WORKING PAPERS IN ECONOMICS

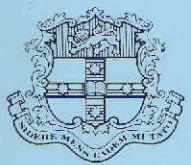
Moral rhetoric in the face of strategic weakness: modern clues for an ancient puzzle

by

Yanis Varoufakis

No. 211 **November 1994**

DEPARTMENT OF ECONOMICS



The University of Sydney
Australia 2006

Moral rhetoric in the face of strategic weakness: modern clues for an ancient puzzle

by

Yanis Varoufakis

No. 211 November 1994

Abstract

Moralising is a venerable last resort strategy. The ancient Melians presented the Athenian generals with a splendid example when in a particularly tight corner. In our Western philosophical tradition moral rhetoric is often couched in the form of reasons for action either external to preference and desire (eg.Kant) or internal to the agent's calculus of desire (eg. Hume, Gauthier). A third tradition dismisses such rhetoric as the last recourse of the weak (eg. Aristotle, Nietzsche) whereas a fourth calls for an examination of the social context (eg. Socrates, Marx, Wittgenstein, Habermas). This paper reports on an experiment which throws some empirical light on these debates and which offers a surprising twist to the interpretation of the Melian's plea.

Acknowledgment

I wish to thank Michael Jackson for his idea of a frame in which to place the discussion, Shaun Hargreaves-Heap for many suggestions during our joint experimental project, Wassily Kafouros, Gabrielle Meagher and Chris Rauchley for assistance with the experiments themselves, and Louis Haddad for comments on an earlier draft. Martin Hollis provided invaluable philosophical restraint and guidance. This project was supported by Australian Research Council grant no.24657 and Faculty of Economics, University of Sydney grant no. 25663. All errors are mine.

National Library of Australia Card Number and ISBN 0 86758 860 8

CONTENTS

1. The Melians' Plea: moral principle or clever tactic?

Morality has been hailed variously as a product of enlightened selfishness, the greatest proof of our autonomy, a social construct, an elaborate illusion; the list goes on. Regardless of perspective, its relation with strategy and justice has a long lineage. Thucydides reports that in the course of its geopolitical struggle against Sparta, Athens dispatched a fleet with the specific order that the independently minded island-state of Melos be subdued or razed to the ground. In the dialogue entered into by representatives of the two sides following the arrival of the Athenian troops, the interplay between moral principles and strategic concerns underscored the rhetoric.

In an opening speech anticipating Aristotle's infamous pronouncement that "The weaker are always anxious for justice and equality. The strong pay heed to neither" [**Politics**[1], s1318], the Athenians demanded Melos' surrender. After all, they decried,

> "...on the one hand the principles of justice, encompassed in human reason, hinge on the equal capacity to compel, yet on the other hand, the strong actually do what is possible and the weak suffer what they must." [Thucydides, **The History of the Peloponnesian War**, Book 5, s89]

The Melians played their only card. They demanded that they are allowed to remain neutral and free for Athens' *own* sake:

> "Then in our view (since you force us to base our arguments on self-interest, rather than on what is proper) it is useful that you should not destroy a principle that is *to the general good - namely that those who find themselves in the clutches of* misfortune should be justly and properly treated, and should be allowed to thrive beyond the limits set by the precise calculation of their power. And this is a principle which does not affect you less, since your own fall would be visited by the most terrible vengeance, watched by the whole world." [Thucydides, **The History of the Peloponnesian War**, Book 5, s90]

Necessity invented a splendid, and highly prophetic, argument: When in the dominant position do to others what you would like to be done to you when weak. If your behaviour is unconstrained by such a *principle*, you will live to regret it. Years latter, these words may have resonated in Athenian ears as the Spartans were scaling the walls of Pireus intent on destruction.

Of course Thucydides does not elaborate on the precise philosophical content of the Melians' argument. Were they envisaging principled behaviour as the solution to the calculus of long term Athenian preferences, or were they canvassing a universalisable principle to be activated by a pro-active reason (eg. Kant's advice: "Act only on that maxim which you can at the same time will that it should become a universal law")? To help unravel this ancient mystery, let us pay a visit to a laboratory where modern subjects are observed while interacting. Their behaviour, this paper contends, may hold clues to the Melians' strategy.

---

[1] All translations from the Greek text are the author's.

2. The experiment

A recent experiment (fully discussed elsewhere, see Varoufakis and Hargreaves-Heap, 1994) inadvertedly produced some interesting insights concerning the Melians' plea. It involved one hundred and thirty eight volunteers (75 men and 63 women) who played each of the following games four times. (They also played another two games which are not relevant here.) Most were University students from different faculties of Australian, Austrian, Greek and Hong Kong Universities. A small proportion of participants were professional people, most of whom, with University degrees. None had been exposed to game theory before.

To get a handle on how the games were played, consider Game 1 below. Suppose you are assigned role R: you are asked to choose a row strategy from the set {R1,R2,R3}. At the same time, someone else (whom you cannot see) is choosing among the set of column strategies {C1,C2,C3}. The payoff matrix translates such a pair of choices into payoffs; for example, if you chose R2 and your partner chose C2, then you win nothing and your opponent wins 5. Note that the first (second) payoff in each cell belongs to the row (column) player.

At close inspection, the three games appear similar in structure: for example, if both players choose their third strategy (R3 or C3) they collect the same payoff of 6. No other payoff can be better for either player, except if his or her counterpart receives a negative payoff [as in the case of outcome (R1,C3)]. If players do not choose R3 and C3, then one player will get (at best) 5 while the other will receive zero [outcomes (R1,C1) or (R2,C2)]. At worst they will both suffer a negative payoff [(R1,C2), (R2,C1)]. Thus the combination of strategies (R3,C3) corresponds to the only mutually beneficial outcome. However, the prospects of (R3,C3) being selected are seriously undermined by the fear that, while one is trying to bring (R3,C3) about by playing R3 or C3, the other will take advantage of this and go for the maximum payoff of 10. I will refer to this type of behaviour as *cheating* since it involves a premeditated attempt at short-changing the player who aims at the mutually beneficial outcome. By the same token, it makes sense to refer to strategies R3 or C3 as *cooperative* since they are only played by persons who disregard the fear that they will be cheated and who target the mutually beneficial outcome.

The three games played by subjects in the laboratory

	C1	C2	C3
R1	'5,0*	-1,-1	'10,-1
R2	-1,-1	'0,5*	-1,-2
R3	-1,10*	-2,-1	6,6

Game 1

	C1	C2	C3
R1	'5,0*	-1,-1	'10,-1
R2	-1,-1	'0,5*	-1,-2
R3	-1,1	-2,-1	6,6*

Game 2

	C1	C2	C3
R1	'5,0*	-5,-1	'10,-1
R2	-5,-1	'0,5*	-1,-2
R3	-1,1	-2,-1	6,6*

Game 3

Notice the (*,#) signs next to payoffs; they mark the instrumental rational responses of players R and C respectively. An instrumentally rational action is defined as that action which best serves the given goals of the agent. For example, if you have the R role in Game 3 and you expect your opponent to play C3, your counterpart to play C1, your highest payoff is given by strategy R1. Similarly if you have the C role, then your best response to R1 is to play C1. Thus the * and # marks. When these marks fall in the same cell, we have what game theorists call a Nash equilibrium in pure strategies. Their argument is that, unless a cell is a Nash equilibrium, then instrumentally rational players will have a tendency to switch to other strategies. Notice that in Game 1 the cooperative (R3,C3) outcome is incompatible with instrumental rationality since there are no * or # marks associated with it (ie. neither R3 or C3 is a rationally playable strategy; or, put differently, they are not 'clever' replies to anything one's counterpart may have played). In Games 2 and 3 C3 is a best reply to R3 [note the # next to the R3,C3 payoff outcome] but the opposite is not true. Hence a rational C (who expects that R will choose rationally) does not expect R to choose R3 - therefore, she will never play C3 either. So once more outcome (R3,C3) is not recommended by game theory as it is not part of any Nash equilibrium.

Table 1

So judging from the strategic structure of these games, instrumentally rational players will shun strategies R3 and C3. Player R, for instance, can always gain more by not playing R3 *regardless of what she may expect player C to do.* It follows, that C should infer this and harbour no hope of cooperation from R, in which case she will never play C3 either. Game theorists thus spot the (instrumental) irrationality of cooperative moves immediately. Of course the irony is that the moment cooperative moves are ruled out, so is the possibility that each player will gain payoff 6. This seems like one of these (prisoner's dilemma type) cases in which the rational pursuit of material gain is self-defeating.

And yet some people may cooperate regardless. Before observing our subjects' behaviour, let us recount some well-rehearsed explanations of why they do so (see Table 2). From an instrumental viewpoint, there three explanations

worth considering. One is that they have failed to recognise what is in their interest to do; that they have misread the situation and did not realise that cooperating is a dominated strategy. In this case [(1a) in Table 2] agents are expected to pay more attention to the strategic structure of the interaction (and thus cooperate less) the greater the stakes (ie. the larger the numbers in the payoff matrix) and the more experienced they are.

The second (1b) requires a shared future so that a good reputation for cooperativeness may yield long term benefits. Then the instrumental agent may bite her tongue and cooperate in the short run in order to enjoy a string of 6-payoffs in the medium run (what game theorists refer to as trigger strategies or tit-for-tat). The reason why this does not qualify as moral (or principled) behaviour is that when we reach the last interaction, our agent will immediately shuffle off her reputation and abandon the pretence of being a cooperative soul. [Indeed an interesting line of thought applies here according to which the finiteness of most interactions wrecks the chances of such enlightened selfishness regardless of the size of the long term gains. For a discussion see Pettit and Sugden (1989), Hollis (1991) and Varoufakis (1993).]

Six explanations for acts which defy the agent's strategic interests:

1. Instrumental

(1a) Execution errors (eg. mainstream game theory)

(1b) An investment in an agreeable reputation (eg. game theory again)

1(c) Natural sympathy (eg. Hume)

2. Instrumental-cum-moral

Moral action via hypothetical reasoning (eg. Gauthier)

3. Non-instrumental

(3a) Moral action via categorical reasoning (eg. Kant)

(3b) Social context (eg. Socrates, Hegel, Marx, Wittgenstein, Habermas)

Table 2

The final instrumental explanation (1c) takes us to Hume's (1740) *Treatise* where he argues that the agent's morality is to be found in her passions, inclinations or preferences[2] (as opposed to her reason who remains a slave to her passions). The moral agent shows some natural sympathy to the preferences of others and can rationally act on them in a manner which cannot be explained if we only take into consideration her personal gains. This would mean that the payoffs in the matrices above do not reflect the true preferences of individuals. For instance, players may derive an additional 5 'psychic' units from the cooperative outcome because they value a cooperative outcome *per se*. Thus it would be instrumentally rational to cooperate.

2 Here I conflate passions (which motivate in Hume's original account) and preferences (which motivate in the Hume-derived models of economists and rational choice theorists). This conflation can be misleading as it conceals the attempt to rid (unsuccessfully) choice from any moral psychology. See Hollis and Sugden (1993) for a discussion of why Hume's passions are closer to the mark than rational choice theory's preferences.

A radical extension of (1b) and (1c) above [corresponding to (2) in Table 2] has it that, once an agent recognises the value of cooperation, she has a reason to develop a cooperative *disposition* (as compared to simply acting cooperatively). Gauthier (1986) is the source and argues that the recognition of the value of principled behaviour becomes an independent reason for cooperating. In a sense, it literally pays to be moral. And since the instrumental meaning of 'rational' is grounded on how efficiently higher payoffs are secured, this type of instrumental-cum-moral explanation does not stray far from instrumental rationality. Nevertheless it challenges rather strongly the conventional instrumental approach by driving a wedge between rational choice and naked preference. Of course its weakness is, as Hollis (1993) explains, that unless a person undergoes a deeper ontological change (so that she can act on reasons external to her desires) she will not be able to sustain that *disposition*. Indeed being a person capable of self-restraint (eg. capable of overcoming the temptation to play R1 when you expect C to play C3) may pay more than being a straightforward payoff maximiser, but the best strategy will always be to dissemble as a principled agent and then cheat. And of course, when people do this, we are back to a world without moral dispositions.

For such an ontological transformation we need non-instrumental reasons for action. One clear suggestion [see (3a) in Table 2] comes from Kant (1788,1959) for whom the distinction between rational and moral choice recedes. Unlike Gauthier who calls for the development of a moral inclination internal to the agent, Kant's moral psychology distances her from her inclinations; instead it empowers her to trump her desires when they are incompatible with a universalisable (moral) principle. In this light, when people cooperate in our games this is seen as worthy action because it is activated by the 'right' motives and independently of consequences; even if players who cooperate gain more dollars. Greater gain is a welcome by-product and not the cause of moral action. However we are still in the realm of rational action because it is reason which, according to Kant, motivates the will in this particular way.

Finally we have a melange of explanations (3b) which lead us away from an individualist perspective. Returning for a moment to ancient Greece, Socrates suggests that prior to action we ought to ask ourselves: How should we live in order to achieve ευδαιμονια (loosely translatable into 'good living')? He suggests that our goal must be a successful life (as opposed to an enjoyable one) and that, crucially, it is 'others' who are to judge whether we have achieved our task. So, while purposely sidestepping the conceptual minefield of morality and virtue, Socrates introduces 'others' into our calculation of what it is rational for us to do.

More recently, Gilbert (1989) has commented that agents can coordinate their actions (and avoid the temptation to cheat) in interactions like those above provided they find a mutually beneficial and *shared* line of reasoning. This is different to Hume's natural sympathy argument because it is our reason which is responsible for the coalescence; not our passions. Suddenly our players see each other as partners, rather as opponents (nb. it is interesting that game theorists

always talk of 'opponents' even if a superior, mutually beneficial outcome such as (R3,C3) is available). If players manage to conceive of themselves as one decision-making unit (again notice the difference with Hume), then cooperative moves in our games cease to be paradoxes in need of de-mystification. However, to sustain this view it is important to explain how it might be rational to conceive of the 'other' as part of a unit to which you also belong. One way to do this is to follow Hurley (1989) in her attempt to establish some Archimedian vantage point from which to judge who can qualify (rationally) as partners. Clearly, expanding the borders of our 'self' to include others is one way in which cooperation in our games can be understood. But does this mean that we are expanding these borders because doing so is an end in itself? Not necessarily.

For the ancient Greeks moral action, as understood by Western philosophy, was not an issue (see Rowe, 1993). In the earlier description of our games I tended to describe the choice of strategy R3 in terms of a tension between cooperative (or moral) and self-interested (or instrumental) action. For Socrates this would be non-sense: to choose anything other than R3 when in the R role is shameful - regardless of whether any one is watching us! The crux in his thinking is the derivation of 'shame' from the realm of the Polis. Indeed it has nothing to do with morality depending on the extent to which the self has some natural sympathy for the preferences of others - as Hume (1740) suggested. Socrates might have agreed with Hume that we care about ourselves quite fundamentally. Yet the great difference is that to be rational in Socratic terms is not to be slaves to our preferences but rather to seek our own ευδαιμονια. And whereas preferences are private, the concept of ευδαιμονια is to be determined by the community.

So, the reason why we should cooperate is because we will not be leading the good life if we do not. What distinguishes Socrates from Kant, even though their recommendations to our players are the identical, is the same point which distinguishes him from Hume: whereas for Kant the perspective from which the right principles are drawn is that of the individual (who can derive these rules from the data of the game alone, without reference to social context), for Socrates it is the perspective of the 'others' which counts. It is through the eyes of the community that we must try to see whether our actions correspond to those which are constitutive of the good life.

Others have followed Socrates down this path. Hegel's (1963) conception of a reason which evolves as the 'self' rationally reflects on the 'other' and which ultimately reflects the progress of political society, is but one example. Marx (1963) - a spiritual child of the ancient Greeks and of Hegel - denounced any attempt at defining the meaning of rationality outside the specific social context as shaped by the technology and social organisation of the community. Wittgenstein (1953) rejects that action (such as choosing the cooperative strategy in our games) should be assessed by exclusive reference to the data of the game, or the mental state of the agents. Instead he would suggest that the moment the game is *described*, a process of interpretation of each strategy begins; players try to attach meaning to their available strategies. If they conceptualise strategies R3/C3 by means of the

linguistic signifier 'cooperative', then whether they will play them or not depends on whether there is an institution of cooperating in the community from which they have been abstracted and which has created their language. Finally, Habermas (1990) completes this greco-german group with his famous definition of rationality as communicative action.

It is now time to return to the experiment and see whether it sheds any light on all this. The task of sifting through the six explanations of cooperative behaviour is assisted firstly by the experimental design and secondly by the observed behaviour. Let us start with the former. Our players were divided in groups (group size ranged from 8 to 16) and punched their choices simultaneously into a computer terminal without knowing who they were playing against; a computer network assigned them at random to some other player in their group. Moreover, in each round they played against another random draw from the group. However the computer did not allow for the same pair to play a game twice in a row. Each game was played four times before proceeding to the next. Another constraint to the randomisation of pairs within the four repetitions of the same game was that each player was assigned the role of R twice and the role of C twice. Subjects were made aware of all this at the outset.

At the end of each round players were informed of their opponents' (or should I say 'partner's'?) choice, and thus of their score, as well as of the frequency with which different strategies and outcomes eventuated in their group. At the end of the session, their payoffs from each round were summed up and translated into Australian dollars. For instance, if during some round of Game 1 outcome (R1,C3) occurred, then the person with the R role was credited with A$10 while the one with the C role lost A$1. At the outset we guaranteed our subjects a minimum (final) payment of A$10 in order to dispel any fears that they would make a net loss; nevertheless this floor never became binding. Instead players netted an average profit from these games in excess of A$40 - a reasonable payoff (especially so for students) for forty five minutes' work.

Evidently the experimental design ruled out explanation (1b) outright: since the games were played anonymously and the chances of meeting the same player in the next round were zero, there is no room for explaining cooperative moves as an investment in reputation. As we will see shortly, the results themselves cast doubt on some of the other explanations. Table 3 offers an overview of the aggregate results. Clearly, there was a great deal of cooperation. Thus explanation (1a) demands that we dismiss more than half of our subjects as rationally defective, especially in view of the relative high stakes involved (recall how a successful cheating move rewarded a subject with A$10). Even though there is no way of *proving* that this is not the case (albeit that conclusion should cause panic in our Universities), the reader will agree that (1a) offers a weak explanation of what is happening here.

	Outcome (R1,C1)	Outcome (R2,C2)	Outcome (R3,C3)
Game 1	77	0	67
Game 2	51	0	69
Game 3	58	0	54

A summary of the incidents of asymmetrical distributions [(R1,C1),(R2,C2)] and of successful cooperation (R3,C3). Each row of the table is the sum of the observations from the four rounds of each game.

Table 3

The aggregate data of Table 3 leaves explanations (1c), (2),(3a) and (3b) in play. If we are to discriminate between them, we need to look more closely. Table 4 does this by breaking down the data in two sub-tables: one for the R role and one for the C role. The first column lists the raw number of observed cooperative attempts. The second column (labelled P3P3, standing for: predicted strategy 3, played strategy 3) tells us how many times a cooperative strategy was played *when the person who played it was anticipating cooperation*. [To compile the data of Table 4 we had to ask players to punch in their keyboard a prediction of which strategy their partner would choose prior to making their own strategy choice.] The third column relates the cases in which a player expected cooperation (that is, expected the third strategy) but chose not to reciprocate and played strategy 1 instead - the cheating (or 'zapping') option.

Row Players

ROW PLAYERS	Cooperated- ie. played R3	P3P3 - ie. predicted C3 and then played R3	Cheat - ie. predicted C3 and then played R1
Game 1	122	89	75
Game 2	89	74	130
Game 3	72	58	147

Table 4(a)

Column Players

COLUMN PLAYERS	Cooperated - ie. played C3	P3P3 - ie. predicted R3 and then played C3	Cheat - ie. predicted R3 and then played C1
Game 2	134	96	62
Game 3	194	105	N/A
Game 5	190	127	N/A

Table 4(b)

Total number of choices in Games 1,2 and 3 = 138x4x3 = 1656; Total number of P3P3 choices = 549; Average of P3P3 incidence = 33%; Average for Rs = 27%; Average for Cs = 40%. The 'Cheat' column does not report data for Games 3&5 because in those games C-players who predict R3 are best off replying with C3; that is, they have no incentive to cheat.

The most interesting column is the one labelled P3P3 since it reports on the extent to which subjects transcended the strategic structure of the game (which, let us not forget, advises players to avoid cooperation) and cooperate in response to an expectation that their opposite number will do likewise. Encouragingly, the overall proportion of such occurrences is a high 33%. However a puzzle sets in when we break this down further. The proportion for those in the R-role is only 27% whereas for those in the C-role a much higher 40% is registered. Why?

Notice that the explanation cannot be person-specific since, given the experiment's design, each player occupied the role of R and the role of C exactly as often. Thus the R-players and the C-players are the very same persons! It is therefore doubtless that this difference is role-specific. So, what is it that makes the two roles different? Looking at Game 1, in strategic terms the two roles are different since payoff A$10 corresponds to the same row strategy (R1) as payoff A$5 for player R, whereas this is not so for C. Games 2&3 feature a greater asymmetry since, unlike R, C cannot even aim at cheating. However the prediction which emanates from this strategic asymmetry is that the players will be converging towards strategies (R1,C1) and away from (R2,C2) - which we do observe. What it does not explain is why those in the C-role cooperate more (recall that if a C player expects R1 her best response is C1). Is this phenomenon by any of our explanations in Table 2?

Let us begin with explanation (2) in Table 2. For Gauthier (1986), rational agents should foresee that unconstrained maximisation will, at best, lead them to the rather poor payoffs resulting from strategy combinations (R1,C1) or (R2,C2). By contrast, if they could acquire a cooperative *disposition* they would be captivated by strategies (R3,C3) and, as a result, boost their rewards. And since they know well that they are alternating between the R and C positions, the strategic asymmetry mentioned above between the Rs and the Cs should make no difference: they should act as constrained maximisers who regardless of role opt for the third strategy. [The only room allowed by Gauthier for different propensities to cooperate in games such as ours is if one type of player has a greater tendency than the other to recognise the benefits from moral restraint or, alternatively, is more likely to avoid detection for cheating. However given our setting, neither are plausible explanations.]

Another explanation which seems unsatisfactory is (3a). If our subjects' cooperative moves are due to some type of categorical reasoning *a la* Kant, why did they cooperate more when in the C role? Surely a universalisable imperative ought to be just that and to demand that the agents' principle trumps calculations of strategic advantage consistently. Which leaves us with explanations (1c) and (3b). According to the former, the role-specificity of cooperative behaviour must be due to the different passions (eg. one for money, another for equity etc.) brought into the laboratory by the players. What distinguishes this Humean interpretation from the conventional rational choice model is that Hume makes no assumption regarding the commensurability of these passions. It may very well be the case that the agent is torn in a manner which cannot be easily settled by means of a clean ordering of preferences.

This would explain why some times a player cooperates while at others she does not, even when the objective (ie. the strategic) data remain unchanged. And (if Hume was right) this has nothing to do with the person's rationality - if the passions are unruly it is not reason (the passions' slave) who is to blame. A similarly contextual interpretation is offered by the melange of thinkers under (3b). The crucial difference is that they, unlike Hume, do not conceptualise motivation as clinically separable between an impartial, static, asocial reason on the one side and the non-rational passions (in which the will lies) on the other. For Socrates, Hegel, Marx, Wittgenstein and others, people create their reason as they create the rest of their lives: socially. For them, the three strategies in our games are first *interpreted* in terms of social data they have brought with them, and then selected on the basis of that interpretation. And since interpretation is inherently haphazard due to persons' diverse social location, different things and ideas motivate different people in our laboratory regardless of our attempts to impose on them (through the experimental design) uniform ends. The different propensity of players to cooperate depending on whether they are assigned the R or the C role is thus explained by the interpretative differences in motivations (their ideology as Marx would insist) engendered by the payoff matrices' asymmetries. Indeed subjects may plausibly come to the conclusion that, when a C, it is *better* to be more cooperative than when an R; where *better* is of course to be understood independently of instrumental concerns.

The most remarkable observation has been left last: as time went by and our players moved from one game to the next, the frequency of P3P3 dropped for those in the R role fast while it rose (even faster) for those in the C position! In Game 1 the frequency of P3P3 was 89 for the Rs and 96 for the Cs. By Game 3 these frequencies had become 58 and 127 respectively. Why? To remain consistent with Hume we need to resort to his theory of conventions. When the passions under-determine choice, agents achieve consistent behavioural patterns through trial and error. Thus a social convention emerges which agents learn to observe because doing so reduces wasteful indeterminateness. So although it is within reason to follow conventions, the specific convention people end up following is not uniquely rational. As Hume (1740) put it: "Tis not, therefore, reason, which is the guide of life but custom." In our laboratory our subjects generated endogenously a social convention according to which those in the better strategic

position (the Rs) cooperate less and those in the weaker position (the Cs) more[3].

Naturally Hume's separation of reason from society is highly controversial and not everyone's favourite move (see Varoufakis, 1991). Under (3b) we find a host of alternative explanations of the dynamics of cooperative behaviour as socially and historically determined. Socrates introduced the idea of reason as argument in the public sphere and Hegel saw reason develop within the evolution of social norms (unlike Hume's reason which is a disinterested, unchanging part of the agent's mind). Marx tied that evolution to the history of the social organisation of production and Wittgenstein introduced us to the mutual constitution of action and structure in the practices of a community of persons. Any one of these traditions could shed light on the social convention which was generated endogenously in our laboratory: namely that, **those in the strategically weaker position develop a tendency towards expectations, actions and rhetoric which can be seen as morally motivated**[4].

3. Back to the Melians: Imperialism and the moral authority of the weak

The actual events depicted in the *Peloponnesian War* make the questions in this paper look rather academic. The Melians' plea was doomed from the start regardless of its elegance, rationality or moral content. It failed because Athens did not aim at a reputation for magnanimity in victory. Indeed its objective was the opposite: a reputation for ruthlessness towards those 'allies' who absconded its sphere of influence (an ironic twist on explanation (1b) of Table 2). Melos' fate is testimony to the impotence of abstract morality against the logic of imperialism[5]. Yet it offers no evidence that moral rhetoric is irrelevant: at least one Athenian (that is, Thucydides) was impressed by their argument.

---

3 Recently evolutionary game theory (see Friedman, 1991 and Hargreaves-Heap and Varoufakis, 1995) has introduced conventions into a modern theoretical context. However it assumes that conventional behaviour cannot encompass strategies which are dominated (eg. R3,C3 in our games). Yet there is nothing in Hume to suggest that this exclusion is called for. Our results confirm this.

4 The actual experience of observing subjects play these games was particularly insightful. From the operator's console which allowed me access to everyone's terminal, I would often notice players who had just chosen non-cooperatively while in the R-role (ie. the dominant role) switch to the cooperative strategy when in the C-role. Frequently a moral indignation would be painted on their faces when the R-player to whom they were randomly assigned refused to cooperate (even though they had just acted in an identical fashion as R-players moments before). What a pity such 'data' is unquantifiable!

5 The Athenians were disarmingly honest on this. Asked why they could not accept an independent, yet friendly, Melos they replied: "No, because we are not injured by your hostility; rather we are worried that, if we were on friendly terms with you, those whom we have already subjugated would regard this as a sign of weakness in us, whereas your hostility is evidence of our power." [Thucydides, **History of the Peloponnesian War**, Book 5, 95]

Our ... three thoughts to the assessment of that argument:

> (a) Deeds with a moral appea... ...ted expedience: In an environment which abolished the u... ...se of any anticipated gains to be had from a reputatio... ...te proved remarkably resilient (see Table 3).
> (b) A moral disposition is unlikely to be acquired instrume... (1986) suggests. Additionally, and contrary to Kant's view, principle... seems neither universalisable nor independent of strategic motivation.
> (c) The weaker strategic role produced a greater expectation of good deeds (cooperation in our case) as well as a propensity to carry them out. Moreover these propensities evolved with time.

Thus the experiment does not lend straightforward insights to the Melian case. On their side they have the first finding: there is room for a moral stance such as theirs irreducible to strategic self-interest. However the next two are less sympathetic. Finding (b) reinforces scepticism about the chances that Athens' imperialist plans would be shelved, trumped by moral (internal or external) reasons which were supposed to have been activated by the Melian representative's fiery speech. To make things worse, had the Athenians had access to finding (c), they would be tempted to dismiss the Melian speech as the inevitable moralising of the feeble.

In many ways, the Athenian general foreshadowed this by implying that a re-run of history following a reversal of fortunes would offer conclusive evidence on the insincerity of the Melian position: "...[W]e know that you or anybody else with the same power as ours would be acting in precisely the same manner" [s104]. Was he right to think so? Unlike history which is not obliging on this, the verdict from the laboratory leans in his favour. At the end of some experimental sessions, subjects were asked to talk about their choices. Those who were asked questions about their behaviour, and who were assigned the R role in the last round of the session, made frequent use of words such as 'strategy', 'gain', 'opportunity'. By contrast, the rest tended to use terms such as 'right', 'mutual' and 'common'.

There is one perspective which has been so far confined conspicuously to a quotation by Aristotle in the Introduction. It takes its strongest form in the words of Nietzsche:

> "...there is master morality and slave morality...those qualities which serve to make easier the existence of the suffering will be brought into prominence and flooded with light...Slave morality is the morality of utility" [Nietzsche, 1973]

As the Melians tragically found out, and our experiment confirmed, a world systematically segregated between the dominant and the lesser social roles may indeed evolve into a world of slave and master moralities. This is where Aristotle and Nietzsche were right. Where I hope they were wrong is in their conviction that such segregation is due to *natural* differences between people. Thankfully our experiment cast doubt on this interpretation: Table 4 shows clearly that one's

moral disposition depends on one's *social location* rather than on an intrinsic strength or weakness (since the Rs and the Cs were the same persons). Therefore Nietzsche's separation between the weak and the strong may well be as artificial in society as it was in our laboratory - in which case all that is needed to undo it is a re-designed social context. If this is so, there is hope that Nietzsche was also wrong to think that the Will to Exploit is "...a consequence of the Will to Power, which is after all the Will to Life" [Nietzsche, 1973]. The Melians certainly thought so.

References

Aristotle, **Politics**

Friedman, D. (1991), 'Evolutionary Games', **Econometrica**, 59, 637-66

Gauthier, D. (1986), **Morals by Agreement**, Oxford: Oxford University Press

Gilbert, M. (1989), **On Social Facts**, London: Routledge

Habermas, J. (1990), **Moral Consciousness and Communicative Action**, trans. C. Lenhardt and S. Nicholsen, Oxford: Polity Press

Hegel, G.W.F. (1953), **Reason in History**, trans. J. Baillie, London: Macmillan

Hollis, M. (1989), 'Honour Among Thieves', **Proceedings of the British Academy**, LXXV, 163-180

Hollis, M. (1991), 'Penny Pinching and Backward Induction', **Journal of Philosophy**, 88, 473-88

Hollis, M. (1993), 'The Agriculture of the Mind' in D. Gauthier and R. Sugden (eds), **Rationality, Justice and the Social Contract**, Hemel Hempstead: Wheatsheaf

Hollis, M. and R. Sugden (1993), 'Rationality in Action', **Mind**, 102, 1-35

Hume, D. (1740), **A Treatise on Human Nature**, 1978 edition, Oxford: Clarendon

Kant, I. (1788), **Critique of Practical Reason**, in L. Beck (trans. and ed.) **Critique of Practical Reason and other writings in Moral Philosophy**, Cambridge: Cambridge University Press

Kant, I. (1959), **The Fundamental Principles of the Metaphysic of Morals**, London: Longmans

Marx. K. (1963), **The Poverty of Philosophy**, New York: International Publishers

Nietzsche, F. (1973), **Beyond Good and Evil: prelude to a philosophy of the future**, Hammondsworth: Penguin

Pettit, P. and R. Sugden (1989), 'The Paradox of Backward Induction', **Journal of Philosophy**, 86, 169-82

Rowe, C. (1993), 'Ethics in Ancient Greece' in P. Singer (ed), **A Companion to Ethics**, Oxford: Blackwell

Thucydides, **The History of the Peloponnesian War**

Varoufakis, Y. (1991), **Rational Conflict**, Oxford: Blackwell

Varoufakis, Y. (1993), 'Modern and Postmodern Challenges to Game Theory', **Erkenntnis**, 38, 371-404

Varoufakis, Y. and S. Hargreaves-Heap (1994), 'Dominated Cooperative Strategies and Equilibrium Selection: some additional evidence', mimeo, University of Sydney

Wittgenstein, L. (1953), **Philosophical Investigations**, Oxford: Blackwell

Working Papers
in Economics

163	Y. Varoufakis	Freedom within Reason from Axioms to Marxian Praxis; August 1991
164	D.J. Wright	Permanent vs. Temporary Infant Industry Assistance; September 1991
165	C. Karfakis & A.J. Phipps	Covered Interest Parity and the Efficiency of the Australian Dollar Forward Market: A Cointegration Analysis Using Daily Data; November 1991
166	W. Jack	Pollution Control Versus Abatement: Implications for *Taxation Under Asymmetric Information*; November 1991
167	C. Karfakis & A. Parikh	Exchange Rate Convenience and Market Efficiency; December 1991
168	W. Jack	An Application of Optimal Tax Theory to the Regulation of a Duopoly; December 1991
169	I.J. Irvine & W.A. Sims	The Welfare Effects of Alcohol Taxation; December 1991
170	B. Fritsch	Energy and Environment in Terms of Evolutionary Economies; January 1992
171	W.P. Hogan	Financial Deregulation: Fact and Fantasy; January 1992
172	P.T. Vipraio	*An Evolutionary Approach to International Expansion: A Study* for an Italian Region; January 1992
173	C. Rose	Equilibrium and Adverse Selection; February 1992
174	D.J. Wright	Incentives, Protection and Time Consistency; April 1992
175	A.J. Phipps, J. Sheen & C. Wilkins	The Slowdown in Australian Productivity Growth: Some Aggregated and Disaggregated Evidence; April 1992
176	J.B. Towe	Aspects of the Japanese Equity Investment in Australia; June 1992
177	P.D. Groenewegen	Alfred Marshall and the Labour Commission 1891-1894; *July* 1992
178	D.J. Wright	Television Advertising Regulation and Programme Quality; August 1992
179	S. Ziss	Moral Hazard with Cost and Revenue Signals; December 1992
180	C. Rose	The Distributional Approach to Exchange Rate Target Zones; December 1992
181	W.P. Hogan	Markets for Illicit Drugs; January 1993
182	E. Jones	The Macroeconomic Fetish in Anglo-American Economies; January 1993
183	F. Gill	Statistics in the Social Sciences A Mixed Blessing? March 1993
184	Y. Varoufakis & S. Hargreaves-Heap	The Simultaneous Evolution of Social Roles and of Cooperation; April 1993
185	C. Karfakis & D.M. Moschos	The Information Content of the Yield Curve in Australia; April 1993
186	C. Karfakis & A. Parikh	Uncovered Interest Parity Hypothesis for Major Currencies; May 1993
187	C. Karfakis & A.J. Phipps	Do Movements in the Forward Discount on the Australian Dollar Predict Movements in Domestic Interest Rates? Evidence from a Time Series Analysis of Covered Interest Parity in Australia in the late 1980s; May 1993
188	J.B. Towe	Citation Analysis of Publications on the Australian Tariff Debate, 1946-1991; August 1993

189	C. Karfakis & S-J Kim	Exchange Rates, Interest Rates and Current Account News: Some Evidence from Australia; September 1993
190	A.J. Phipps & J.R. Sheen	Unionisation, Industrial Relations and Labour Productivity Growth in Australia: A Pooled Time-Series/Cross-Section Analysis of TFP Growth; September 1993
191	W.P. Hogan	Market Value Accounting in the Financial Sector; November 1993
192	Y. Varoufakis & W. Kafouros	The Transferability of Property Rights and the Scope of Industrial Relations' Legislation: Some Lessons from the NSW Road Transport Industry; November 1993
193	P.D. Groenewegen	Jacob Viner and the History of Economic Thought; January 1994
194	D. Dutta & A. Hussain	A Model of Share-Cropping with Interlinked Markets in a Dual Agrarian Economy; March 1994
195	P.E. Korsvold	Hedging Efficiency of Froward and Option Currency Contracts; March 1994
196	J. Yates	Housing and Taxation: An Overview; March 1994
197	P.D. Groenewegen	Keynes and Marshall: Methodology, Society and Politics; March 1994
198	D.J. Wright	Strategic Trade Policy and Signalling with Unobservable Costs; April 1994
199	J. Yates	Private Finance for Social Housing in Australia; April 1994
200	L. Haddad	The Disjunction Between Decision-Making and Information Flows: The Case of the Former Planned Economies; April 1994
201	P.D. Groenewegen & S. King	Women as Producers of Economic Articles: A Statistical Assessment of the Nature and the Extent of Female Participation in Five British and North American Journals 1900-39; June 1994
202	P.D. Groenewegen	The French Connection: Some Case Studies of French Influences on British Economics in the Eighteenth Century; June 1994
203	F. Gill	Inequality and the Wheel of Fortune: Systemic Causes of Economic Deprivation; July 1994
204	M. Smith	The Monetary Thought of Thomas Tooke; July 1994
205	A. Aspromourgos	Keynes on the Australian Wages System; July 1994
206	W. Kafouros & Y. Varoufakis	Bargaining and Strikes: From an Equilibrium to an Evolutionary Framework; July 1994
207	A. Oswald & I. Walker	Rethinking Labor Supply: Contract Theory and Unions; July 1994
208	J.B. Towe & D.J. Wright	The Research Output of Australian Econometrics and Economics Department: 1988-93; *July* 1994
209	F. Gill & C. Rose	Discontinuous Payoff Functions under Incomplete Information; August 1994
210	S-J Kim	Inflation News in Australia: Its Effects on Exchange Rates and Interest Rates; October 1994
211	Y. Varoufakis	Moral Rhetoric in the Face of Strategic Weakness: Modern Clues for an Ancient Puzzle

Copies are available upon request from:

Department of Economics
The University of Sydney
N.S.W. 2006, Australia

Working Papers in Economics Published Elsewhere

2	I.G. Sharpe & R.G. Walker	*Journal of Accounting Research*, *13(2)*, Autumn 1975
3	N.V. Lam	*Journal of the Developing Economies*, *17*(1), March 1979
4	V.B. Hall & M.L. King	*New Zealand Economic Papers*, *10*, 1976
5	A.J. Phipps	*Economic Record*, *53*(143), September 1977
6	N.V. Lam	*Journal of Development Studies*, *14*(1), October 1977
7	I.G. Sharpe	*Australian Journal of Management*, April 1976
9	W.P. Hogan	*Economic Papers*, 55, The Economic Society of Australia and New Zealand, October 1977
12	I.G. Sharpe & P.A. Volker	*Economics Letters*, 2, 1979
13	I.G. Sharpe & P.A. Volker	*Kredit and Kapital*, *12*(1), 1979
14	W.P. Hogan	*Some Calculations in Stability and Inflation*, A.R. Bergström et al (eds.), J. Wiley & Sons, 1978
15	F. Gill	*Australian Economic Papers*, *19*(35), December 1980
18	I.G. Sharpe	*Journal of Banking and Finance*, *3*(1), April 1978
21	R.L. Brown	*Australian Journal of Management*, *3*(1), April 1978
23	I.G. Sharpe & P.A. Volker	*The Australian Monetary System in the 1970s*, M. Porter (ed.), Supplement to Economic Board 1978
24	V.B. Hall	*Economic Record*, *56*(152), March 1980
25	I.G. Sharpe & P.A. Volker	*Australian Journal of Management*, October 1979
27	W.P. Hogan	*Malayan Economic Review*, *24*(1), April 1979
28	P. Saunders	*Australian Economic Papers*, *19*(34), June 1980
29	W.P. Hogan I.G. Sharpe & P.A. Volker	*Economics Letters*, 6 (1980), 7 (1981)
30	W.P. Hogan	*Australian Economic Papers*, 18(33), December 1979
32	R.W. Bailey, V.B. Hall & P.C.B. Phillips	*Keynesian Theory, Planning Models, and Quantitative Economics*, G. Gandolfo and F. Marzano (eds.), 1987
38	U.R. Kohli	*Australian Economic Papers*, *21*(39), December 1982
39	G. Mills	*Journal of the Operational Research Society* (33) 1982
41	U.R. Kohli	*Canadian Journal of Economics*, *15*(2), May 1982
42	W.J. Merrilees	*Applied Economics*, 15, February 1983
43	P. Saunders	*Australian Economic Papers*, *20*(37), December 1981
45	W.J.Merrilees	*Canadian Journal of Economics*, *15*(3), August 1982
46	W.J. Merrilees	*Journal of Industrial Economics*, *31*, March 1983
49	U.R. Kohli	*Review of Economic Studies*, *50*(160), January 1983
50	P. Saunders	*Economic Record*, *57*(159), December 1981

53	J. Yates	AFSI, *Commissioned Studies and Selected Papers*, AGPS, IV 1982
54	J. Yates	*Economic Record*, *58*(161), June 1982
55	G. Mills	*Seventh Australian Transport Research Forum-Papers*, Hobart 1982
56	V.B. Hall & P. Saunders	*Economic Record*, *60*(168), March 1984
57	P. Saunders	*Economic Record*, *59*(166), September 1983
58	F. Gill	*Économie Appliquée*, *37*(3-4), 1984
59	G. Mills & W. Coleman	*Journal of Transport Economics and Policy*, 16(3), September 1982
60	J. Yates	*Economic Papers*, Special Edition, April 1983
61	S.S. Joson	*Australian Economic Papers*, *24*(44), June 1985
62	R.T. Ross	*Australian Quarterly*, *56*(3), Spring 1984
63	W.J. Merrilees	*Economic Record*, *59*(166), September 1983
65	A.J. Phipps	*Australian Economic Papers*, *22*(41), December 1983
67	V.B. Hall	*Economics Letters*, 12, 1983
69	V.B. Hall	*Energy Economics*, *8*(2), April 1986
70	F. Gill	*Australian Quarterly*, *59*(2), Winter 1987
71	W.J. Merrilees	*Australian Economic Papers*, *23*(43), December 1984
73	C.G.F. Simkin	*Singapore Economic Review*, *29*(1), April 1984
74	J. Yates	*Australian Quarterly*, *56*(2), Winter 1984
77	V.B. Hall	*Economics Letters*, 20, 1986
78	S.S. Joson	*Journal of Policy Modeling*, *8*(2), Summer 1986
79	R.T. Ross	*Economic Record*, *62*(178), September 1986
81	R.T. Ross	*Australian Bulletin of Labour*, *11*(4), September 1985
82	P.D. Groenewegen	*History of Political Economy*, *20*(4), Winter 1988 and *Scottish Journal of Political Economy*, *37*(1) 1990
84	E.M.A. Gross, W.P. Hogan & I.G. Sharpe	*Australian Economic Papers*, *27*(50), June 1988
85	F. Gill	*Australian Bulletin of Labour*, *16*(4), December 1990
94	W.P. Hogan	*Company and Securities Law Journal*, 6(1), February 1988
95	J. Yates	*Urban Studies*, *26*, 1989
96	B.W. Ross	*The Economic and Social Review*, *20*(3), April 1989
97	F. Gill	*Australia's Greatest Asset: Human Resources in the Nineteenth and Twentieth Centuries*, D. Pope (ed.), Federation Press, 1988
98	A.J. Phipps	*Australian Economic Papers*, *31*(58), June 1992
99	R.T. Ross	*Australian Bulletin of Labour*, *15*(1), December 1988
100	L. Haddad	*Hetsa Bulletin*, (11), Winter 1989
101	J. Piggott	*Public Sector Economics - A Reader*, P. Hare (ed.), Basil Blackwell, 1988
102	J. Carlson & D. Findlay	*Journal of Macroeconomics*, *13*(1), Winter 1991
102	J. Carlson & D. Findlay	*Journal of Economics and Business*, *44*(1), February 1992

104	P.D. Groenewegen	*Decentralization, Local Government and Markets: Towards a Post-Welfare Agenda*, R.J. Bennet (ed.) Oxford University Press, 1990
107	B.W. Ross	*Prometheus*, 6(2), December 1988
108	S.S. Joson	*Rivista di diritto valutario e di economia internazionale*, 35(2), June 1988
112	P. Groenewegen	*NeoClassical Economic Theory 1870 to 1930*, K. Hennings and W. Samuels (eds.), Boston Kluwer-Nighoff, 1990
113	V.B. Hall T.P. Truong V.A. Nguyen	*Energy Economics*, 12(4) October 1990
114	V.B. Hall T.P. Truong & V.A. Nguyen	*Australian Economic Review*, (87) 1989(3)
115	F. Gill	*Australian Journal of Social Issues*, 25(2), May 1990
116	G. Kingston	*Economics Letters*, 15 (1989)
117	V.B. Hall & D.R. Mills	*Pacific and Asian Journal of Energy*, 2(2), December 1988
118	W.P. Hogan	*Abacus*, 25(2), September 1989
120	P. Groenewegen	*Flattening the Tax Rate Scale: Alternative Scenarios & Methodologies*, (eds.) J.G. Head and R. Krever, 1990
122	W.P. Hogan & I.G. Sharpe	*Economic Analysis and Policy*, 19(1), March 1989
123	G. Mills	*Journal of Transport Economics and Policy*, 23, May 1989
126	F. Gill	*The Australian Quarterly*, 61(4), 1989
128	S. Lahiri & J. Sheen	*The Economic Journal*, 100(400), 1990
130	J. Sheen	*Journal of Economic Dynamics and Control*, 16, 1992
135	Y. Varoufakis	*Économie Appliquée*, 45(1), 1992
136	L. Ermini	*The Economic Record*, 69(204), March 1993
138	D. Wright	*Journal of International Economics*, 35, (1/2) 1993
139	D. Wright	*Australian Economic Papers*, 32, 1993
141	P. Groenewegen	*Australian Economic Papers*, 31, 1992
143	C. Karfakis	*Applied Economics*, 23, 1991
144	C. Karfakis & D. Moschos	*Journal of Money, Credit and Banking*, 22,(3), 1990
147	J. Yates	*Housing Studies*, 7, (2), April 1992
158	W.P. Hogan	*Economic Papers*, 10(1), March 1991
159	P.Groenewegen	*Local Government and Market Decentralisation: Experiences in Industialised, Developing and Former Eastern Block Countries*, R. J. Bennett (ed.) UN University Press, 1994
160	C. Karfakis	*Applied Financial Economics*, 1(3), September 1991
162	Y. Varoufakis	*Erkenntnis*, 38, 1993
163	Y. Varoufakis	*Science and Society*, 56(4), 1993
173	C. Rose	*The Rand Journal of Economics*, 24(4), Winter 1993
177	P. Groenewegen	*European Journal of the History of Economic Thought*, 1(2) Spring 1994
189	C. Karfakis & S-J Kim	*Jounal of International Money and Finance*, 14(4) August 1995
190	A.J. Phipps & J.R. Sheen	*Labour Economics and Productivity*, 6(1), March 1994