The conception of the individual in non-cooperative game theory

John B. Davis
University of Amsterdam and Marquette University

Abstract

This paper examines the conception of individuals as being of certain types in Harsanyi's transformation of games of incomplete information into games of complete information. It argues that while the conception of the individual in games of complete information offers potential advances over the problematic neoclassical conception of the individual, Harsanyi's more realistic incomplete information games framework essentially re-introduces the difficulties from the neoclassical conception. A further argument of the paper is that fixed point equilibrium existence proof theorems and individual existence proofs function in an analogous manner, and can consequently been seen as both dependent upon one another. Thus the inadequacy of Harsanyi's conception of individuals raises questions about Nash equilibrium approaches to equilibria in games.

Keywords conceptions of individuals, Harsanyi, games of incomplete information, fixed point theorems, Nash equilibrium

This paper investigates the conception of the individual in non-cooperative, one-off or single play game theory. Previously I examined the conception of the individual in neoclassical economics (Davis, 2003). The conclusion of that analysis was that the neoclassical conception fails two basic identity tests required of any conception of the individual. The conception of the individual in non-cooperative single play game theory is similar to that of neoclassical theory in that in both theories individuals are highly isolated. But game theory's conception of the individual also differs from the neoclassical one in its treatment of individuals as interactive. The non-cooperative game theory conception in single play games, as the closest kin to the methodological individualist neoclassical conception, may thus be seen as both an extension and a redevelopment of the traditional neoclassical conception. The motivation for this paper, then, is to investigate whether non-cooperative, single play game theory produces a new, viable conception of the

individual for the linked neoclassical and non-cooperative game theory research programmes. To do this I introduce a new set of perspectives on the Nash equilibrium concept and its refinements in terms of the so-called 'Harsanyi doctrine' and its Aumann extension that generate Bayesian (Nash) equilibria in single play, non-cooperative games. Much of the attention devoted to this research programme focuses on such issues as the plausibility of mixed strategies and problems of multiple equilibria. I put these sorts of issues aside, however, to rather consider how the core ideas of the Nash-Harsanyi-Aumann equilibrium programme are interrelated with assumptions about the nature of individuals, particularly in connection with John Harsanyi's view of individuals as being of certain types, as developed in his transformation of games of incomplete information into complete information games. This paper consequently does not investigate non-cooperative repeated and cooperative games, because I believe they often rely upon additional assumptions about individuals that go beyond or are at odds with the basic Nash view, and because it seems that appraising these additional assumptions is best done in light of an understanding of the conception of the individual in the basic, non-cooperative play research programme.[1] It is this conception, in other words, that most clearly develops the traditional methodological individualist view, and thus is the natural starting point for an examination of the individual in game theory.

Section one begins by comparing the logic of mathematical fixed point theorems used in equilibrium existence proofs in neoclassical economics and non-cooperative game theory to the logic used in philosophical personal identity arguments, in order to elicit the main features of the conception of the individual employed in both research programmes. Section two reviews and comments on the difficulties that arise in connection with explaining individual identity in terms of preferences in the neoclassical conception of the individual. Section three compares and distinguishes the game theory conception of the individual in games of complete information. Section four examines the conception of the individual in games of incomplete information, where this involves the Harsanyi method for transforming incomplete information games into complete information games, explains players as being of certain types, and employs the common priors assumption – the so-called Harsanyi doctrine – and its Aumann extension in the elaboration of Bayesian (Nash) equilibria. Section five evaluates the incomplete information game conception of the individual as being of a type relative to the conception of particular individuals employed in complete information games. Section six returns to the themes of the first section to

[1] In repeated game, for example, the role of standards of fairness cannot be ignored. The seminal paper on the subject is Kahneman, Knetsch, and Thaler (1986). See Mirowski (2002) for the divide between non-cooperative and cooperative game theory and John Nash's role in creating it.

re-consider the relationship between equilibrium existence proofs and the view of individuals in single play non-cooperative game theory.

1 Fixed point theorems and personal identity analysis

Since equilibrium concepts are closely intertwined with the conceptions of economic agents they show to be in equilibrium, the particular equilibrium concept used in game theory should provide a key to understanding the particular conception of the agent in game theory. Fixed point theorems constitute the primary method employed in economics for establishing the existence of solutions to equilibrium systems of equations or inequalities (Giocoli, 2003). Brouwer-Kakutani-type fixed point theorems have been used to demonstrate the existence of a set of equilibrium prices for a Walrasian competitive economy and the existence of an equilibrium point of n-tuple strategies in a many-person non-cooperative game.[2] A fixed point theorem is a mathematical proposition which states that a mapping f that transforms each point x of a set X to a point $f(x)$ within X has a fixed point x^* that is transformed to itself, so that $f(x^*) = x^*$. Thus a fixed point theorem demonstrates the existence of some system of relationships by anchoring that system in one self-identical relationship within that system. This is essentially the same logic employed in philosophical personal identity analysis to establish the existence of a person in terms of one unchanged or self-identical characteristic amidst change in other characteristics of the person. The mapping f that transforms each point x of a set X to a point $f(x)$ within X can be understood as some process of change in the person understood in terms of a set of characteristics X. The fixed point x^* that is transformed to itself, so that $f(x^*) = x^*$, is that unchanging or self-identical characteristic of the person that allows us to say that the person exists. From this perspective, philosophical personal identity analysis is a form of existence analysis, and philosophical theories that aim to demonstrate personal identity aim to demonstrate the existence of the person. Two conclusions may be drawn from this connection.

First, if we are to say, based on the application of fixed point theorems, that an economy exists when represented as a system of equations, then extending that same fixed point logic to the personal identity of economic agents requires that we also demonstrate that the agents in that economy exist when represented in terms of collections of characteristics. That is, economic equilibrium depends not just upon the existence of a set of equilibrium prices or the existence of an equilibrium point of n-tuple strategies, but also on the existence of market agents. In the standard neoclassical conception of the individual, change exists in one set of characteristics as individual endowments are transformed into the individual's commodities. But an individual's

[2] See Leonard (1992) for John von Neumann's original application of fixed point methods to games.

unchanging preferences can be construed as that characteristic which acts as a self-identical fixed point assuring the individual's existence. Thus in principle the market economy and the individuals who populate it can both be said to exist for the same kind of reason. I return to this argument in Section 6.

Second, the comparison between fixed point theorems and personal identity analysis offers an important insight into the latter for economics. To say that a fixed point x^* is transformed to itself, or that $f(x^*) = x^*$, is to characterize x^* reflexively. That is, in all transformations, x^* always reproduces itself and only itself. Thus a personal identity analysis understood in fixed point terms would explain the existence of the person or individual in terms of one specific type of characteristic, namely, one exhibiting reflexivity. That is, what would be unchanging about individual economic agents amidst change in other characteristics is their being able to take themselves as an object. Much existing personal identity analysis, however, ignores reflexivity, and simply focuses on individual characteristics which might be thought to be unchanging, for example, such as that a person can never have someone else's body. Applying fixed point thinking to the identity of individual economic agents, then, leads us to interpret the characteristic of unchangingness specifically in reflexive terms. Below I will argue that both the neoclassical and game theory conceptions of the individual seek to establish the existence of the individual reflexively, but differ in the ways in which they interpret this reflexivity. However, together – whether successful or not – they take a different approach to personal identity than is typically found in philosophical investigations.

2 The neoclassical conception of the individual

The role of reflexivity in the neoclassical conception of the individual can be introduced by comparison with John Locke's view of the individual, which I have previously argued was a model for the neoclassical conception (Davis, 2003).[3] Locke, who is generally agreed to have had the first account of personal identity, asserted the following:

For it is by the consciousness it has of its present thoughts and actions, that it is a *self to itself* now, and so will be the same self, as far as the same consciousness can extend to actions past and to come. [Locke, 1975 [1694], 2.27.10]

[3] There are close affinities between the neoclassical conception of the individual and Hume's thinking, but Hume rejected the idea of personal identity, or that there is a subject, whereas neoclassical economics treats individuals as unitary agents.

That is he explained the self as a single entity by claiming that one recognized oneself in one's own thoughts. One's thoughts, in effect, were transparently one's *own* thoughts. To make his case Locke used memory to unite the different episodes of consciousness. When remembering a past experience, one remembered it explicitly as one's own past experience. This uniting of past and present experiences gave the self the unity required to justify personal identity.

However, it was quickly pointed out by Locke's most famous critic, Bishop Butler (1896 [1736]), that Locke's argument was circular since it presupposed the self in order to prove the existence of the self. That Locke's observed his *own* consciousness and *own* memories meant that he assumed the existence of what he intended to demonstrate. Contemporary philosophers accept that Locke's argument failed, but have not necessarily abandoned having a single consciousness as a possible criterion of personal identity. Those that have considered this criterion, however, have taken the circularity problem seriously, and suggested other methods for identifying a single consciousness besides simple self-inspection. I suggest, however, that the problem lies not in reflexivity but rather in an exclusively subjectivist understanding of it, since as we will see in the next section the Harsanyi refinement of Nash equilibrium employs reflexivity across individuals to produce a conception of the individual.

The neoclassical conception of the individual replaces Locke's contents of consciousness with individual preferences. Preferences are understood to have a structure whose axiomatic representation justifies the ascription of utility functions to individuals. Having a single utility function not only provides an identity to the individual agent, but also disciplines the individual's choices so as to make individuals countable, and allow aggregation procedures across individuals. But there is a subtle step passed over in this account that directly recalls Locke. Strictly speaking, there is nothing a in a given set of preferences that makes them necessarily an individual's *own* preferences. Preferences do not come with someone's name on them – unless one has presupposed that preferences must always be 'own preferences.' This may be tempting to assume on the surface, but quickly becomes less persuasive when one begins to think about social influences on tastes. The point is that, as in Locke's argument, giving an account of the individual in terms of own experiences is circular, and thus the neoclassical conception does not succeed in explaining the individual as an individual.[4] At the same time, as with Locke, the criterion purported to explain the individual is a reflexive one. Individuals are thought to be individuals in virtue of their ability to appraise themselves – here, recognize their preferences. How, then, is game theory's conception of the individual any different?

[4] I extend this argument to the more sophisticated neoclassical human capital model in Davis (2003).

3 The conception of the individual in games of complete information

The game theory conception of the individual shares the utility-based view of the individual with the neoclassical conception. More accurately, game theory shares expected utility reasoning with the neoclassical conception, though it goes beyond the latter's focus on lotteries and choice under risk with their relatively well understood (though not unproblematic) principles to emphasize the additional complications created by the individual's interactions with others as a source of uncertainty in the world. What does this involve? Since Harsanyi (1967/1968, 1995) argues that games of incomplete information can always be transformed into games of complete information,[5] let us begin with these as more basic. In the first place, then, players in a game need to know the rules of the game being played (what kind of game it is) as well as the utility payoffs for the other players. In the second place, players need to know the basis on which other players select their actions, namely that they act rationally rather than, say, habitually. This latter requirement is the common knowledge of rationality assumption regarded as indispensable to all of game theory.[6] As Robert Aumann describes it:

It is not enough that each player be fully aware of the rules of the game and the utility functions of the players. Each player must also be aware of this fact, i.e., of the awareness of all the players; moreover, each player must be aware that each player is aware that each player is aware, and so on ad infinitum. In brief, the awareness of the description of the game by all players must be a part of the description itself (Aumann, 1989, p. 473).

Note the fundamental departure from standard neoclassical theory that this involves. Neoclassical theory sidesteps the interactive, game-theoretic aspects of economic behavior by postulating perfect competition and price-taking behavior, so that individuals need not be aware of what other players are aware of, nor aware that other players are aware of what they are aware of, and so on. This means that individuals need only consult their own preferences and prices in deciding what to do. In contrast, in interactive settings individuals need to know the rules of the game including the utility payoffs for both themselves and all the other players in the game, and they need to know this *per* the description above. Thus ultimately, whereas in the standard

[5] Strictly speaking, he produces games that function like or as if they were complete information games, since they are still games in which "players have less than full information about each other's *payoff functions.*" Or, "the new game G^* will be one with *complete* information because its basic mathematical structure will be defined by the probabilistic model for the game, which will be fully known to both players." (Harsanyi, 1995, pp. 293, 295).

[6] First given clear formulation by Lewis (1969).

neoclassical model individuals need know only their own preferences, in the basic game-theoretic setting everyone needs to know everyone's preferences!

That this be the case depends on the common knowledge requirement. The common knowledge requirement might seem to constitute a barrier to any coherent account of games, since the regress it potentially implies appears to make an impossible demand on what individuals can know. Nash, however, provided an escape from the regress problem in his formulation of an game theoretic equilibrium concept of strategies taken as individuals' best replies to one another. To say that all individuals in a game make only their best replies to one another means that only rational or rather rationalizable strategies are implemented. Rationalizable strategies are those which confirm the expectations of each player about each other player's choice, and may thus also be termed self-confirming strategies. Their effect is to incorporate the pattern of reflexivity and interaction exhibited in a game's common knowledge structure into the decision-making of each player of a game. Individuals then know their own payoffs and strategies, which are known by others, whose own payoffs and strategies are known in turn, which are also known by others, etc., and all this enters into the decision-making of each individual. Thus, Nash's disciplining of the common knowledge assumption makes possible a significant redevelopment of the standard neoclassical view of the individual, since it makes what are uniquely private preferences on that view into publicly known preferences, albeit ones that still attach to separate individuals.

Like the traditional neoclassical conception, the game theory conception of the individual is still framed reflexively, because like that conception it characterizes individuals as beings who consult their own preferences in deciding what to do. But clearly the form of reflexivity in the game theory conception is different, since each individual's preferences are not only identified by each individual as their own, but are now also identified by everyone else as their own as well. Does this represent a sufficient enough departure from the neoclassical conception to escape the circularity problem described in the previous section? Locke's (reflexive) account of the individual – which is the model for the neoclassical conception – is circular, because individuals can only identify themselves in terms of their experiences by presupposing that those experiences are their own experiences. The problem here, however, is not the simple fact of circularity, since circularity can be self-defeating or benign (Davis and Klaes, 2003). Rather the problem seems to be that the individual lacks the authority to claim observed experiences are own experiences, because no one else is in a position to affirm or dispute this association. Thus suppose that another individual could somehow affirm that all an individual's experiences claimed to be own experiences were indeed that particular individual's own experiences. Then it might be argued that individuals can be defined as collections of experiences, and the circularity problem

consequently dismissed. The individual would still be defined reflexively in terms of own experiences, but this reflexivity would not be circular, or constitute a self-defeating reflexivity.

This argument may be transferred to the game theory emendation of the neoclassical conception of the individual. Whereas the neoclassical conception is circular in a self-defeating sense, the game theory conception of the individual appears to escape this problem by making each individual's preferences known by all other individuals. Single individuals lack the authority to recognize observed preferences as own preferences, but individuals' preferences can nonetheless be seen to be their own preferences in a common knowledge, Nash equilibrium-type game setting. That is, when game theory transforms uniquely private preferences into publicly known preferences, it introduces what might be called a game-objective account of the individual, where neoclassicism had previously employed what might be called a preference-subjective account of the individual. This revised conception of the individual appears to meet the objections that apply to the earlier conception, then, by recasting the subjectivist character of the traditional view in new terms. In effect, incorporating the pattern of reflexivity and interaction exhibited in a game's common knowledge structure into the decision-making of each player of a game builds up a structure of individual preferences constituted out of all players' preferences that recognize and distinguish each individual player's preferences. Individual preferences are still subjective, but they are now couched in an intersubjective framework of interactive games.

Games understood in this way, however, are complete information games. In particular, they require that all individuals know all individuals' payoffs. But in incomplete information games players may lack information about other players' and even their own payoff functions. Since clearly games of incomplete information constitute the more realistic sort of case, complete information games may be regarded as an idealization in the sense that this is how games *would* be understood *were* certain conditions regarding information to be satisfied. I approach the analysis of these conditions by way of the thinking involved in the equilibrium conception employed in game theory, which relies on fixed point logic and indirect proof method (Giocoli, 2003, pp. 20ff). The indirect proof method (IPM) is a non-constructive form of argument that assumes that an object whose existence is to be demonstrated does not exist, and then proceeds by showing that this assumption leads to a contradiction.[7] If we apply this IPM form of reasoning to the complete information game theory account of the individual above, we would assume that individuals thus understood do not exist, and then proceed by demonstrating that this assumption

[7] In contrast, a constructive proof "outlines a procedure or algorithm leading to the mathematical object whose existence we aim to assert. In other words, it is a demonstration technique based upon the 'calculability' of the object under scrutiny" (Giocoli, 2003, p. 21).

leads to a contradiction. That is, we would assume that individuals rather play games of incomplete information, and then ask how this produces a contradiction, thereby implying that individuals *do* play games of complete information. Harsanyi's well-known arguments provide the template for this examination in that he shows how games of incomplete information can function as if they were games of complete information. The way that he does this is to identify various conditions involved in individuals being able to classify other individuals as being of certain types. Satisfaction of these conditions not only transforms games of incomplete information into games of complete information, but also in principle creates a basis for treating the complete information account of the individual as a viable idealization. In the following section these conditions are identified. I return to the larger argument returning IPM reasoning in the last section.

4 The conception of the individual in games of incomplete information

Harsanyi's arguments regarding how games of incomplete information function as if they were games of complete information simultaneously solved two problems. One was that most games are presumably games of incomplete information, but the Nash framework required complete information. The other involved the rationale for mixed strategies.[8] Nash equilibria always exist in finite games when mixed strategies are allowed, but are absent in a large number of games in pure strategies. At the same time, mixed strategies lack plausibility, both because they do not seem to be common in real world decision making, and because it can be shown that "a player will not lose if he abandons the randomization and uses instead any arbitrary one of the pure strategy components of the randomization" (Aumann, 1985, p. 44). In Harsanyi's (1967/1968) formulation of incomplete games, however, a player employs pure strategies, but appears to other players to be using mixed strategies. These apparent mixed strategies are other players' best guesses or conjectures as to which pure strategy a player is playing. This device made it possible to suppose that the Nash framework had wide application, since it allowed one to say that players play pure strategies, but that games could nonetheless still be analyzed as if they played mixed strategies.

Harsanyi used his analysis to show that players might regard each other as being of certain types. Before turning to that account, however, we should note the two conditions under which his formulation holds: there must be common knowledge of players' conjectures, and players must have common priors on the set of the states of the world.[9] The common knowledge

[8] Mixed strategies involve players playing a number of pure strategies with certain probabilities.
[9] This latter condition was labeled 'Harsanyi doctrine' by Aumann (1976).

of players' conjectures condition follows from the common knowledge assumption. The common priors condition means that individuals or players share the same probability distribution over all possible states of the world. Whereas individuals differ in their preferences, "probabilities reflect information, so that prior to information being received, probabilities should be the same" (Rizvi, 1994, p. 18; cf. Aumann, 1987, pp. 13-14). Once information is received, probabilities can then differ, such that in principle informational differences would be important to explaining the outcomes of games. However, Aumann's (1976) extension of the 'Harsanyi doctrine' eliminates informational differences by arguing that if individuals begin with common priors, and if their beliefs concerning any given event are common knowledge, then their posterior probabilities concerning that event must be the same, since rational individuals would revise their beliefs through Bayesian updating when faced with informational differences.[10] Common knowledge thus trumps asymmetric information, so that, as Aumann famously put it, players cannot agree to disagree.

There have been many questions raised about this analysis, but I rather focus on the view of individuals as being of certain types that it enables Harsanyi to elaborate, thus treating the view's main assumptions – Harsanyi's initial two points and Aumann's extension – as conditions that need to be satisfied should an IPM-type reasoning be used to tell us that we may understand individuals as types. Harsanyi's focus on types of individuals has become more central to his thinking over time. In his more recent work, he begins by saying that while a complete information game (or C-game) is always analyzed on the assumption that "the centers of activity" are players or individuals, incomplete information games (or I-games) are more clearly formulated as having types of players as their "centers of activity" (Harsanyi, 1995, p. 295). At the same time, in I-games the player-centered representation and type-centered representation are taken as being ultimately equivalent, and in important expositions of his thinking on the subject (e.g., Harsanyi, 1995, pp. 298-9), Harsanyi relies on the player-centered language to provide what he regards as the more intuitive understanding of his argument. Are these two forms of representation, then, truly equivalent?

The means by which Harsanyi treats them as such is his reinterpretation of games of incomplete information in accordance with the probabilistic model (that is, by adding a lottery to the game), such that facts about players not known to all players are replaced by probability assessments regarding players' characteristics that are known to all players. Players are then represented as types because they may be represented in terms of certain sets of characteristics. Broadly speaking, they may be represented in terms of certain characteristics and as being certain

[10] See Hargreaves Heap and Varoufakis (1995, p. 26) for a brief summary of Aumann's argument.

types, because of "causal factors" or "social forces" in the world that determine what characteristics different individuals are likely to possess (Harsanyi, 1995, p. 297).[11] In particular games, then, before any moves are made, players estimate the probabilities that other players have certain characteristics (just in the way as would any outside observer), and act on the assumption that every other player will estimate these probabilities much in the same way (the common priors assumption). Harsanyi also assumes that players know which type they themselves each represent – "know their own identities" (*Ibid*., p. 296) – and rely on this information to assess the probability that other players are of certain types. This makes each player's assessment that another player is of a certain type a conditional probability assessment, one conditional, that is, upon knowing one's own type. All players make such assessments, and consequently any given player (player 1) will act "so as to protect his interests not only against his unknown *actual* opponent … but … against all M types of player 2 because, for all he knows, *any* of them could now be his opponent in the game" (*Ibid.*, p. 299). Thus, each player's expected payoff depends not just on the strategy of the actual unknown opponent, but also on the strategies of any one of M potential opponents.[12] Then, regarding types of players as "the real 'players'" and their payoff functions as the "real payoff functions, one can easily define the *Nash equilibrium*… of this C-game G^* (*Ibid.*, p. 300).

Aumann's extension of the Harsanyi doctrine does not add anything to the latter's theory of types, but rather lends reinforcement to the view by explaining how individuals would respond to inevitable differences in information about the world. Were informational differences to persist, though players began with common priors, their differing views of the world's "causal factors" and "social forces" would lead to inconsistent probability assessments regarding each other's types. Aumann's argument that individuals would come to have the same information through a Bayesian up-dating process – that their posterior probabilities would be the same – means that they would have end up with the same probability distributions regarding players' characteristics, so that informational differences ultimately do not exist. This effectively removes all discussion of belief revision and learning processes from the analysis so as to restrict the focus to how types of individuals may be in equilibrium when engaging in non-cooperative strategic behavior. In the language of Section 1, individuals of certain types exist, and this can be shown to imply that equilibria in games between them exist. The following section, however, asks whether the change in the view of the individual from complete to incomplete information games

[11] Drawing on game theory applications to the Cold War, Harsanyi's example distinguishes American and Russian types whose causal factors pertain to their locations in the United States and the Soviet Union.
[12] These are labeled semiconditional payoff functions.

sustains or undermines the idea developed above for complete information games that individuals can be understood in terms of preferences.

5 Difficulties in the conception of the individual in games of incomplete information

Critics of the Harsanyi-Aumann view (e.g., Rizvi, 1994, pp. 15-20; Hargreaves Heap and Varoufakis, 1995, pp. 25ff) have argued that its three conditions – common knowledge of players' conjectures, common priors, and Bayesian updating – are implausible on a variety of epistemic grounds. Their questions are important, but my focus is rather on how the complete information conception of the individual set out in Section 3 is transformed by the Harsanyi-Aumann view of the individual as being of a certain a type in games of incomplete information. The complete information conception was labeled a game-objective account of the individual to contrast it with the traditional neoclassical account labeled a preference-subjective conception of the individual. I suggested that this game-objective conception constituted an advance on the preference-subjective conception, because with a game's common knowledge structure each individual player's preferences are recognized as their own by all, whereas in the traditional neoclassical conception is asymmetrical in this regard in that individuals only recognize their own preferences. In effect, individual subjectivity is made objective in games of complete information when everyone knows everyone else's preferences, and this arguably removes the self-defeating circularity present in the traditional neoclassical account of the individual as a set of preferences.

Note, then, that in games of incomplete information everyone does not know everyone else's preferences, and that the old neoclassical asymmetry between what individuals know is essentially reintroduced with the assumption that each individual alone knows their own type. Whereas other players know that an individual is one of M types, each individual knows specifically which of the M types they are. With respect to everyone but themselves, players can only say that a given player is of a certain type with a particular probability, is of another certain type with another particular probability, and so on. Alternatively, with respect to everyone but themselves, players can only ascribe one set of preferences to a given player with one probability, another set of preferences to that player with another probability, and so on. Might we, however, finesse this asymmetry between what players know about themselves and others by emphasizing the overall probabilistic framework? Consistency would seem to imply that we should say that each player knows what type they are or what preferences they themselves have with a probability of one, while they know what type other players are or what preferences others have with a probability of less than one. Certain knowledge, on this view, is just a limiting case of uncertain knowledge, and is not a fundamentally different kind of knowledge. Looking at the

matter this way would then seem to undercut the suggestion that individuals have special self-access to themselves different in kind from their access to others. We might argue on this basis that just as in complete information games everyone knows everyone else's preferences – thereby validating the view of individuals as sets of preferences – so in incomplete information games everyone would know everyone else's preferences in a probabilistic sense – thereby again validating the view that individuals may be identified as sets of preferences.

Unfortunately, this interpretation of the probabilistic framework is questionable. While one can indeed assert that players know their own types and preferences with a probability of one, the character of their knowledge about themselves is still different in character from their knowledge of others. When we say that certain states of the world have particular probabilities, we suppose this to be contingently true. But it is not contingently true in the Harsanyi-Aumann model that individuals' knowledge of their own states has a probability of one, since the analysis depends entirely on the conditional probability formulation of individuals' expected payoff functions that assumes that each individual must know their own type. Basically, as in the traditional neoclassical conception, individuals cannot *not* know their own types, because knowing their own types is a matter of knowing their own payoff functions, and this is a matter of knowing their own preferences. Harsanyi thus refers to utility payoffs and payoff functions, and unhesitatingly asserts that players "know their own identities" (Harsanyi, p. 296). Thus the incomplete information game framework really restores the old neoclassical asymmetry, and it is not the case, as in complete information games, that individuals can be validated as sets of preferences by way of everyone knowing everyone else's preferences – at least in the way that they know their own preferences.

It might be replied to this that Harsanyi's theory could be reinterpreted to eliminate all talk of preferences, allowing us to then see individuals as simply certain types, or as sets of characteristics. Then saying that individuals' knowledge of their own types has a probability of one and their knowledge of others' types has a probability of less than one would not be problematic nor reflect any asymmetry between oneself and others, because the characteristics that underlie type identifications reflect states of the world rather than individuals' subjective states. As Aumann essentially points out, utilities are personal, but information is public (Aumann, 1987). This sort of view, moreover, likely reflects the thinking behind many applied uses of incomplete information games, which are typically concerned only with explaining the

outcomes of games between different types of players.[13] Indeed, it might be said that preferences are only paid lip service in the incomplete information game approach, and really have no functional role in an analysis in which the key idea is rather that players know their own types and are able to determine other players' types with a certain probability. From the point of view of the overall development of game theory, given a choice between the realism of incomplete information games and the foundational game theory principle that all individual players must know all individual players' subjective payoffs, the latter is readily given up with the substitution of types of individuals for individuals.

Such a view, however, while reasonable as an understanding of the uses and value of non-cooperative game theory, does not offer an alternative account of individuals, since there is nothing by itself in the idea that individuals have certain characteristics or are the products of "social forces" to explain what makes them individuals. This is consistent with the general argument I have previously advanced (Davis, 2003) that the history of neoclassical economics from cardinal to ordinal utility theory to revealed preference theory involved a progressive elimination of subjectivity in the explanation of choice that had as an inadvertent consequence the elimination of the theory's original (and only) basis for understanding individuals as individuals (namely, as sets of preferences). The evolution of non-cooperative game theory itself, then, replicates this earlier history in that in this part of the Nash equilibrium refinement project the treatment of individuals as subjective basically goes by the wayside. In seeing game theory as a whole as successor research programme to neoclassical theory, then, an inadequate subjectivist account of the individual in neoclassical theory is remedied in early non-cooperative, complete information game theory only at the cost of a lack of realism, which once addressed in games of incomplete information games either fails as an account of the individual or leads to the abandonment of the subjectivist strategy for explaining individuals. Non-cooperative, single play game theory, it must thus be concluded, fails to produce a new, viable conception of the individual for the linked neoclassical and non-cooperative game theory research programmes.

6 Existence of equilibrium and existence of individuals

In the neoclassical and non-cooperative game theory research programmes, demonstrating the existence of equilibrium and demonstrating the existence of individuals can be argued to depend upon one another in that failure to demonstrate the existence of equilibrium casts doubt on the

[13] For example, Fudenberge and Tirole (1989) describe an incomplete information game between a monopoly supplier of an energy source and an monopoly producer of electricity. See Hargreaves Heap and Varoufakis (1995, pp. 62-4) for a summary.

associated understanding of individuals as isolated subjective beings, and failure to demonstrate the existence of individuals makes the notion of an equilibrium between them meaningless. This is not the standard view. The postwar development of neoclassical general equilibrium theory and Nash equilibrium non-cooperative game theory has made demonstration of the existence of equilibrium an exclusive concern, and overlooked the issue of the existence of individuals on the grounds of being 'methodologically' individualist. But taking individuals as given, it seems, only makes sense if individuals can be understood in the same way in and out of equilibrium. The question that arises, that is, is whether in going from an initial distribution of resources to an equilibrium distribution of commodities or payoffs individuals themselves can be shown to exist – in effect have equilibria of their own – just as the system as a whole can be shown to exist in this transformation.

In the neoclassical and non-cooperative game theory research programmes, the general logic by which these two existence proofs should proceed derives from the fixed point approach to demonstrating equilibrium. The parallel existence proof for the individual then employs a reflexivity-based personal identity type of argument, where just as in the equilibrium proof a fixed point x^* is transformed to itself such that $f(x^*) = x^*$, and x^* operates reflexively in the proof, similarly for the individual in all transformations some characteristic always transformed into itself and only itself. In the subjective conception of the individual this characteristic is the individual's unchanging preferences, and the personal identity proof of individual existence requires that the individual reflexively be able to recognize own (unchanging) preferences. Unfortunately, neither the traditional neoclassical conception nor the more realistic incomplete information game theory conception of the individual is successful in providing an adequate proof of personal identity or individual existence, because in both the circularity they involve is self-defeating.[14] The complete information game theory account escapes this charge, but does so at the cost of a lack of realism and by making the subjective objective.

I close, then, with a review of the form of the argument implicit in the Harsanyi strategy for transforming games of incomplete information into games of complete information. My claim at the end of Section 3 was that we may understand this transformation in terms of the IPM, non-constructive form of argument employed in fixed point proofs that works by assuming what one thinks to be false, and then demonstrating that this assumption produces a contradiction. Though used almost exclusively for equilibrium existence proofs in the neoclassical and game theory research programmes on account of their ease of formulation, the IPM does not actually

[14] Reflexive accounts need not be self-defeating. I offer an alternative, non-subjectivist account that uses collective intentionality analysis in Davis (2003).

show something exists – as would constructive proofs which 'calculate' their objects – but rather demonstrates that it is contradictory to suppose they do not. This makes existence proofs sensitive to the conditions they assume, which in the case of the Harsanyi-Aumann account are common knowledge of players' conjectures, common priors, and Bayesian updating. Thus in Harsanyi's argument we assume games are games of incomplete information, and then employ these three conditions to show that individuals can be understood to be of certain types. This produces a contradiction in the sense that these games function as if they were complete information games, enabling us, as Harsanyi says, to "define the *Nash equilibrium*… of this C-game G^* (Harsanyi, p. 300).

What the arguments of this paper imply, however, is that Harsanyi's IPM contradiction has not been demonstrated, because the key conditions that he and Aumann require are tied up with and inseparable from a conception of the individual which fails. In effect, the conditions they require do not hold. In fact, non-cooperative single play game theory only employs an adequate conception of the individual when one assumes from the outset that games are complete information games (as argued in Section 3). But then we retreat from the Harsanyi view of individuals as being of certain types to the more traditional neoclassical subjectivist view of the individuals. Thus the failure to employ a viable conception of the individual undermines the IPM demonstration of the existence of a Nash equilibrium in single play non-cooperative games of any realism, namely, those of incomplete information. These individuals do not exist; therefore Nash equilibria based on their interaction do not exist either.

Non-cooperative single play game theory, then, as an extension and redevelopment of neoclassical thinking does not appear to escape a fundamental limitation of that original framework associated with its account of individuals. Non-cooperative repeated games and cooperative games, however, cannot be easily confined to the view that individuals are highly isolated, essentially subjective beings, and they may offer more promising accounts of individuals in equilibrium settings. Investigation of these accounts, it seems, should begin by asking how they depart from the non-cooperative single play case.

References

Aumann, R. (1976) "Agreeing to disagree," *Annals of Statistics*, 4: 1236-39.

Aumann, R. (1985) "What is game theory trying to accomplish?" in K. Arrow and S. Honkapohja (eds.), *Frontiers of Economics*, Oxford: Blackwell.

Aumann, R. (1987) "Correlated equilibrium as an expression of Bayesian rationality," *Econometrica*, 55: 1-18.

Aumann, R. (1989) "Game theory," in J. Eatwell, M. Milgate, and P. Newman (eds.), *The New Palgrave: a Dictionary of Economics*, London: Macmillan.

Butler, J. (1896 [1736]) *The Analogy of Religion*, in *The Works of Joseph Butler*, W. Gladstone (ed.), 2 vols., Oxford: Clarendon.

Davis, J. (2003) *The Theory of the Individual in Economics*, London: Routledge.

Davis, J. and Klaes, M. (2003) "Reflexivity: Curse or Cure?" *Journal of Economic Methodology*, forthcoming.

Fudenberg, D. and Tirole, J. "Non-cooperative game theory for industrial organisation: an introduction and overview," in R. Schmalensee and R. Willig (eds.), *Handbook of Industrial Organization*, Amsterdam: North Holland.

Giocoli, N. (2003) "Fixing the point: the contribution of early game theory to the tool-box of modern economics," *Journal of Economic Methodology* 10 (1): 1-39.

Hargreaves Heap, S. and Varoufakis, Y. (1995) *Game Theory: A Critical Introduction*, London: Routledge.

Harsanyi, J. (1995) "Games with Incomplete Information," *American Economic Review* 85 (3): 291-303.

Harsanyi, J. (1967/1968) Games with incomplete information played by Bayesian players," *Management Science* 14: 159-82, 320-34, 486-502.

Kahneman, D., Knetsch, J., and Thaler, R. (1986) "Fairness as a Constraint on Profit Seeking: Entitlements in the Market," *American Economic Review*, 76 (4): 728-41.

Leonard, R. (1992) "Creating a Context for Game Theory," in R. Weintraub (ed.), *Toward a History of Game Theory*, Durham: Duke University Press: 29-76.

Lewis, D. (1969) *Convention*, Cambridge, MA: MIT Press.

Locke, J. (1975 [1694]). *An Essay concerning Human Understanding*, ed. P.H. Nidditch, Oxford: Clarendon Press.

Mirowski, P. (2002). *Machine Dreams: Economics Became a Cyborg Science*, Cambridge: Cambridge University Press.

Rizvi, S. (1994) "Game theory to the rescue?" *Contributions to Political Economy* 13: 1-28.