Origins of Epistemic Game Theory*

Adam Brandenburger†
J.P. Valles Professor
New York University

Version 07/25/10

1 Introduction

In a 1935 article titled "Perfect Foresight and Economic Equilibrium," Oskar Morgenstern wrote about how members of a social system form expectations, expectations about expectations, and the like:

> [T]here is exhibited an endless chain of reciprocally conjectural reactions and counter-reactions.... The remedy would lie in analogous employment of the so-called Russell theory of types in logistics. This would mean that on the basis of the assumed knowledge by the economic subjects of theoretical tenets of Type I, there can be formulated higher propositions of the theory; thus, at least, of Type II. On the basis of information about tenets of Type II, propositions of Type III, at least, may be set up, etc. [14, 1935, pp.174-176]

Morgenstern had written earlier [13, 1928, p.98] about a "battle of wits" between Sherlock Holmes and Professor Moriarty in which each tries to think about what the other is thinking, about what the other is thinking he (the first) is thinking, and so on. In 1935, Morgenstern added the suggestion that such ingredients could be captured via formal mathematical-logical methods.

It took approximately fifty years for a formalization of Morgenstern's bold but more-or-less forgotten idea to appear–in the modern subfield of game theory called epistemic game theory.

What accounts for the delay?

2 The Protective (Maximin) Criterion

I suggest that the answer, at least in part, is that von Neumann, the intellectual giant with whom Morgenstern embarked on the systematic construction of game theory, put different considerations center-stage.

*I am indebted to Sergei Artemov, Amanda Friedenberg, Rena Henderson, Willemien Kets, Alex Peysakhovich, Ariel Ropek, Gus Stuart, and Natalya Vinokurova for important input and to Harold Kuhn and Robert Leonard for kindly providing a copy of Wolfe [22, 1955]. Financial support from the Stern School of Business is gratefully acknowledged. el5q-07-25-10

†Address: Stern School of Business, New York University, 44 West Fourth Street, New York, NY 10012, adam.brandenburger@stern.nyu.edu, www.stern.nyu.edu/∼abranden

A key theme in von Neumann's famous 1928 paper [20, 1928] is the payoff that an individual player–or group of players–can be guaranteed to get. In adopting a maximin strategy, a player "is protected against his adversary 'finding him out'" [20, 1928, p.23]. This is von Neumann's theory of play in two-player zero-sum games. It is also his theory for n-player zero-sum and general-sum games. For these games, von Neumann introduces the characteristic function of a game, defined by assigning to each subset of players the total payoff that it can guarantee itself via coordinated choice of actions–regardless of the coordinated actions that the players outside the subset might choose. In brief, each subset is assigned the maximin payoff of that subset.

The maximin criterion obviates Morgenstern's epistemic considerations. Each player (or set of players) considers the 'worst-case' scenario in terms of which strategies the other player (or players) might choose. No player (or players) tries to predict what other players will do by putting him/herself in their shoes and thinking about what they might be thinking, and so on. In *Theory of Games and Economic Behavior* [21, 1944], von Neumann and Morgenstern made this very explicit: "Nor are our results for one player based upon any belief in the rational conduct of the other" [21, 1944, p.160].

It seems that von Neumann's agenda dominated Morgenstern's.

One more observation from *Theory of Games and Economic Behavior*: Does the game model determine how a game is played? Von Neumann and Morgenstern said no:

> [W]e shall in most cases observe a multiplicity of solutions. Considering what we have said about interpreting solutions as stable 'standards of behavior' this has a simple and not unreasonable meaning, namely that given the same physical background different 'established orders of society' or 'accepted standards of behavior' can be built.... [21, 1944, p.42]

Alternatively put, the reason for multiplicity is that outcomes are under-determined by the game model. Additional factors–of a more 'intangible' kind–also matter. In this way, the von Neumann-Morgenstern philosophy can be thought of as a kind of indeterminism.

3 The Equilibrium Criterion

Nash's reformulation of the n-player theory removes both the cooperative and the maximin aspects of the von Neumann-Morgenstern theory. He puts the question of what rational individual play is firmly on the table. We might even say "–back on the table" since we could argue that this question is closer to what Morgenstern was talking about in 1935. But, we will see a big difference in how the two saw the question.

Nash wrote:

> We proceed by investigating the question: what would be a 'rational' prediction of the behavior to be expected of rational[ly] playing the game in question? By using the principles that a rational prediction should be unique, that the players should be able to deduce and make use of it, and that such knowledge on the part of each player of what to expect the others to do should not lead him to act out of conformity with the prediction, one is led to the concept of a solution defined before. [16, 1950]

The key components of Nash's argument are that: (i) associated with each game is a unique correct way to analyze that game; (ii) this way is accessible to the players themselves; and (iii) each player makes the best choice of strategy for him/herself.

In (ii), Nash is saying that a player can step outside the game, so to speak, and adopt the role of observer or analyst. If players do this, they see the game exactly the way an observer does. The player vs. observer issue is very interesting–and we will come back to it.

Right now, we focus on (i). The equilibrium-selection program recognizes the importance of this step for Nash, and tries to narrow down the set of equilibria for any given game to a single point. (The position can be taken that without a theory of equilibrium selection, Nash's argument breaks down. As an aside, we note that Hillas and Kohlberg [10, 2002, Section 8.2] distinguish the refinements program from the selection program. They define the former as concerned with identifying necessary–not sufficient–conditions on candidates for the 'right' equilibrium of a game.)

In any case, Nash supposes uniqueness and (via step (ii)) concludes that each player will reach a correct conclusion about the strategies that the other players choose. Add rationality (step (iii)), and we arrive at Nash equilibrium.

Nash's uniqueness assumption is very different from the multiplicity that von Neumann and Morgenstern saw as a natural and desirable state of affairs.

Also, Nash's conclusion that each player is correct about the others' choices is very different from how Morgenstern–who wrote about "faulty, heterogeneous foresight" [14, 1935, p.174]–appears to have been thinking.

It is well known that von Neumann did not receive Nash's idea positively. (See Shubik [19, 1992, p.155].) Was von Neumann convinced later? It seems not. Multiplicity remained central to von Neumann's picture of game theory. Here is a report of what he said at a 1955 conference on game theory at Princeton:

> The discussion opened with a statement by von Neumann in justification of the enormous variety of solutions which may obtain for n-person games. He pointed out that this was not surprising in view of the correspondingly enormous variety of observed stable social structures; many differing conventions can endure, existing today for no better reason than that they were here yesterday. [22, 1955, p.25]

However, with the rise of the Nash equilibrium concept, von Neumann-Morgenstern multiplicity–which can also be called von Neumann-Morgenstern indeterminism–was pushed aside.

4 Types

Harsanyi [9, 1967-8] wanted to analyze uncertainty about the structure of a game–specifically, about the players' payoff functions. To this end, he introduced the fundamental concept of a player's "type," which, he asserted, could be used to encode what the player believes the payoff functions to be, what the player believes other players believe the payoff functions to be, and so on indefinitely. (This brilliant idea was given a formal justification much later, by Armbruster and Böge [1, 1979], Böge and Eisele [5, 1979], Mertens and Zamir [12, 1985], and others.)

For a finite set X, write $\mathcal{M}(X)$ for the set of probability measures on X. Given sets X^i, for $i = 1, \ldots, n$, let $X = \times_i X^i$, and, for each i, $X^{-i} = \times_{j \neq i} X^j$. Now fix, for each player i, a finite set S^i of strategies for player i. Harsanyi's formalism consists of, for each i,

- a finite set T^i of types for player i;
- a map $f^i : T^i \to \mathcal{M}(T^{-i})$;
- a map $g^i : T^i \to S^i$;

- a map $h^i : S \times T \to \mathbb{R}$ (the reals).

(Sometimes, the map g^i is to $\mathcal{M}(S^i)$ rather than S^i, but, we can 'purify' these maps by defining new type spaces $U^i = T^i \times [0,1]$. Of course, this entails extending the framework to infinite type spaces. We omit the details in this short piece.)

Example 4.1 Figures 4.1 and 4.2 are simple illustrations (taken from Myerson [15, 1985, p.241]) of Harsanyi's formalism. Start with Figure 4.1. Ann has one possible type t^a, while Bob has two possible types t^b and v^b. Ann's type t^a assigns probability 3/5 to t^b and probability 2/5 to v^b. Also, type t^a plays U and has payoff function h^a. Of course, both types t^b and v^b for Bob assign probability 1 to t^a. Type t^b plays L and has payoff function h^b. Type v^b plays R and has payoff function $\tilde{h}^b$. The payoff functions are depicted in Figure 4.2.

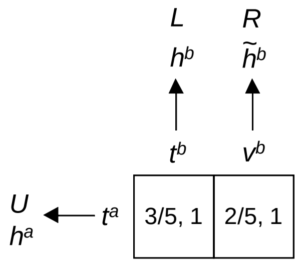


Figure 4.1

		Bob L	Bob R
Ann	U	1, 2	0, 1
	D	0, 4	1, 3

(h^a, h^b)

		Bob L	Bob R
Ann	U	1, 3	0, 4
	D	0, 1	1, 2

$(h^a, \tilde{h}^b)$

Figure 4.2

In this example, Ann has a simple induced hierarchy of beliefs about the payoff functions. She assigns probability 3/5 to Bob's having payoff function h^b and probability 2/5 to his having payoff function $\tilde{h}^b$. She assigns probability 1 to Bob's assigning probability 1 to her having the payoff function h^a. And so on to higher levels.

One more item from Harsanyi's formulation: He required each type to optimize as follows. (Given maps $\varphi^i : X^i \to Y^i$, let $\varphi^{-i} = \times_{j \neq i} \varphi^j$.) For each player i and each type t^i for i,

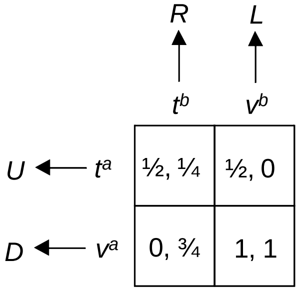
$$\sum_{t^{-i} \in T^{-i}} f^i(t^i)(t^{-i}) h^i(g^i(t^i), g^{-i}(t^{-i}), t) \geq \sum_{t^{-i} \in T^{-i}} f^i(t^i)(t^{-i}) h^i(s^i, g^{-i}(t^{-i}), t) \tag{4.1}$$

for all $s^i \in S^i$. This, of course, is the condition for Bayesian equilibrium. It is easy to check that Ann's type and both of Bob's types in Example 4.1 satisfy the condition. We will refer back to this item shortly.

5 Epistemic Game Theory

Here is another example that uses Harsanyi's formalism.

Example 5.1 There are two types t^a and v^a for Ann, and two types t^b and v^b for Bob. Figure 5.1 depicts the probabilities associated (via the maps f^i) with each type for either player. For example, type t^a assigns probability 1/2 (resp. 1/2) to Bob's being type t^b (resp. v^b). The diagram also depicts the strategy associated (via the maps g^i) with each type. There is only one payoff function for each player–namely, the payoff function depicted in Figure 5.2–and so we have suppressed the maps h^a and h^b.



	t^b	v^b
t^a	½, ¼	½, 0
v^a	0, ¾	1, 1

Figure 5.1

		Bob	
		L	*R*
Ann	*U*	2, 2	0, 0
	D	0, 0	1, 1

Figure 5.2

Figure 5.1 is a simple example of the type structures used in epistemic game theory. Suppose that the actual state of the world is (U, t^a, R, t^b). We can then calculate Ann's hierarchy of beliefs over the strategies chosen. We see that Ann assigns probability 1/2 to Bob's choosing R and probability 1/2 to Bob's choosing L. Ann also assigns: (i) probability 1/2 to the event "Bob chooses R and assigns probability 1/4 to her (Ann's) choosing U and probability 3/4 to her choosing D"; and (ii) probability 1/2 to the event "Bob chooses L and assigns probability 0 to her (Ann's) choosing U and probability 1 to her choosing D"; and so on to higher levels. We can likewise calculate Bob's hierarchy of beliefs over the strategies.

We can also talk about the "rationality" or "irrationality" of a type. The rationality criterion is exactly inequality (4.1) above. Thus, each player is rational: Type t^a for Ann optimally chooses U, and type t^b for Bob optimally chooses R. We also see that Ann assigns probability 1/2 to Bob's being rational (type t^b) and probability 1/2 to Bob's being irrational (type v^b). Bob assigns probability 1/4 to Ann's being rational (type t^a) and probability 3/4 to Ann's being irrational (type v^a). Again, we can continue to higher levels.

There are two important differences between the two examples:

- In Example 4.1, we did not calculate the hierarchies of beliefs over the strategies. The reason is that, in this example, if Ann were to come to know Bob's payoff function, she would be certain–and correct–about the strategy he chooses. Her hierarchy of beliefs over payoff functions determines her hierarchy of beliefs over strategies. The situation in Example 5.1 is different. Ann knows Bob's payoff function (there is only one), but she does not assign probability 1 to his actual choice of strategy. (Likewise from Bob's perspective.)

- In Example 4.1, all types optimize–i.e., satisfy inequality (4.1). In Example 5.1, some types optimize and some do not. In particular, types v^a and v^b are irrational.

These two new features in Example 5.1 mark the transition from the world of Bayesian equilibrium to the world of epistemic game theory (EGT). The first new feature introduces into game theory the idea of uncertainty about the strategies in a game–in addition to uncertainty about the structure of a game. The second new feature is the introduction of irrationality, or belief in irrationality, or belief about belief in irrationality, and the like.

It is true that, in principle, both features are expressible in Harsanyi's formalism (more precisely, provided that one does not insist on Bayesian equilibrium). But, in practice, this was not how his formalism came to be used. Indeed, in the numerical examples that Harsanyi himself used in [9, 1967-8] to illustrate his framework, neither feature is present.

A clear break is evident in the papers by Bernheim [4, 1984] and Pearce [17, 1984]. Written during the height of the equilibrium-refinements program, while many people were working on trying to narrow down the set of Nash equilibria in a game, these two papers challenged the view that Nash equilibrium was the inevitable starting point of analysis in the first place.

Rather than banish uncertainty about strategies (as Nash did), Bernheim and Pearce make this uncertainty central. (But, they did not treat irrationality.) Ann has subjectively formed probabilities about Bob's choice of strategy, constrained only by the assumption that she believes him to be rational, she believes he believes her to be rational, and so on. Call this the assumption of "common belief" of rationality. Actually, Bernheim and Pearce assumed "common knowledge" of rationality. On common knowledge, see Aumann [2, 1976] and Lewis [11, 1969], and also the remarkable earlier work by Friedell [8, 1967] (re-discovered by Barry O'Neill). The belief-knowledge distinction is very important in EGT, but we will not go into it here.

It is intuitively clear that the assumption of rationality and common belief of rationality implies that each player chooses an iteratively undominated strategy–i.e., a strategy that survives iterated elimination of strongly dominated strategies. EGT proper began with formal proofs of this assertion–and of an appropriate converse–using type structures like the one in Figure 5.1. This set the foundation for subsequent research in the area. Up to the present, EGT has studied the epistemics of irrationality as well as rationality, the epistemics of game trees, and the epistemics of weak dominance, among other issues.

Here is a general definition of the field today (for which I am grateful to Sergei Artemov): *EGT makes epistemic states of players an input of a game and devises solution concepts that take epistemics into account.*

6 Indeterminism Again

Where do the type structures of EGT come from? For a given game, what determines the appropriate type structure?

The answer is that the type structure is to be understood as part of the description of the situation being studied–on a par with the strategy sets, the outcome map, and the payoff functions (and the information sets in a tree). Remember that a player's payoffs are that player's own evaluation of the possible outcomes of the game. What a player believes, what a player believes other players believe, etc., is also subjective. Both payoffs and beliefs are subjective inputs into the game model. (In decision theory, Savage [18, 1954, p.3] uses the illuminating term "personalistic" in place of "subjective.") Neither payoffs nor beliefs can be deduced from other components of the model; they must both be described.

An epistemic analysis will, in most cases, depend on the particular type structure used. In such cases, the outcome of the analysis will be under-determined by the classical game model. We are back to the same situation as the one von Neumann and Morgenstern found, and for broadly similar reasons. There are 'intangible' as well as 'tangible' components of the model, and we should not expect determinacy from the second kind of components alone.

This said, there are type structures that have a special status. These are type structures that, in one sense or another, contain all possible beliefs. Various notions of such a structure have been given: terminal structures (Böge and Eisele [5, 1979]); canonically-built structures (often called universal structures, see Mertens and Zamir [12, 1985]); and complete structures (Brandenburger [6, 2003]).

Epistemic analysis on such structures can yield sharp results. An example is the important paper by Battigalli and Siniscalchi [3, 2002], which states epistemic conditions on a complete structure and derives from them unique outcomes in a number of games of applied interest.

7 Epistemics and Logic

Most work in EGT has used the tools of Polish spaces, Borel probability measures, etc. But the topic of large type structures turns out to be one where tools from mathematical logic have also proved useful. These tools have the virtue of being explicit about the methods of reasoning being used to think about a game.

Taking a logic perspective on epistemic conditions of the "I believe that you believe that I believe..." kind might lead one to suspect that some type of self-reference could arise in game theory. Similar to occurrences of self-reference in other areas (most famously, of course, in set theory), could this lead to an impossibility result?

Jerry Keisler and I [7, 2006] investigated this question, and found the following theorem: Let the players use a language that includes first-order logic (and symbols for the relations in the type structure). Then, a large such structure–formally, a structure that is complete relative to this language–does not exist.

An interpretation is that the type structure is a tool that the analyst uses to describe the game. If this tool is also available to the players, then a difficulty can arise. Should the language(s) used be restricted to avoid the impossibility? Or, is the key to maintain a sharp distinction between the players and the analyst of a game? Logical tools seem well-suited to investigating these (and other) questions in EGT.

8 Conclusion

Morgenstern's bold idea of using the tools of formal logic to talk about how members of a social system think, about how they think about what other members think, and so on, was far ahead of its time. But now, in the form of EGT, it has found a home.

References

[1] Armbruster, W., and W. Böge, “Bayesian Game Theory,” in Möschlin, O., and D. Pallaschke (eds.), *Game Theory and Related Topics*, North-Holland, Amsterdam, 1979.

[2] Aumann, R., “Agreeing to Disagree,” *Annals of Statistics*, 4, 1976, 1236-1239.

[3] Battigalli, P., and M. Siniscalchi, “Strong Belief and Forward-Induction Reasoning,” *Journal of Economic Theory*, 106, 2002, 356-391.

[4] Bernheim, D., “Rationalizable Strategic Behavior,” *Econometrica*, 52, 1984, 1007-1028.

[5] Böge, W., and T. Eisele, “On Solutions of Bayesian Games,” *International Journal of Game Theory*, 8, 1979, 193-215.

[6] Brandenburger, A., “On the Existence of a ‘Complete’ Possibility Structure,” in Basili, M., N. Dimitri, and I. Gilboa (eds.), *Cognitive Processes and Economic Behavior*, Routledge, 2003, 30-34.

[7] Brandenburger, A., and H.J. Keisler, “An Impossibility Theorem on Beliefs in Games,” *Studia Logica*, 84, 2006, 211-240.

[8] Friedell, M., “On the Structure of Shared Awareness,” Paper #27, Center for Research on Social Organization, University of Michigan, 1967.

[9] Harsanyi, J., “Games with Incomplete Information Played by ‘Bayesian’ Players, I-III,” *Management Science*, 14, 1967-8, 159-182, 320-334, 486-502.

[10] Hillas, J., and E. Kohlberg. “Conceptual Foundations of Strategic Equilibrium,” in Aumann, R., and S. Hart (eds.), *Handbook of Game Theory*, Vol. III, Elsevier, 2002, 1597-1663.

[11] Lewis, D., *Convention: A Philosophical Study*, Harvard University Press 1969.

[12] Mertens, J-F., and S. Zamir, “Formulation of Bayesian Analysis for Games with Incomplete Information,” *International Journal of Game Theory*, 14, 1985, 1-29.

[13] Morgenstern, O., *Wirtschaftsprognose, eine Untersuchung ihrer Voraussetzungen und Mödglichkeiten*, Springer Verlag, 1928.

[14] Morgenstern, O., “Vollkommene Voraussicht und wirtschaftliches Gleichgewicht,” *Zeitschrift für Nationalökonomie*, 6, 1935, 337-357. Reprinted as “Perfect Foresight and Economic Equilibrium,” in Schotter, A. (ed.), *Selected Economic Writings of Oskar Morgenstern*, New York University Press, 1976, 169-183.

[15] Myerson, R., “Bayesian Equilibrium and Incentive-Compatibility: An Introduction,” in Hurwicz, L., D. Schmeidler, and H. Sonnenschein (eds.), *Social Goals and Social Organization*, Cambridge University Press, 1985, 229-260.

[16] Nash, J., “Non-Cooperative Games,” doctoral dissertation, Princeton University, 1950.

[17] Pearce, D., “Rational Strategic Behavior and the Problem of Perfection,” *Econometrica*, 52, 1984, 1029-1050.

[18] Savage, L., *Foundations of Statistics*, Wiley, 1954.

[19] Shubik, M., "Game Theory at Princeton, 1949-1955: A Personal Reminiscence," in Weintraub, E.R. (ed.), *Toward a History of Game Theory*, Duke University Press, 1992, 151-164.

[20] Von Neumann, J., "Zur Theorie der Gesellschaftsspiele," *Mathematische Annalen*, 100, 1928, 295-320.

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

[22] Wolfe, P. (ed.), Report of an informal conference on *Recent Developments in the Theory of Games*, Department of Mathematics (Logistics Research Project), Princeton University, January 31-February 1, 1955.