

“TEAM-THINKING” IN THE WORK OF MARTIN HOLLIS.

by Nicholas Bardsley
School of Social and Economic Studies
University of East Anglia, Norwich (UK).
N.Bardsley@uea.ac.uk

Abstract

A key interest of Martin Hollis was the concept of rationality in action, and a key tool for its analysis was the theory of games. He advocated that puzzles it generates demonstrate the limitations of a purely instrumental account of rationality, and indicate lines along which a fuller account could be drawn. A late development in his thinking is the endorsement of a conception of collective rationality, seemingly at odds with defensible forms of methodological individualism. I argue that problems for “team-thinking” lie elsewhere; it resolves one puzzle about rationality at the expense of posing another, concerning agents’ reasons for membership of collectives.

Introduction

Game theory produces a number of puzzles in which ideally rational agents are required to think and act in ways which are objectionable to their imperfectly rational human counterparts. For Martin Hollis, philosopher of the social sciences, the resolution of these puzzles required, in Hume’s phrase, “a remedy in the judgement and understanding for what is incommodious in the affections”. The puzzles are insoluble for rational beings if reason amounts to prudence, so the question is how else to conceive rationality so as to enable a solution. Famous examples discussed by Hollis are the prisoner’s dilemma and the free-rider problem, where the pursuit of individual interests is collectively self-defeating. Reflection on such cases, and on tensions between economic and sociological accounts of action, prompted his exploration of the notion of “expressive” rationality.¹ Hollis eventually came to the view, however, that there is a more scandalous puzzle in which the ideal agents have no way of deciding what to do, despite the complete congruence of interests between them and the apparent obviousness of a correct course of action. The preferred remedy in this case is *collective* rationality. The problem and the solution are the subject matter of chapter 7 of

¹ Hollis (1987).

“Trust Within Reason”.² I summarise the basic case for “team-thinking” in Hollis’ last work, defend it from a methodological individualist objection, but show how it generates a problem for a rationalistic understanding of the relationship between individuals and collectives.

The Need for Collective Rationality: a Trivial Coordination Problem

The new notion, team agency, is introduced via two person coordination games.³ In a *pure* coordination game there are multiple equilibria, each of which gives both players the best payoffs on offer (an equilibrium in this context is a set of behaviours that are best replies to each other). A natural example is two people deciding which side of a road to drive on; it matters not at all whether they drive on the left or the right, so long as they both do the same, avoiding collision. It seems plausible, in a context of repeated play between many agents, that once people come to expect one pattern of play rather than the other, these expectations will amount to self-fulfilling and self-perpetuating prophecies.

Lewis (1969) offers an analysis of conventions along these lines. However, his analysis is dependent on an extraneous element of irrational or non-rational behaviour. This is because such expectations would always be ungrounded for the agents of game theory, who can always ask why their opponent will not deviate from an established convention, given that were a deviation expected of them, deviation would be the rational strategy. Within the “game theoretic tool box”, an expectation cannot be grounded in anything other than the payoff matrix of the game and common knowledge of players’ rationality. If one concentrates solely on these, game theory can *only ever* advise “keep left if the other will keep left, right if the other will keep right”; because of the symmetry of the game the players cannot advance by putting themselves in the other players’ shoes and asking what their opponent will expect them to do.

² Hollis (1998).

³ Formally, a coordination game is one involving equal payoffs for both players along the leading diagonal of a payoff matrix, with zeros elsewhere.

Hollis presents us with a more surprising example, an *impure* coordination game, that is one with a *pareto dominant* equilibrium, (a pareto dominant equilibrium is one offering better payoffs for both players than the other equilibrium). Suppose for example, that two people have, independently, to choose either “Ten pounds” or “Fifty pounds”. If they choose the same figure they receive the sum chosen, if not they each receive nothing. It seems obvious that each player ought to name the larger sum. Again, however, there is an insoluble problem of how to generate unconditional advice to play one’s part in the better solution, which emerges for exactly the same reasons as in the pure coordination game. If A expects B to choose “Ten pounds” then A ought also to choose this, and the contrary expectation cannot be grounded by any argument that does not also ground its opposite. The impure coordination game sets a tough puzzle. One can increase the payoffs in the better equilibrium as much as one likes without the strategy involved in the dominant equilibrium becoming the game theoretic rational choice; the same problem is produced when the choice is between “One pound” and “One million pounds”. Neither habit nor custom can generate the necessary unconditional advice, because the choice between “The customary (habitual) action” and *is the same coordination problem with different labels attached to the actions.*

“Trust in Miniature: Teamwork”

The solution Hollis offered to this problem proposes that game theoretic agents make no progress because they ask themselves the wrong question. The agents should ask themselves not “What should *I* do?” but rather “What should *we* do?” The answer to the latter question for the impure coordination problem is “We should choose the larger sum.” Team agency gets round conditionality because it involves taking the existence of a plan which it is best for the team to follow as a sufficient reason for action. Apparently this is a form of invalid practical reasoning, since, for example, it does not follow from the fact that A ought to let go of the trapeze and B ought to catch her that A ought to let go. The further premise that is to licence such inferences is that there really are group agents, or teams. That “The good of the team

requires A to do *a* and B to do *b*” is a sufficient consideration for each to play their part in the plan, given that A and B *are* (part of) a team.

Team-thinking is in one respect analogous to Kantian practical reasoning since it issues prescriptions which are not conditional on others’ performance. It requires that team membership be a peculiar kind of relationship, binding those involved into a single acting unit. It goes beyond action that is in the interests of the team, for this can be conceived individualistically. In the impure coordination game, that is, agents could ask themselves “What should I do to further the interests of both of us?”, but the answer in that case will be “Choose whichever sum you expect the other to choose”, that is, a conditional answer. The contrast between action in the interest of the team and action as a member of it is akin to Rousseau’s distinction between the *will of all* and the *general will*.⁴

If what is rational for an individual to do depends on their beliefs and goals, what is rational for a group to do is analogously dependent on its beliefs and goals. In team agency, then, each member has, and acts on, *group* beliefs and desires. It might be thought that such a notion commits one to a strong concept of a group mind. However, it is notable that philosophers offering analyses of group agency amenable to team-thinking typically claim to have avoided any such notion; what is postulated is merely a particular way individuals have of thinking and acting.⁵ These accounts can be read as attempts to flesh out the notion of doing one’s part in an action *as a part*. Group agency, it seems, can be understood in terms of particular features of individuals’ mental states.

If admissible, the team-thinking concept has relevance for a wider class of games than coordination problems. Indeed, without a notion of collective rationality on which individuals

⁴ Such a contrast also provides one reason for distinguishing sharply between act and rule utilitarianism.

⁵ See in particular Gilbert (1989) and Searle (1990).

are capable of acting, it is perhaps hard to see why certain topics in game theory, notably the one-shot prisoner's dilemma, have generated so much interest. Certainly, many game theorists would hold that, in the latter, since each individual has a dominant strategy of defection, there is almost nothing of interest to be said.⁶ If we wish to understand why the prisoner's dilemma *has* been seen as a real dilemma, though, without an uncharitable dismissal of this as mere confusion, the notion of collective rationality is extremely helpful. For then both actions can be rationalised to the agent, one as individually, and the other as collectively rational, and between them there is a genuine conflict. In "Trust Within Reason", the immediate motive for the introduction of team-thinking is yet another game theoretic puzzle, the "centipede game", but one chosen precisely because there arises a conflict between individually and collectively rational action. It exemplifies a problem of *trust* since a mutually beneficial outcome is available, but only if each individual somehow chooses to act against the dictates of instrumental rationality. The solution is not to be that trust is grounded in irrational or non-rational behaviours, but in collective reason and team agency.

Critical Reflection

There is a problem, though, for *team-thinking within reason*. This arises when we ask how it is that teams are formed or maintained. Hollis wants his individuals to retain a critical distance from the teams they belong to - otherwise, they could acquire "a wholly local identity and source of reasons for action" (Hollis (1998) p128), and so become unable to deal reasonably with outsiders or refrain from group excesses. This concern reflects another theme in Hollis' work, namely the need for "social relations that reason can endorse". But if there are reasons for belonging to teams, *to whom are these reasons addressed?* Can a sense be given to the seemingly bizarre question "Should I think from the point of view of "I" or "We" in this game?" This question is addressed to an "I", illustrating the point that a reason apparently *presupposes* a unit of agency. If some third point of view existed, neutral between I and We to resolve this problem, it is unclear how the agent could have reason to act

⁶ See Binmore (1994 p310) for example.

according to its dictates, since for all we are told the goals and beliefs attributed to the agent have been exhausted in the two points of view already posited.⁷

So long as one is interested only in the admissibility of the team-thinking concept, this question can perhaps be left aside; it is enough that teams exist, a relief that collective rationality rationalises sensible play in the impure coordination game, and a further question how individuals mediate between personal and team agency. If one is at all concerned to explain actual trust and cooperation though, including the empirics of behaviour in games, this will not do. It seems that people can coordinate successfully, and are more cooperative than game theory predicts, even when there is no history of interpersonal contact between them of the sort that is required for the individuals to constitute a pre-existing team. If they cooperate through team agency in these cases, we need to know how the team is formed. Social psychologists are inclined to view such questions as causal matters.⁸ However, Hollis would not have liked to let matters rest there, since he mistrusted explanations of behaviour which substitute causes for reasons.

Conclusion

In the impure coordination game, it is clear that agents can coordinate on the better equilibrium, and in so doing make the best choices despite the fact that game theory offers them no advice. Hollis' aim in "Trust Within Reason" was to offer an account of rational action which does explain why best choices are best. Consideration of impure coordination naturally led him to the conclusion that to meet this demand it is necessary to introduce a concept of collective rationality, according to which agents behave *as if* moved by a common mind. Whilst this need not imply commitment to a literal concept of a group mind it does make a rationalistic understanding of the relationship between individuals and groups appear highly problematic.

⁷ First person plural and singular would also appear to be the only possible relevant points of view for deciding on one's own action.

⁸ See, for example, Brewer (1989).

References

- Binmore, K.G. (1992): *Fun and Games*, Lexington, Mass.; D.C. Heath.
- Brewer, M.B. (1989): “Ambivalent Sociality: The Human Condition”, *Behavioural and Brain Sciences*, 12, 4, p699.
- Cohen, P. R., Morgan, J. and Pollack, M. E. eds.(1990): *Intentions in Communication*. Cambridge: MIT Press.
- Davis, D. D. and Holt, C. A. (1992): *Experimental Economics*. Princeton, N.J.: Princeton University Press.
- Gilbert, M. (1989): *On Social Facts*. London: Routledge.
- Hollis, M. (1987): *The Cunning of Reason*. Cambridge: Cambridge University Press.
- (1998): *Trust Within Reason*. Cambridge: Cambridge University Press.
- Hurley, S. (1989): *Natural Reasons*, Oxford University Press.
- Lewis, D. K. (1969): *Convention: A Philosophical Study*. Cambridge, Mass.: Harvard University Press.
- Schelling, T. C. (1963): *The Strategy of Conflict*. London: Oxford University Press.
- Searle, J. R. (1990): “Shared Cooperative Activity” in Cohen et al (1990).