A. Blinov[1]

RATIONALITIES IN CONFLICT: COMPENSATORY LOGICO-COGNITIVE IRRATIONALITY IN INTERACTIVE CONTEXTS

The aim of the article is to make a couple of steps toward a theory of conflicts between two different varieties of rationality, - namely: (i) logico-cognitive, or *epistemic* rationality[2] and (ii) rationality, all things considered, or *aggregative* rationality.

The main interest of such a theory, as I perceive it, may be that, when applied to interactive contexts, it provides a basis for plausible explanations of some kinds of empirically observed irrationalities in human thought and behaviour.

1) I will begin with a brief exposition of a theory of the two rationalities. (2) Then I will construct and discuss a paradigmatic interactive situation in which for all the participants it is rational, all things considered, to jointly indulge in an epistemic irrationality, because such a choice restores Pareto-efficiency of the initially Pareto-inefficient situation. (3) Finally, I will discuss the significance and scope of possible applications of the paradigmatic model. In particular, I will argue that the paradigmatic model can provide explanations for the persistence of at least some sorts of ideologies.

1. Epistemic vs aggregative rationality

I borrow from Richard Foley his characterisation of the distinction between epistemic and aggregative rationality[3]. The characterisation is this:

All judgments of rationality are judgments about how effectively an individual is pursuing some goal. However, such judgments are commonly elliptical. For one thing, they commonly fail to make

[1] University of New England, NSW, Australia.

[2] For an enlightening discussion of reasons for assimilating a person's logical principles with his or her cognitive [= epistemic] rationality, see, e.g., Korner (1984), pp.42-62. One pivotal point of the discussion is this: "The principles which determine a person's conception of logical consistency, i.e., the principles of his logic, and the cognitively supreme principles which determine his categorial framework, including his logic, are his standards of cognitive rationality." (p.44)

[3] Foley (1987)

explicit what goals are in question.[4] To avoid confusion one should strive, when pronouncing a judgment of rationality, explicitly to relativise it to a goal. When the goal in question is epistemic, then the judgment is one of *epistemic* rationality.

What goals are epistemic? Foley takes it that there is only one *purely* epistemic goal, namely, that of *now believing true beliefs and now not believing false beliefs*. I think that I am not prepared to agree with the 'only one' part of Foley's claim, but it does not matter for my purposes here. What matters is that we all seem to have more or less clear intuitions about which goals are epistemic and which not. For example, two further goals are unmistakably epistemic, though not necessarily purely so: (ii) *acquiring as much knowledge as possible*; (iii) *developing one's reasoning ability (or more generally: cognitive abilities at large) as high as possible*.

On the other hand, one can have more than one goal simultaneously. Then judgments about how effectively she is pursuing the whole constellation of her several (weighted) goals are judgments of *rationality, all things considered*, or, to have a regular adjective, *aggregative rationality*.

It should be clear that the two notions of rationality - epistemic and aggregative - are distinct. More than that, it is *prima facie* possible that on some occasions the two clash with one another: say, a belief which is aggregatively rational for an individual on a specific occasion to maintain may not be epistemically rational for him, on the same occasion, to maintain.

2. The game of the Good Jailer: An epistemic dilemma for the prisoners

It is relatively safe to ignore the distinction between the two rationalities when treating one-agent contexts. Actually, there is a decision-theoretic result, namely, Savage-Good theorem that guarantees impossibility of a conflict between aggregative rationality and one variety of epistemic rationality under some well-specified conditions: In a situation where a single utility-maximiser is to take a decision, new information can never be harmful for her, given that the information is correct and costless.

Admittedly, even remaining within the domain of one-agent contexts, one can think of a situation like that of Pascal's Wager where it is rational, all things considered, for the individual to come to maintain a belief which is epistemically irrational for her to maintain. But

[4] One further thing, by Foley's lights, is the perspective of the judgment, but for my purposes here we can forget about it, - at least at the first stages.

what makes this possible is the unusual assumption that there exists a being who (i) is endowed with the supernatural ability of having immediate access to the agent's mind, and (ii) who can reward or punish the agent for having this or that belief.

The picture changes dramatically when we move from one-agent to many-agent [= interactive] situations, that is, to the domain of Game Theory. The fact that the value of knowledge can be negative in an interactive context is well-documented in the literature.[5]

Let me come up with a situation that is quite paradigmatic in this respect, but which, to my knowledge, was never discussed in the literature. The situation is a variation on the famous Prisoners' Dilemma. The Prisoners' Dilemma is this: On suspicion of having jointly committed a crime, two persons, say Ann and Peter, have been detained and put into separate cells so that they are unable to communicate. Common knowledge for both is at least this: If one confesses while the other does not, he who has confessed will be immediately set free for helping the investigator. The other will be put away for ten years. If both confess, both will be put away for nine years. If both keep silence, both will be locked up for a year for a misdemeanour, since there is not enough evidence to support the more serious suspicion.

The names of the two strategies on Table 1 are abbreviations for 'cooperate' and 'defect', respectively. As is well known, the unique Nash equilibrium for this game is that both players should defect. This implies that it is rational for each player to defect, which is also supported by the fact that, for each player, D strictly dominates C. So if they are rational, they will both defect and spend in jail nine years each.

		Peter	
		C	D
Ann	C	-1, -1	-10, 0
	D	0, -10	-9, -9

Table 1.

Such is the standard Prisoners' Dilemma. My variation is this: Suppose that Ann and Peter are members of a gang which is governed

[5] See, among others, Hirshleifer (1971), Kamien a.o. (1990: 1), Kamien a.o. (1990: 2), Neyman (1991), Bassan and Scarsini (1995), Gossner (1997), Korilis a.o. (1999).

in a democratic fashion. In particular, a couple of days after Ann and Peter's arrest there took place a general meeting of the gang. The only item of the agenda was a proposal to consider collaboration of a jailed member of the gang with the investigator as a capital offence which is to be punished by death. If the proposal has been adopted by the meeting, and this has become common knowledge between the two players, then of course this knowledge should result in a drastic change of their strategic situation. Suppose, for the sake of smooth calculation, that each of the two players assesses the negative utility of their own death as equal to 50 years in jail. Then the new situation is represented by Table 2:

		Peter	
		C	D
Ann	C	-1, -1	-10, -50
	D	-50, -10	-59, -59

Table 2.

Now both the logic of Nash equilibrium and that of strict dominance recommend that each should cooperate. So if they are rational they will both cooperate (that is, keep silence) and spend in jail one year each.

Unfortunately, the players do not know the poll's result, but being old-standing members of the gang as they are, they know the mentality of their fellow gangsters, so that they share the belief that is represented by subjective probability of .5 that the meeting has adopted the proposal and subjective probability of .5 that the proposal was not adopted. The fact that they share this belief is common knowledge between them. As can be easily calculated, this still leaves them, *qua* maximisers of expected utility, with recommendation that each should cooperate.

Let us dub the resulting game ‘The PD/CP under the Veil of Ignorance’, where ‘PD’ stand for ‘Prisoners’ Dilemma’ and ‘CD’ for ‘Capital Punishment’. Of course, the PD/CP under the Veil of Ignorance is a typical game with incomplete information in Harsanyi’s sense.

So far so good. But this is not the end of the story. Suppose now that the two prisoners are offered one more option. It happens that one of their jailers, out of sheer sympathy with the two hapless creatures, comes up with a suggestion. He can inquire and report to them about

the meeting's result. To handle the issue with perfect equity, though, he will report either to both - if each opts to learn, or to neither - if at least one of the two opts to remain ignorant. The offer and the fact that both detainees completely trust the jailer's information are common knowledge between them.

Call the resulting game ‘The Kind Jailer’. To grasp its formal structure, consider first the game ‘PD/CP without the Veil of Ignorance’ which is exactly like ‘PD/CP with the Veil of Ignorance’ except that the veil of ignorance [= the two-member information set] is removed: whichever option (that is, the PD or the CP) the Nature chooses, the two prisoners will learn the choice. Now, the Kind Jailer is the game in which the two prisoners start with having a choice between playing the PD/CP with, or without, the Veil of Ignorance. Their initial (simultaneous) move is voicing their preferences between the two options. After that move, they proceed to play the PD/CP without the Veil of Ignorance iff both preferred to do so at the initial move. Otherwise, they play the PD/CP with the Veil of Ignorance.

Now, as to the Kind Jailer, the main question is 'What is it rational for each prisoner to do: accept the jailer's offer or refuse it?' The question is elliptical, which cannot be tolerated given the situation at issue. We should make the goal explicit, and this will result in at least two different complete (non-elliptical) questions:

Q1 What is it rational for each prisoner - say, for Ann, - to do relative to the epistemic goal of acquiring as much knowledge as possible?

Q2 What is it rational for each prisoner - say, for Ann, - to do relative to the whole constellation of her goals, that is, what is it rational for her to do, all things considered?

If each values knowledge positively, but sufficiently lower than freedom and/or life, and this is common knowledge between the two, then the answers to the two questions differ, the answer to Q1 being 'Accept the offer', and the answer to Q2, 'Refuse it'. This is so because under ignorance, aggregative rationality recommends each to cooperate, which results in one year in jail for each. On the other hand, given their subjective probabilities, the offer brings the 50-out-of-100 risk that they will come to common knowledge that the meeting failed to introduce capital punishment, and then it will be aggregatively rational for each to defect, which will keep each in jail for nine years. The risk being too high, aggregative rationality recommends each to remain

ignorant.[6] Thus, shared ignorance, and even shared epistemic irrationality, can easily be a boon rather than a bane, all things considered, in an interactive situation.

3. The game of the Good Jailer: Possible generalisations and applications

The paradigmatic situation of the Good Jailer derives its significance from the fact that it seems to be generalisable along no fewer dimensions than the original Prisoners' Dilemma. Let me cite some crucial dimensions:

(1) It generalises to other epistemic goals, that is, to other *varieties of epistemic rationality*. For example, there is a result[7] to the effect that if, in a finitely repeated Prisoners' Dilemma, there are bounds (possibly very large) to the complexity of the strategies that the players may use, then there is a Nash equilibrium that yields a payoff close to the cooperative one. Now, if we try and put a real-life interpretation on this mathematical result, then one realistic reason why the players' strategies should be of limited complexity may be that the players have the epistemic imperfection of being of low intelligence: they are just not intelligent enough to think of and implement very complex strategies. Under this interpretation of the result at issue, simple-mindedness is on a par with incompleteness of knowledge in the sense that it is an epistemic imperfection that, when shared by all the participants, can be beneficial in Pareto-inefficient interactive contexts. Consequently, it may occur, under suitable interactive circumstances, that it is aggregatively rational for all the participants to jointly indulge in a corresponding variety of epistemic irrationality, e.g., that of refraining from developing one's intelligence as high as possible.

(2) Secondly, exactly in the same way in which its core component, the Prisoners' Dilemma, does, the Good Jailer generalises to the situations with *more than two players*, which makes it relevant for the whole range of problems of collective action.

(3) Thirdly, the *precise pattern of the situation* can vary, the only invariant required being *Pareto-inefficiency* of the core situation. It is a straightforward observation that for every Pareto-inefficient interactive situation, there exists a way of augmenting it with a stage of epistemic preplay such that the augmented situation is Pareto-efficient,

[6] Technically, it means that opting for the PD/CP with the Veil of Ignorance at the initial move is part of the unique Nash equilibrium of the Kind Jailer. This unique Nash equilibrium is the strategy profile in which each player's complete strategy is 'First, reject the jailer's offer; then, cooperate under the veil of ignorance'.

[7] See Neyman (1985).

but the cost of restoration of Pareto-efficiency is that part of the Pareto-efficient Nash-equilibrium path of the augmented game is for all the participants to jointly commit an epistemic irrationality of some sort or other[8].

As to possible applications, my contention is that, given all possible generalisations of the Good Jailer, the formal models of its kind can provide a clue for explaining some important sorts of empirically observable irrationalities in human thought and reasoning – in the same way in which the formal model of the Prisoners' Dilemma provides a clue for explaining some important kinds of real-life strategic situations.

Given the limitations of this paper, I will cite just one, but important, area of possible application. There is a problem in current economic theorising which is highly relevant both to cognitive science and to the theory of rationalities in conflict I am discussing here. The problem is that, more often than not, real-life markets are imperfect in that sense of perfection that has been ascribed to them by neoclassical theory. One implication is that the agents' beliefs begin to matter, whereas they were irrelevant under the assumption of perfection. Now, the question is 'Why is it that very often belief systems that determine the choices of real-life market agents happen to be less than rational epistemically, being myths, taboos, prejudices and other such theories that can be grouped under the umbrella term of *ideologies*?'[9]

I think that an interesting answer can be found along the lines of the theory of rationalities in conflict, the rough outline of the answer being this: Imperfect markets fail to guarantee Pareto-efficiency. But as we have seen, if an interactive situation is Pareto-inefficient, then a jointly committed epistemic irrationality of the right sort can be a remedy. Ideologies may happen to be exactly such sort of epistemically irrational belief systems that, when maintained by all or most of the participants, compensate for the Pareto-inefficiency of the initial market situation.

In other words, some sorts of empirically observed epistemic irrationality may happen to have an impeccable rationale: they render services to aggregative rationality. And there seems to be no reason why this format of explaining irrationality could not be transferred from imperfect markets to further areas of collective action and even to coordination problems with several equilibria.

[8] For some formal aspects of the issue, see Blinov (2001).

[9] For an enlightening discussion of this question, see North (1998), pp.713-721. North seeks an answer along different lines than mine, though.

REFERENCES

1. Bassan, B, and Scarsini, M., 'On the value of information in multi-agent decision theory', *Journal of Mathematical Economics* 24, 557-576, 1995.
2. Blinov, A. 'Games with common belief on payoff function', *Logical Investigations*, Moscow: Nauka, 2001, pp.278-281.
3. Foley, R., *The Theory of Epistemic Rationality*, Harvard University Press, 1987.
4. Gossner, O., 'Comparison of information structures', *Games and Economic Behavior*, 1997.
5. Hirshleifer, J., 'The private and social value of information and the reward to inventive activity', *American Economic Review* 61, 561-574, 1971.
6. Kamien, M. I., Tauman, Y., and Zamir, S., 'On the value of information in a strategic conflict', *Games and Economic Behavior* 2, 129-153, 1990.
7. Kamien, M. I., Tauman, Y., and Zamir, S., 'Information transmission', in Ichiishi, T., Neyman, A., and Tauman, Y. (eds.), *Game Theory and Applications*, Academic Press, 273-281, 1990.
9. Korilis, Y. A., Lazar, A. A. and Orda, A., 'Avoiding the Braess paradox in non-cooperative networks', *Journal of Applied Probability* 36, 211-222, 1999.
10. Korner, S., *Metaphysics: Its Structure and Function*, Cambridge University Press, 1984.
11. Neyman, A. 'Bounded complexity justifies cooperation in the finitely repeated Prisoners' Dilemma', *Economic Letters*, 19 (1985).
12. Neyman, A., 'The positive value of information', *Games and Economic Behavior* 3, 350-355, 1991.
13. North, D. C., 'Institutions and economics', in *Blackwell Companion to Cognitive Science*, W. Bechtel and G. Graham, eds., Blackwell Publishers, 1998