insists that salience does enable persons to coordinate their actions. He argues that salience enables persons to 'substitute, for their original conception of the situation, a more restricted conception', given which the situation is no longer problematic. In this new conception of the situation, each person has to choose between two options: (1) go for the salient point; (2) randomly choose, with equal probabilities to each, among all the available strategies. Gauthier then argues 'coordination is easily achieved' once the parties so reconceive the situation, which has its own, derived payoff matrix.

In his explicit reasoning Gauthier appeals to the idea that the reconceived situation has 'but one best equilibrium'. And he takes it that when this is so, successful coordination is achieved by acting in accordance with what he calls the 'principle of coordination'.[17] Now I have in fact argued against this principle as a general principle elsewhere.[18] But for present purposes we can waive Gauthier's own general characterization of how reconceiving the situation helps, noting that in at least some cases the reconceived situation allows for a straightforward argument as to what one should do, which makes appeals only to the payoff numbers. (Indeed, each party may now have a dominant strategy.) Let me now ask, then, whether this shows that at least these cases are solvable for all rational agents in the way Gauthier suggests.

The logical structure of Gauthier's suggestion seems to be as follows. In certain circumstances, a rational agent can reason starting thus: 'If each of us reconceives the situation in such-and-such a way (and this is open to all), then it will be obvious to each of us that he should do his part in the salient combination of actions, and then we will coordinate (more or less) as desired.' From this it is presumably supposed to follow that each *should* reconceive the situation in the relevant way. Given that both do so, and this is obvious to all, coordination at the salient point will be achieved. But this form of argument is open to the very type of objection that, I have argued, can be levelled at a direct or ground-level appeal to salience in the first place. It begins with a *conditional* about what will happen (a good outcome) if each player

does a certain thing. But you only have reason to play your own part here if you have reason to think the other player will. But what reason can this be? He is in the same position that you are in.

Once again, then, we can agree that in certain cases it would be *good* if *all* rational agents as such would do a certain thing (in this case reconceive their situation in a certain way). All rational agents can recognize this, without that generating any confidence in a given rational agent that his rational opposite number will act accordingly on this or any other occasion. (Note what might be referred to as the *influence-independence* of the participants. The fact that a good outcome would be reached if *both* did something cannot by itself be a reason for either one individually why he should do it. For his doing it cannot itself ensure that the other does it.)

Someone might defend Gauthier as presenting an account of how salience *could* work (if everyone decided to abide by a certain policy) or an implicit psychological account as to how it is salience often seems to be *used.* (People may tend to reconceive the situation as suggested.) His own discussion is perhaps slightly ambiguous on this score. But he does imply that 'successful coordination' which 'depends on conceiving a situation' in a certain way, is a 'matter . . . of reason'. In any case, in the context of a discussion of game-theoretic rationality as characterized here, I think we must conclude that his idea does not show that rational agents as such will derive a reason for acting directly from the fact that one solution is salient.

A final point. The psychological nature of salience, and so perhaps the non-reason-providing nature of salience, might be queried for some cases. For instance, in a colour-matching case in which there are more than two colours it may seem that if one colour has been mentioned over a loudspeaker just before the choice had to be made, choosing that colour now would be the salient choice *for all rational agents.* (Contrast here the case where each party only has two choices. Then whereas one colour can be singled out as being the colour which has just been mentioned, the other can be singled out as being the colour which has *not* just been mentioned.) The psychological nature of salience may seem even more open to question for a type of case noted by Schelling, in which the matrix itself contains an observable asymmetry. It may be, for instance, that one of the more preferred outcomes is the worst,

several tie for best.[19] Possibly in such a case the worst of these points will, *for all rational agents*, "stand out from the rest by its uniqueness in some conspicuous respect" (Lewis).

Even if this point is well taken — it may be — it leaves the crucial question whether in these cases there is a reason for doing one's part in the salient combination of actions. Given that rationality does not in general dictate that one do one's part in a solution by virtue of its salience, it is hard to see how it can do so here.

I conclude that rational agents as such cannot be expected to do their parts in a salient solution in a coordination problem. This point is worth stressing, and not just because we could run into rational beings who are not equipped with human psychologies. For one thing, it is clear from the literature in this area that it is tempting to deny it. (I myself was quite surprised when I first saw it was true.) Meanwhile, it is important to understand the scope and limits of rationality. This is so for more than one reason. We may prefer that our lives be structured as far as possible by the exercise of reason. Then we need to understand which are the situations in which reason can indeed move us. We may also want to know how we actually work. As my opening quotation suggests, it seems that human agents tend to do their parts in a salient combination of action in an otherwise intractable coordination problem. According to the argument of this paper, this tendency cannot be explained as an expression of each agent's knowledge that, if a combination obviously stands out for all, then *ipso facto* each one has reason to do his part in it.

III

What, then, might explain it? There are various possibilities. In this concluding section I focus on one general idea of what goes on — an idea contemplated by David Lewis — and two possible developments from it. In my kidnapping case, perhaps we would choose red under those conditions because we are made that way. That is, because each of us feels pulled towards a combination of actions that obviously stands out for all.[20] Evidently it is useful to be made that way, if all one's fellows are. So there could be some evolutionary explanation of how this tendency developed.

There is more than one way things could develop given the seed of an instinctive pull, on the part of each, to choose the salient. Suppose that each of us feels an instinctive pull or tendency to choose red. Given an understanding that others are made the same way, this will not be a pull there is any obvious reason to resist. Indeed, there is every reason to go along with it. Hence, from the seed of common instinct, a mutually acceptable pattern of behavior could emerge.

Lewis points out that it would be enough for each agent in a coordination problem simply to ascribe to the others a tendency to choose the salient 'as a last resort, when they have no stronger ground for choice', and hence choose the salient himself.[21] In this case it is not necessary that anyone in fact has a tendency to choose the salient as a last resort. All that is necessary is that each one believes the others have such a tendency, and that in this case each of the others is so placed he will act on this tendency. Each will then have sufficient reason to do his part in the salient combination of actions. In the event, each could be acting purely on the basis of reasons in choosing the salient.

There is clearly something disturbing about such a state of affairs. For one thing, everyone is wrong about everybody else. (Everyone in fact acts on the basis of reasons, rather than on the basis of his unreasoned tendency, but everyone thinks that everyone else will act on his unreasoned tendency.) Again, everyone regards his own situation as unique; he is acting on the basis of reasons, the others are not.

I do not say that some people may not sometimes operate in such terms. (I can report that one participant in an informal experiment I conducted explained his choice of the salient combination in terms of *everyone else's* 'suggestibility'.) However, in the model I am now considering the main force operating in each case is not the agent's reasoning (about others or anything else) but his own unreasoned impulse. Reasonably ascribing a similar impulse to others (either through observation of their actions or simply through knowledge that they are the same kind of creature) he sees no reason to struggle against this tendency. As Lewis in effect points out, if he does ascribe a similar impulse to others, this will provide a strong prima facie reason to do his part, irrespective of his own impulse. However, it seems plausible to suggest that generally the process starts with the inclination of the agent in the face of salience, and proceeds with reason allowing inclination to take its course.

Another way things could develop from the seed of impulse is suggested by the ease with which many have felt salience considerations could be fitted within the framework of game theory. Not only do we in fact tend to choose the salient, but apparently the combination of actions that obviously stands out for all may come to seem *the thing one should (obviously) do one's part in.* Given the natural impulse model under consideration, a Humean (and also Wittgensteinian) idea comes to mind. Recall how Hume famously talks of the human mind's tendency to expect that a B will follow an A, once Bs have followed As in the past, and suggests that it is this tendency in us that produces our (dubious) belief in some kind of necessary connection between As and Bs: we project the tendency of our minds on to objects 'out there' (and then attempt to justify our tendency by appeal to the (perceived) nature of things). Recall also how, according to Wittgenstein, in 'obeying a rule' for the use of a word we actually act 'blindly', 'without reasons'.[22] Meanwhile there is an evident temptation to suppose (wrongly) that we are relying on some knowledge that *justifies* our going on as we do. In other words, we are tempted to suppose that we do what we do *for a reason.*

We may or may not accept the 'debunking' arguments of Hume and Wittgenstein on the matters that concerned them. In particular, we may not be convinced that the proposed justification of the practice in question is indeed unavailable.[23] Meanwhile, in the salience case, if we insist that the salient combination is what *should* be chosen, it seems that we may well be involved in the projection of an unreasoned compulsion on to reason. If this makes it easier for each of us to choose the salient, then perhaps it is just as well, as a matter of practice, if not of theory.[24]

NOTES

[1] Here I echo Wittgenstein, 1953, Part 1, section 219. I say some more about the relation of my concerns in this paper to Wittgenstein's in the final section.

[2] In all such contexts, 'his', is to be read as 'his or her', and so on.

[3] For notice of some problems with David Lewis's game-theoretic definition of 'co-ordination problem' see Gilbert (1981).

[4] Doing so is somewhat complicated by the fact that more than one technical definition of 'common knowledge' has been given. The commonest definition in philosophy and economics is as follows: For any two agents A and B, it is common knowledge between A and B that p if and only if A knows that p, B knows that p, A knows that B knows

that p, B knows that A knows that p, and so on. See David Lewis (1969) and (1975), Stephen Schiffer (1972) and Robert Aumann (1976). I discuss the issue of definition further in *On Social Facts*, forthcoming. See also Heal (1978). The details of this issue need not concern us here.

[5] In the passage quoted above Lewis (1969) writes of "standing out from the rest by its uniqueness in some conspicuous respect"; he also uses a characterization which seems considerably weaker, in terms of "being unique in some way everyone will notice, expect the others to notice, and so on". It is not clear that several different combinations of actions could not be salient in this sense. The definition of salience in this context should, evidently, be such that there can only be a single salient combination of actions in a given occurrence of a coordination problem.

[6] See Lewis (1969), p. 35. On the presence of nonrational factors in Lewis's model of the origin of conventions, see my (1983ms) and *On Social Facts*, forthcoming. See section III below, also.

[7] See for instance Ullmann-Margalit (1977) (p. 112); Schiffer (1972) pp. 145—147. See also G. Postema (1982), p. 174.

[8] This example resembles the case on which Heal focusses in Heal, 1978. In her case salience is a result of a previously successful precedent. I deliberately eschew examples in which precedent gives rise to salience here, since precedent involves special features and we are concerned with salience as such and in general. I discuss the case of precedent separately elsewhere (1983ms; *On Social Facts*, forthcoming).

[9] The following way of arguing regarding salience is an application of a type of argument I have previously used in discussion of successful precedent as such, and other facts common knowledge of which some have deemed to give rational agents a reason for action. See Gilbert (1981), and, with the focus on precedent (1983ms). See also *On Social Facts*, forthcoming, Chapter Six, and my discussion of Gauthier, below.

[10] Sally might simply assign a high prior probability to Joe doing his part, but, if one does not engage in circular reasoning here, she might just as well assign a high prior to one of Joe's other actions. Actually, it seems to me that any assumption of the assignment of such arbitrary priors to particular actions rather than of the deduction of probabilities from the assumed rationality of the other agents is dubious in a discussion of purely rational agents with common knowledge of rationality. This position conforms with that of classical game theory. (See Robert Aumann (1987) p. 1236, 'Headnote'.) I shall not attempt to debate this issue here.

[11] I am indebted to an anonymous referee whose comments suggested this line of argument. Thanks also to Simon Blackburn for an example that suggested to me the next case in the text.

[12] Heal, 1978, p. 129, footnote.

[13] This is very similar to Schiffer, 1972, p. 146. See also Postema, loc. cit.

[14] I do not deny that there could be a use in a different context for a conception of the salient as that which one is 'obviously' called upon do do. Cf. John McDowell (1979) p. 335. I thank Joel Kupperman for the reference.

[15] See for instance, in David Gauthier (1975) pp. 207—209, Gauthier's first example of a way in which one of the best equilibria in a situation may be salient. (It is not clear that we need to bring consideration of expected utility and expected indifference into play here, as Gauthier does. We can appeal to the rationality of maximizing one's security level, ceteris paribus. Cf. Schelling (1960) p. 297.) See also Arthur Kuflik (1982).

[16] 'Ceteris paribus' is to rule out any cases where there is somehow a conflict between the salient solution and what rationality dictates given the payoff matrix. Cf. Gauthier (1975) p. 209, footnote 11.

[17] Gauthier presents this principle on p. 201.

[18] Gilbert (1983ms).

[19] See Schelling (1960), p. 295—6.

[20] Cf. Lewis (1969) p. 35.
[21] See Lewis (1969) pp. 35—6.
[22] The term 'blindly' is used at op. cit., section 219, and 'without reasons' at section 211.
[23] There is a sceptical discussion of Wittgenstein in my *On Social Facts*, forthcoming, Chapter Three.
[24] A version of this material was presented to the Philosophy Department at the University of Connecticut, Storrs, October 9th 1985. I am grateful for the comments I received then. Thanks also to Simon Blackburn, Peter Hammond, Saul Kripke and John Troyer for comments on written material. The current version was completed while I was a Visiting Fellow in the Philosophy Department at Princeton University. I thank the Department for its hospitality.

REFERENCES

Aumann, Robert: 1976, 'Agreeing to Disagree', Annals of Statistics 4, pp. 1236—1239.
Aumann, Robert: 1987, 'Correlated Equilibrium as an Expression of Bayesian Rationality', Econometrica 55, pp. 1—18.
Gauthier, David: 1975, 'Coordination', Dialogue 14, pp. 195—221.
Gilbert, Margaret: 1981, 'Game Theory and Convention', Synthese 46, pp. 44—93.
Gilbert, Margaret: 1983abs., 'Some Limitations of Rationality', abstract, Journal of Philosophy 53, p. 615.
Gilbert, Margaret: 1983ms, 'Some Limitations of Rationality', paper presented at the meetings of the American Philosophical Association and elsewhere.
Gilbert, Margaret: forthcoming, On Social Facts, to be published by Routledge, London and New York.
Heal, Jane: 1978, 'Common Knowledge', Philosophical Quarterly 28, pp. 116—131.
Kuflik, Arthur: 1982, 'Coordination, Equilibrium, and Rational Choice', Philosophical Studies, 42, pp. 333—348.
Lewis, David: 1969, Convention: A Philosophic Study (Harvard University Press).
Lewis, David: 1975, 'Languages and Language', Minnesota Studies in the Philosophy of Science Volume 7.
McDowell, John: 1979, 'Virtue and Reason', Monist 62, pp. 331—350.
Postema, Gerald J.: 1982, 'Coordination and Convention at the Foundations of Law', Journal of Legal Studies XI, pp. 165—203.
Schelling, Thomas: 1960, The Strategy of Conflict, Harvard University Press.
Schiffer, Steven: 1972, Meaning, Oxford University Press.
Ullman-Margalit, Edna: 1977, The Emergence of Norms, Oxford University Press.
Wittgenstein, Ludwig: 1953, Philosophical Investigations, The Macmillan Company.

Philosophy Department,
U-54,
University of Connecticut,
Storrs, CT 06268,
U.S.A.