Patrick Suppes and Game Theory

Ken Binmore
Philosophy Department
University of Bristol
Cotham House
Bristol BS6 6JL, UK
K.Binmore@ucl.ac.uk

Abstract: This article is a contribution to a symposium celebrating the life of Patrick Suppes. It describes the context in which he made contributions relevant to two extremes of the game theory spectrum. At one extreme, he made an experimental study of whether laboratory subjects learn to use Von Neumann's minimax theory in games of pure conflict. At the other extreme, he invented a theory of empathetic identification that lies at the root of an approach to making interpersonal comparisons needed for the study of games in which cooperation is central rather than conflict. These pieces of work are peripheral to his major interests, but they nevertheless illustrate how it is possible to be an academic success without conceding anything to current academic fashion.

Patrick Suppes and Game Theory

by Ken Binmore

1 Introduction

Where should the line be drawn between philosophy and science? Aristotle and Descartes seemingly saw no reason to draw a line at all. Nor did Galileo or Newton. So Patrick Suppes was in good company when he wrote on whatever topic took his fancy with little regard either to disciplinary boundaries or to academic fashion.

To appreciate the range of interests of this extraordinary man, it is necessary to review the whole spectrum of work covered in the symposium in his honor of which this paper is a part. Here we only look at two of his pioneering contributions that are relevant to game theory.

It is typical of Patrick Suppes that he was working on these issues in the late 1950s and early 1960s when game theory was at its lowest ebb—after the initial enthusiasm that greeted the 1944 publication of Von Neumann and Morgenstern's [34] *Theory of Games and Economic Behavior* had dissipated into disillusion, and before there was an inkling of the revival of the late 1970s that would convert the original theory into a workhorse that is now regarded as the fundamental tool in economic theory. I do not think it would be useful for me to discuss the contents of the two pieces of work in detail since the world has moved on a great deal since those early days. Instead, I propose to review the context in which they were written and what the literature has to say about these issues in the present.

2 Zero-Sum Games

The first piece of work is described in a book *Applications of a Markov Model to Multiperson Interactions* written with Richard Atkinson, which belongs in a list of pioneering works in at least three areas (Suppes and Atkinson [31, 32]).

The first area to which it makes a contribution is the Psychology of Learning. A few years before Suppes and Atkinson's [31] original publication, the psychologists Bush and Mosteller [7] published a book documenting their research at the University of Michigan into stochastic models of stimulus-response learning. Their finding was that such models can be surprisingly effective in tracking the behavior over time of laboratory subjects engaged in simple tasks. However, their research was

not embraced with open arms by the psychology profession at the time, presumably because their approach seemed too much like the philosophical version of behaviorism promoted by Watson and Skinner, whose initial sky-high reputation was already falling into disrepute.[1] Unconcerned with such worldly considerations, Suppes and Atkinson exploited their comparative advantage in mathematical matters to formulate their own stochastic versions of stimulus-response learning models, which they then ambitiously set out to test experimentally in multi-person situations, thereby arguably initiating the subject of Evolutionary Economics, on which there is now a very large literature (Roth [26]).

The multi-person interactions in Suppes and Atkinson's experiments were framed as zero-sum games, and it is this aspect of their work on which I shall focus. Does Von Neumann's minimax theory of rational play in two-person zero-sum games work in practice? In seeking to answer such a question, Suppes has to be counted as one of the founders of the now flourishing subject of Experimental Economics.

Minimax theory. A game is being played whenever the actions of different people interact so that the players have to take account of the strategies that other players may choose when making their own choice of a strategy. In games of pure coordination, what is good for one player is good for them all, and so the problem the players face is how best to cooperate in order to achieve an outcome that all agree is optimal. Zero-sum games are games of pure conflict that lie at the other end of the cooperative spectrum. They are games in which the interests of the two players (Alice and Bob) are diametrically opposed, so that whatever is good for one is bad for the other. Chess and Poker are typical examples.

Von Neumann and Morgenstern's [34] book is divided into two parts. The first part is a rigorous analysis of zero-sum games, which establishes that rational players in such games will each randomize their choice of strategy so as to achieve their maximin expected payoff. Alice's randomized choice of strategy will then maximize her average payoff on the assumption that Bob will successfully predict what she is doing and act to minimize her average payoff on the assumption that his prediction is correct.[2] When both players use maximin strategies in a zero-sum game, the result is a Nash equilibrium, which means that each player is making a personally best reply to the strategy of the other.[3]

Perfect rationality? Economists of the 1950s were sold on a methodology that treated people as rational optimizers no matter how difficult it might be to work

[1] The behaviorism of Watson and Skinner needs to be sharply distinguished from the behavioral economics associated with the names of Kahneman and Tversky.

[2] Von Neumann's [33] Minimax Theorem says that a player's minimax and maximin payoffs are equal when players are allowed to randomize. But it does not follow—as one sometimes reads—that rational players should choose their minimax strategies rather than their maximin strategies.

[3] General games often have many Nash equilibria among which it can be difficult to make a rational choice, but all Nash equilibria in two-person zero-sum games are interchangeable and payoff-equivalent so that no equilibrium selection problem arises.

out what strategy might be optimal. In their defense, this methodology turns out to generate predictions that work exceedingly well in the perfectly competitive markets with which they were mostly concerned in those days. Nowadays we recognize that people are subject to all kinds of irrational biases and computational inadequacies, and the reason that idealizations of neoclassical economics sometimes work well is that market institutions provide an evolutionary environment in which trial-and-error experimentation can lead to very fast convergence on a Nash equilibrium in which everybody ends up behaving as if they had worked out the optimal thing to do by rational computation. The Chicago wheat market is a stereotypical example.

But things are not so easy in zero-sum games. In the first place, the reasoning that leads to the use of a maximin strategy is a lot more complicated than the "buy cheap and sell dear" philosophy that makes markets work. This point is nicely illustrated by the fact that Von Neumann was anticipated in the study of zero-sum games by the great mathematician Emile Borel. The record shows that Borel formulated the Minimax Theorem but decided that it was probably false. It would therefore seem pointless to run experiments designed to test the hypothesis that laboratory subjects are capable of duplicating Von Neumann's reasoning. Insofar as Von Neumann's minimax theory succeeds in predicting the behavior of laboratory subjects playing zero-sum games, it is not because they are all cleverer than Emile Borel! It is because their trial-and-error attempts to improve their past play converge on the minimax outcome. But under what conditions will such such convergence occur, if at all? And how long will it take? One certainly cannot hope for the fast convergence observed in some markets that allows the appearance of rational behavior to seem almost simultaneous with the onset of trading.

But this is what economists thought they should be looking for when experimental economics had finally become respectable in the late 1980s, and some pioneers turned their attention to zero-sum games. It is therefore not surprising that their experimental results were distant from the minimax prediction (except in games like Matching Pennies for which such an outcome is easily explained on other grounds).[4] The exception was a paper with repeated play by O'Neill [21] that he found hard to get published because it was out of line with the current literature. When finally published, it was denounced by Brown and Rosenthal [6] on the grounds that it failed a test which treated any evidence that the players were learning to play their maximin strategies as evidence $against$ Von Neumann's theory![5]

This wave of experimentation on zero-sum games came some 30 years after the original wave of which the paper by Suppes and Atkinson was a part.[6] The results of this earlier wave were no more supportive of the minimax hypothesis than the later wave, but what we know now is that if Suppes and Atkinson had been able to present their games to their subjects with computerized graphic displays

[4] For example, Rapaport and Boebel [23], Mookherjee and Sopher [19, 20] McCabe $et\ al$ [18].

[5] Because maximin players will randomize $independently$ each time they play, but the choices of players who are learning will be correlated as they find their way to equilibrium.

[6] The other papers were by Frenkel [9], Estes [8] and Malcolm and Lieberman [17].

(instead of using a pencil-and-paper implementation) and understood that it makes a difference whether you pay your subjects for their efforts, then their design would have supported minimax. The reason is that their design got the two important things right. Their experiment made provision for adequate trial-and-error learning, and their subjects were told in a user-friendly way precisely what game they were playing.

None of the second wave—not even O'Neill—were safe on the first point. As for the second point, the very first experiment on zero-sum games was carried out by the distinguished psychologist Estes [8] who reported that the minimax theory may have theoretical virtues but it has no practical application. However, his subjects did not even know they were playing another human being. Even if they guessed that they were playing a game, they did not know its payoffs. In particular, they did not know the game was zero-sum. So why should they play minimax?

My own contribution to the experimental literature on zero-sum games followed the second wave mentioned above (Binmore $et\ al$ [5, 4]). My coauthors and I ticked all the necessary boxes by following the methodology of Suppes and Atkinson using modern experimental techniques, and so obtained results that support the minimax hypothesis. However, as with O'Neill, our paper was initially rejected, partly because our results were not consistent with the accepted literature, but mostly because we took issue with the test of Brown and Rosenthal that had become standard by this time. However, when we finally got round to submitting a revised version several years later, the whole atmosphere had changed because of the appearance in the interim of several other papers in which convergence on the Nash equilibria of a variety of games had been demonstrated in laboratory experiments. Nowadays it is well established that adequately motivated subjects given adequate time for trial-and-error learning in games with a unique Nash equilibrium will eventually converge on the Nash equilibrium most of the time.[7] Evolutionary biologists have found that corresponding results in games played by animals can sometimes be very successful in explaining behavior that would otherwise be very mysterious (Pollock $et\ al$ [22]). Indeed, such biological field studies are arguably the area in which game theory has made its greatest contributions to scientific endeavor.

In summary, Patrick Suppes deserves to be credited as a pioneer in the Psychology of Learning, in Evolutionary Economics, and in Experimental Economics.

3 Comparing Utilities

The second part of Von Neumann and Morgenstern's [34] book is also mostly about zero-sum games, but in multi-person situations that are distant from the two-person games of pure conflict studied in its first part. The emphasis now is on cooperation. The players are assumed to be able to agree on binding pre-play contracts on how

[7] A few behavioral economists deny this fact, even in the case of the Prisoners' Dilemma where the evidence is overwhelmingly strong. (See Ledyard's [15] survey of the very large literature).

the game will be played. Axioms are then proposed that supposedly govern the nature of such contracts for rational players.

While Suppes was engaged in experiments on two-person zero-sum games in the 1950s and 1960s, those few scholars who retained an interest in game theory were mostly working on developing variants of Von Neumann and Morgenstern's axioms—a field that become known as cooperative game theory (as opposed to noncooperative game theory, in which pre-play contracts are not permitted and the basic concept is that of a Nash equilibrium, as in the theory of two-person zero-sum games). An immediate problem is how players' payoffs or utilities are to be compared when pre-play contracts are being negotiated.[8]

The second part of Von Neumann and Morgenstern's [34] book not only takes it for granted that interpersonal comparisons of utility can be made, but that players can trade units of utility (utils) between them. The assumption that utils can be traded is a step too far for most modern critics, but the idea that utils can sensibly be compared across individuals is now a routine assumption, although it was once hotly denied that such comparisons could ever be meaningful. If this were true, I share the view expressed by Hammond [10], Harsanyi [12] and many others that rational ethics would then become a subject with little or no substantive content.

In the remainder of this article, I plan to review the assumptions that Harsanyi [13] added to Von Neumann and Morgenstern's [34] theory of expected utility to justify interpersonal comparisons of utility. In Harsanyi's approach, they arise from the study of what I call empathetic preferences (Binmore [3]). Such preferences originate with Patrick Suppes' [30] "Some Formal Models of Grading Principles", and were studied by Sen [28] and Arrow [1] under the name of "extended sympathy preferences" before being taken up by Harsanyi.

What is Utility? The word *utility* has always been difficult. Even the arch-utilitarian Jeremy Bentham [2] opens his *Principles of Morals and Legislation* by remarking that his earlier work would have been better understood if he had used *happiness* or *felicity* instead. Bentham's position, later taken up by John Stuart Mill, is that utility should be interpreted as the balance of pleasure and pain experienced by an individual human being.

Victorian economists adopted Bentham's idea and incorporated it into their models without paying much attention to its doubtful philosophical and psychological foundations. However, once economists discovered (in the "marginalist revolution" of the early part of the twentieth century) that they did not need to attribute utility functions to economic agents in order to prove most of the propositions that seemed important at the time, all of the baggage on utility theory inherited from the Victorian era was swept away. By the late 1930s, it had become fashionable for

[8] The same problem does not arise in two-person zero-sum games, for which it is only necessary that the players have opposing preferences over all pairs of alternatives. Their utility functions can then be normalized to add up to zero without assuming anything about the relative strength of their preferences.

economists to denounce cardinal utility theory as meaningless nonsense.[9] Even now, the jeremiads to this effect of the immensely influential Lionel Robbins [25] are still occasionally quoted. It is characteristic that when Suppes made his contribution to this subject in the 1960s, Robbin's views were almost unchallenged in the economics profession.

Modern philosophers largely share Robbins' misapprehension that utilities are necessarily measures of pain or pleasure, so that comparing utilities between individuals is like quantifying how much more Alice suffers in the dentist's chair than Bob. Such a misinterpretation of the modern neoclassical concept of utility is sometimes referred to as the Causal Utility Fallacy because it takes for granted that Alice chooses a over b $because$ her utility for a exceeds her utility for b. However, the modern Theory of Revealed Preference[10] disclaims any explanatory role and settles for a merely descriptive role. In particular, modern economists assign a utility function to Alice with $u(a) > u(b)$ because—for whatever reason—she would choose a over b if the opportunity arose.

Modern utility theory famously began when Oskar Morgenstern turned up at Von Neumann's house in Princeton one day in the early forties complaining that they didn't have a proper basis for the payoffs in the book they were then writing together. So Von Neumann invented a theory on the spot that measures how much a rational person wants something by the size of the risk he or she is willing to take to get it. The rationality assumptions built into Von Neumann's theory simply require that people make decisions in a consistent way, but his conclusions are surprisingly strong. Anyone who chooses consistently in risky situations will look to an observer as though he or she were trying to maximize the average value of something. This abstract "something" is what is called utility in the modern theory.

Comparing utils. Why does a rich man hail a taxicab when it rains while a poor man gets wet? Economists answer this traditional question by making an intrapersonal comparison. They argue that an extra dollar in Alice's pocket would be worth more to her if she were poor than it would be if she were rich. But how are we to quantify such a comparison? The answer offered by the Von Neumann and Morgenstern theory is that we can count the number of utils by which her total utility is increased when she gains an extra dollar. Because a rational decision-maker in the Von Neumann and Morgenstern theory acts as though maximizing average utility, she behaves as though each util is "worth" the same as any other.

But what of the utils acquired by a rational Bob? In the absence of a Tiresias with experience of both roles, to whom do we appeal when asked to compare Alice and Bob's utils? The rationality postulates of Von Neumann and Morgenstern provide no justification for such comparisons—which fact provided extra ammunition

[9] A cardinal utility scale operates like a temperature scale, with utils replacing degrees. It is normally contrasted with an ordinal utility scale, in which the amount by which the utility of one outcome exceeds the utility of another outcome is held to be meaningless.

[10] I would prefer to say the Theory of Attributed Preference.

for the die-hard followers of Robbbins in denying that rational interpersonal comparison is possible at all. However, as Harsanyi pointed out, what is to prevent our adding further postulates that do allow interpersonal comparison? Harsanyi's [13] additional assumptions are built on the notion of what I call empathetic preferences as introduced by Patrick Suppes [30].

In surveying the history of utilitarianism, Russell Hardin [11] dismisses Hume's emphasis on the importance of sympathetic identification between human beings as idiosyncratic. Although Adam Smith [29] followed his teacher in making human sympathy a major plank in his *Theory of Moral Sentiments*, Hardin is doubtless broadly right in judging that later moral philosophers appeal to human sympathy only when in need of some auxiliary support for a conclusion to which they were led largely by other considerations. Suppes was therefore again treading where others chose not to go.

What Hume and Adam Smith called sympathy is nowadays usually known as *empathy;* psychologists tend to reserve the word *sympathy* for a stronger notion. Alice empathizes with Bob when she imagines herself in his shoes in order to see things from his point of view. She sympathizes with Bob when she identifies with him so strongly that she is unable to separate her interests from his. The theory of revealed preference has no difficulty in describing Alice's behavior in such a case of sympathetic identification. One simply writes altruistic (or spiteful) preferences into her utility function.

It is easy to see why the forces of biological evolution might lead to our behaving as though we were equipped with sympathetic preferences. Mothers commonly care for their children more than they do for themselves. In such basic matters as these, it seems that we differ little from crocodiles or spiders. However, humans do not sympathize only with their children; it is uncontroversial that they also sympathize to varying degrees, with their husbands and wives, with their extended families, with their friends and neighbors, and with their sect or tribe. Modern behavioral economists are willing to proceed as though we all sympathize with everybody to much the same degree that we sympathize with our near and dear. If this were true, then the Von Neumann and Morgenstern theory would be adequate all by itself to determine a standard of interpersonal comparison, because Alice would only need to consult her own sympathetic utility function to find out how many utils to assign to a change in Bob's situation as compared with some change in her own situation. But Harsanyi's [13] approach is less naive. He argues that, alongside our personal preferences (which may or may not include sympathetic concerns for others), we also have empathetic preferences that reflect our ethical concerns.

Empathetic preferences. When Alice *empathizes* with Bob, she does not necessarily identify with him so closely that she ceases to separate her own preferences from his. We weep, for example, with Romeo when he believes Juliet to be dead. We understand why he takes his own life—but we feel no particular inclination to join him in the act. Similarly, a confidence trickster is unlikely to sympathize with

his victims, but he will be very much more effective at extracting money from them if he is able to put himself in their shoes with a view to predicting how they will respond to his overtures. I think we unconsciously carry out such feats of empathetic identification on a routine basis when playing the game of life each day with our fellow citizens.

It seems evident to me that empathetic identification is crucial to the survival of human societies. It provides a tool for predicting the behavior of those around us, and so helps us find our way to equilibria in the games we play that would otherwise require a slow and clumsy process of trial and error. However, it is not enough for the viability of a human society that we be able to use empathetic identification to recognize the equilibria of commonly occurring games. The games we play often have large numbers of equilibria. As Hume [14] saw several hundred years ago and Schelling [27] and Lewis [16] have emphasized in modern times, society therefore needs commonly understood coordinating conventions that select a particular equilibrium when many are available. Sometimes the conventions that have evolved are essentially arbitrary, as in the case of the side of the road on which we drive. However, in circumstances that are more deeply rooted in our social history, we usually overlook the conventional nature of our equilibrium selection criteria. We internalize the criteria so successfully that we fail to notice that selection criteria are in use at all.

I think we are particularly prone to such sleepwalking when using those conventional rules that we justify with airy references to "fairness" when asked to explain our behavior. In saying this, I do not have in mind the rhetorical appeals to fairness that typify wage negotiations or debates over taxation. Nor do I have in mind the abstract notions of justice proposed by philosophers like Rawls [24]. I am thinking rather of the give-and-take of ordinary life. Who should wash the dishes tonight? Who ought to buy the next round of drinks? How long is it reasonable to allow a bore to monopolize the conversation over the dinner table? We are largely unconscious of the fairness criteria we use to resolve such questions, but the degree of consensus that we achieve in so doing is really quite remarkable.

Many attempts have been made to model the fairness norms we use to solve the equilibrium selection problems posed by the multitude of coordination games of which our daily social life largely consists. When such fairness norms take account of the fact that different people have different preferences,[11] they necessarily require some degree of empathetic identification.

The leading proposal for such a fairness norm is Rawls' [24] device of the original position, which was independently proposed by Harsanyi [13] at around the same time. To work out what is fair using this device, Alice and Bob should imagine the bargaining outcome they would reach if their identities were concealed behind a "veil of ignorance" that gave neither any more reason than the other to think themselves Alice or Bob. Both Rawls and Harsanyi offer a Kantian defense of the

[11] So that Alice needs to qualify the Golden Rule—that she should do unto Bob as she would that he do unto her—by accepting that his preferences may not be same as hers.

original position, but my guess is that the real reason the original position appeals so strongly to our intuition is that, in working through its implications, we recognize that it epitomizes the basic principle that underlies the fairness criteria that have evolved to adjudicate our day-to-day interactions with our fellows.

A fairness norm like the original position would obviously be worthless if one person were to imagine how it would feel to be another person without substituting the other person's personal preferences for their own. But more than this is necessary. In order to make fairness judgments, people must be able to say *how much* better or worse they feel when identifying with others. Empathetic identification by itself is not sufficient for this purpose. An essential prerequisite for the use of a device like the original position is that we be equipped with the empathetic *preferences* introduced by Suppes [30].

It seems uncontentious that we actually do have empathetic preferences that we reveal when we make fairness judgments. Ordinary folk are doubtless less than consistent in the empathetic preferences they reveal, but Harsanyi [13] idealized the situation by taking *homo economicus* as his model of man. In his model, everybody therefore has consistent empathetic preferences, which Harsanyi takes to mean that they satisfy the Von Neumann and Morgenstern rationality postulates. An empathetic preference can then be described using a Von Neumann and Morgenstern utility function.

An orthodox personal utility function of the kind we have considered hitherto simply assigns a utility to each situation that the person in question might encounter. For an empathetic utility function we have to pair up each such situation with the person whom we are considering in that situation. One such pair might consist of Alice drinking a cup of coffee. Another might be Bob eating a doughnut. An empathetic utility function assigns a utility to each such pair. It is, of course, precisely such pairs of possibilities that must be evaluated when people imagine themselves in the original position behind a veil of ignorance that hypothetically conceals their identity. That is to say, in the case of only two people, the set of outcomes that need to be evaluated in the original position is a bunch of possibilities of which one might be that Alice ends up with a cup of coffee and Bob with a doughnut.

The next step in Harsanyi's argument is another idealization. He assumes that when people empathize with Alice or Bob, they do so entirely successfully. More precisely, if Bob is totally successful in empathizing with Alice, then the preferences he will express when imagining myself in Alice's position will be identical to Alice's own personal preferences.

The rest of Harsanyi's argument is a straightforward application of the properties of Von Neumann and Morgenstern utility functions. The property that matters here is that any two Von Neumann and Morgenstern utility scales that represent exactly the same preferences over risky alternatives must be related in the same way as two temperature scales. That is to say, the two scales can differ only in the placing of their zero and their unit. For example, once one knows the number of degrees that the Centigrade and Fahrenheit scales assign to the freezing and boiling points

of water, then one knows how to translate any temperature on one scale into the corresponding temperature on the other scale.

The two utility scales to which this fact is now applied are Alice's personal scale and an observer's empathetic scale for Alice. Since Harsanyi's second assumption implies that both scales represent the same preferences, the observer's empathetic scale for Alice is exactly the same as her personal scale except that the zero and the unit are changed. In particular, a util on the observer's empathetic scale for Alice is obtained by multiplying a util on her personal scale by some constant number a. Similarly, a util on the observer's empathetic scale for Bob is obtained by multiplying a util on her personal scale by some constant number b.

It follows that the observer's empathetic utility function simply expresses the fact that the observer thinks that Alice and Bob's personal utils can be traded off so that a of Alice's personal utils count the same as b of Bob's personal utils. That is to say, Harsanyi's assumptions imply that holding an empathetic preference is exactly the same thing as subscribing to a standard for making interpersonal comparisons.

Utilitarianism? One slots this theory of interpersonal comparison into the standard theory of revealed preference by assuming that our observer makes fairness judgements in a consistent way, and that we know enough of these fairness judgments to be able to summarize all of them in a single empathetic utility function. Harsanyi [13] continues by postulating an Ideal Observer (like that of Adam Smith), whose judgements are assumed to be "morally binding". Since the Ideal Observer's empathetic utility function satisfies the Von Neumann and Morgenstern postulates, the Ideal Observer makes decisions by maximizing the average value of this function. Harsanyi thereby provides the kind of formal defense of utilitarianism that is usually attributed to John Stuart Mill, but is nowhere to be found in his work.

Evolutionary ethics. In my own work (Binmore [3]), I follow Harsanyi's theory of interpersonal comparison, but I balk at the introduction of an Ideal Observer, whose hypothetical fairness judgements are somehow "morally binding". I prefer Harsanyi's second defense of utilitarianism, in which he joins Rawls in replacing his Ideal Observer by the device of the original position. But my impatience with metaphysical defenses of ethical concepts leads me to continue by asking why we should find the notion of the original position intuitive in the first place. And, given that the original position is to be used, how come we all employ the $same$ standard of interpersonal comparison?

My evolutionary answers to these questions do not find favor with orthodox moral philosophers, but the arguments I offer suggest that the fairness norms which actually work in practice originally evolved as equilibrium selection devices in the game of life that our ancestors played on a daily basis in the far past. If I am right, fairness norms share with language the property that their deep structure is universal in the human species, but their surface structure differs from one society to another. Both biological and cultural evolution must therefore be taken into account. I go on

to offer reasons why the deep structure of fairness norms should coincide with the original position, but argue that the standard of interpersonal comparison needed to operationalize the original position should be expected to vary not only with culture but also with context.

Within such a naturalistic framework, nothing binds anybody to anything, and so Harsanyi's defense of utilitarianism fails.[12] One is led instead to an alternative concept from cooperative game theory called the egalitarian or proportional bargaining solution, which might be said to formalize Aristotle's observation that "What is fair . . . is what is proportional".

4 Conclusion

Patrick Suppes was a scholar like Epicurus or Hume, who delighted in being part of a circle of philosophers in which no subjects were taboo—anything could be discussed without anybody getting upset or feeling patronized. This article looks only at two rather special contributions he made to game theory, but even here his disregard for academic fashion when on the track of something interesting is evident. Nor did age curb his enthusiasm or blunt his joy in a good argument. I hope this symposium will give some feeling, not only of the wide breadth of his interests, but of the pleasure those who knew him took in his company.

References

[1] K. Arrow. Extended sympathy and the problem of social choice. *Philosophia*, 7:233–237, 1978.

[2] J. Bentham. An introduction to the principles of morals and legislation. In *Utilitarianism and Other Essays*. Penguin, Harmondsworth, UK, 1987. (Introduction by A. Ryan. Essay first published 1789).

[3] K. Binmore. *Natural Justice*. Oxford University Press, New York, 2005.

[4] K. Binmore. *Does Game Theory Work? The Bargaining Challenge*. MIT Press, Cambridge, MA, 2007.

[5] K. Binmore, J. Swierzbinski, and C. Proulx. Does minimax work? An experimental study. *Economic Journal*, 111:445–465, 2001.

[6] J. Brown and R. Rosenthal. Testing the minimax hypothesis: A re-examination of O'Neill's game experiment. *Econometrica*, 58:1065–1081, 1990.

[12] Without any need to follow Rawls in his iconoclastic denial of orthodox Bayesian decision theory. However, once governments appear on the scene as enforcement agencies, Harsanyi's arguments for utilitarianism can be revived.

[7] R. Bush and F. Mosteller. *Stochastic Models for Learning*. Wiley, New York, 1955.

[8] W. Estes. Of models and men. *American Psychologist*, 12:609–617, 1957.

[9] O. Frenkel. A study of 78 non-iterated 2×2 games. Staff Report 176, Federal Reserve Bank of Minneapolis, 1973.

[10] P. Hammond. Why ethical measures of inequality need interpersonal comparisons. *Theory and Decision*, 7:263–274, 1976.

[11] R. Hardin. *Morality within the Limits of Reason*. University of Chicago Press, Chicago, 1988.

[12] J. Harsanyi. Cardinal welfare, individualistic ethics, and the interpersonal comparison of utility. *Journal of Political Economy*, 63:309–321, 1955.

[13] J. Harsanyi. *Rational Behavior and Bargaining Equilibrium in Games and Social Situations*. Cambridge University Press, Cambridge, 1977.

[14] D. Hume. *A Treatise of Human Nature (Second Edition)*. Clarendon Press, Oxford, 1978. (Edited by L. A. Selby-Bigge. Revised by P. Nidditch. First published 1739.).

[15] J. Ledyard. Public goods: A survey of experimental research. In J. Kagel and A. Roth, editors, *Handbook of Experimental Economics*. Princeton University Press, Princeton, 1995.

[16] D. Lewis. *Conventions: A Philosophical Study*. Harvard University Press, Cambridge, MA, 1969.

[17] D. Malcolm and B. Lieberman. The behavior of responsive individuals playing a two-person zero-sum game. *Psychonomic Science*, 2:373–374, 1965.

[18] K. McCabe, A. Mukherji, and D. Runcie. An experimental study of learning and limited information in games. Staff Report 176, Federal Reserve Bank of Minneapolis, 1994.

[19] D. Mookherjee and B. Sopher. An experimental study of learning and limited information in constant-sum games. Technical report, Rutgers University, 1994.

[20] D. Mookherjee and B. Sopher. Learning and decision costs in experimental constant-sum games. *Games and Economic Behavior*, 19, 1997.

[21] B. O'Neill. Nonmetric test of the minimax theory of two-person zero-sum games. *Proceedings of the National Academy of Sciences*, 84:2106–2109, 1987.

[22] G. Pollock, A. Cabrales, S. Rissing, and K. Binmore. Suicidal punishment in the ant *Acromyrmex Versicolor*. *Evolutionary Ecological Research*, 14:1–21, 2012.

[23] A. Rapaport and D. Bludescu. Generation of random series in two-person strictly competitive games. *Journal of Experimental Psychology*, 121:352–363, 1992.

[24] J. Rawls. *A Theory of Justice*. Oxford University Press, Oxford, 1972.

[25] L. Robbins. Inter-personal comparisons of utility. *Economic Journal*, 48:635–641, 1938.

[26] A. Roth. Bibliography of learning in games: Theory, experiment, and related subjects. web.stanford.edu/‘alroth/learnbib.html.

[27] T. Schelling. *The Strategy of Conflict*. Harvard University Press, Cambridge, MA, 1960.

[28] A. Sen. *Collective Choice and Social Welfare*. Holden Day, San Francisco, 1970.

[29] A. Smith. *The Theory of Moral Sentiments*. Clarendon Press, Oxford, 1975. (Edited by D. Raphael and A. Macfie. First published 1759).

[30] P. Suppes. Some formal models of grading principles. *Synthèse*, 6:284–306, 1966.

[31] P. Suppes and R. Atkinson. An analysis of two-person game situations in terms of statistical learning theory. *Journal of Experimental Psychology*, 55:369–378, 1958.

[32] P. Suppes and R. Atkinson. *Applications of a Markov Model to Multiperson Interactions*. Stanford University Press, Stanford, 1960.

[33] J. Von Neumann. Zur Theorie der Gesellschaftsspiele. *Mathematische Annalen*, 100:295–320, 1928.

[34] J. Von Neumann and O. Morgenstern. *The Theory of Games and Economic Behavior*. Princeton University Press, Princeton, 1944.