Economic Thought 7.1: 24-49, 2018

The Backward Induction Controversy as a Metaphorical Problem

Ramzi Mabsout, American University of Beirut
rm95@aub.edu.lb

Abstract

The backward induction controversy in game theory flared up and then practically ended within a decade – the 1990s. The protagonists, however, did not converge on an agreement about the source of the controversy. Why was this the case, if opposing sides had access to the same modelling techniques and empirical facts? In this paper I offer an explanation for this controversy and its unsettled end. The answer is not to be found in the modelling claims made by the opposing protagonists, but in the tacit metaphors they operate under. Aristotle defined metaphor as giving a 'thing a name that belongs to something else' (Poetica, 1457b). The meaning of metaphors has not changed much since then – in contrast to models which are comparatively new, and still not well-understood, scientific tools. The controversy of backward induction in game theory provides a test bed for the explanatory power of metaphors. This paper frames the controversy in terms of metaphor choice to provide a common framework for the protagonists. This results in the identification of three different domains – mathematical logic, game theory and the world – each connected to the other via different metaphors. The controversy around backward induction is placed in, and tentatively explained by, this framework.

Keywords: backward induction, common knowledge of rationality, game theory, metaphors, models, rationality

JEL Codes: B4, B16, C72

1. Introduction

> 'How we think about such questions depends on the predilections we bring to an inquiry, on our suppositions about what will count as an answer, on our explanatory preferences' (Keller, 2002).

If metaphors are as important as their advocates argue they are, their function in science is for cognitive purposes – not merely ornamental. Metaphor, as a cognitive concept, should contribute to our knowledge of the world (Boyd, 1979 [1993]) and function 'whenever some phenomenon of cognition is conceptualised or explained through the use of metaphor' (Hoffman, et al., 1990, p. 177). The process by which metaphors work is increasingly well-understood in both natural[1] and social[2] sciences. The advocates of metaphors claim that metaphors can be helpful, across a range of disciplines, to form theory, posit novel

[1] See Emmeche and Hoffmeyer (1991), Hesse (1966), Keller (2002), Pulaczewska (1999), and Jeppson et al. (2012).
[2] See Bicchieri (1989), Boumans (1999), Bronk (2008), Brown (1977), Clarke et al. (2014), Fernandez-Duqe and Jonhson (1999), Hodgson (1995; 2005), King (2012), Klamer and Leonard (1994), Leary (1990), McCloskey (1983; 1994), and Mirowski (1988; 1989; 1994).

hypotheses, construct models, explain phenomena, solve problems, and interpret empirical results.

Metaphors are qualitative and procedural – not exacting logically, 'never complete, precise, or literal mappings'[3], nor even fixed in time and across contexts. Even if not currently acknowledged in economics, metaphors are formal conceptual tools too, in the sense that they can be used in both teaching and research. They operate on various levels. First, economists themselves can be under the spell of a meta-metaphor, also known as a root, constitutive, or theme metaphor, which explains the origins of the other metaphors in use; second, the model itself can be seen as a metaphor; third, metaphors can be found within the model connecting its various inputs to bridge the gap between the world and the model. One implication is that metaphors will not vanish once an idea has been mathematically modelled, 'they are constitutive of scientific discourse' (Bicchieri, 1988a, p. 104).

Metaphors enable mapping between two distinct domains. In science, their primary function is establishing links between scientific language and the world (Kuhn, 1979 [1993, p. 539]). In economics, this mapping connects the (mathematical) model, or syntactic structure, to the economic world (which, it may be noted, economic theory could have helped create performatively). Without the metaphor, the structure consists of un-interpreted syntactical relationships. There is no need to think further than the supply and demand cross to illustrate the working of metaphors. Within this textbook example, the metaphor maps the geometric or algebraic properties (specifically, the intercepts and slopes coefficients) to features of the world, namely, the relationship between quantities and prices of some goods. We could, in metaphoric terms, conceive of the metaphor as 'the engine of the model' since it selects the properties that are mapped from one domain to another, abstracting from all other features kept constant with *ceteris paribus*. Without such a mapping, the model is a set of equations and lines silent about the world. The mapping generated by the metaphor permits empirically meaningful explanations, possibly even predictions (Hesse, 1966).

This paper explores a game theoretic controversy on the meaning and use of backward induction (BI) to compute equilibria. The controversy ended in an apparent unexplained deadlock. The back and forth between the protagonists ends abruptly, so to speak, as neither side gives ground or agrees on what was actually the subject of the dispute. The controversy has also spawned a micro industry that feeds on what axioms justify BI. I shall have more to say on this towards the end of the paper. The controversy is amenable to metaphorical interpretation as the protagonists (Aumann, Binmore, and L. Samuelson, among others) do not agree on the use or meaningfulness of BI. The BI equilibrium entails players who will not cooperate or retaliate, failing to reach mutually beneficial payoffs. BI can be considered the textbook game theoretic solution. A controversy arises since such outcomes are not deemed 'rational' especially in repeated games. Furthermore, empirical evidence (corroborative or not of BI) has had no apparent implications on the controversy.

This controversy can elucidate (or not) the workings of metaphors in economic modelling. Metaphors can offer at least a partial explanation of the controversy and why the protagonists do not converge on agreement. There may be other explanations, of course, but here we shall focus on the plausibility of one, namely, the choice of metaphor.

There are reasons to suppose that, when controversies such as this one arise, the source might be metaphorical. Thus, it was pointed out that, 'when theories run into problems, both the problem and their proposed solutions are consequences of the logic of the metaphors that are at work' (Fernandez-Duque and Johnson, 1999, p. 83) and 'many acrimonious debates in the history of economics would have been clarified tremendously if these [metaphorical] tenets had been kept in view' (Mirowski, 1988, p. 139) or that

[3] Hodgson (1995).

> 'controversies in economic discourse can be clarified, and identification slips avoided, if the user of metaphor specifies its type… controversies arising from misunderstanding can be resolved, saving intellectual energy to the defense of the type of resemblance which a particular metaphor is supposed to reveal' (Khalil, 2000, p. 7).

Although not acknowledged by the participants in the controversy, such views suggest that the BI dispute might have its origins in metaphor choice.

The problem pursued in this paper is whether the BI controversy is one that should involve, and would be solved by, the identification metaphor choice. If metaphor choice captures an essential part of the controversy the discussion can move to justifying the legitimacy of different metaphorical choices. To this aim, section 2 defines metaphors and discusses their relationships with models, narrative, and analogies. Section 3 offers an overview of metaphors and their use in economics. Section 4 introduces the game theoretic controversy on BI. In section 5 the controversy is cast and assessed in terms of metaphor choice. Section 6 concludes.

2. Metaphors, Analogies, Models, or Stories?

I argue in this section that metaphors are primitive but simpler and sturdier than models in their ability to explain controversies such as the game theoretic dispute on the relevance of backward induction (BI). A successful metaphorical explanation involves showing that a scientific disagreement is based on the choice of metaphor. It would provide understanding about why certain authors adopt and understand some models, arguments or findings while others – presented with the same models, arguments or findings – fail to be convinced. What is not explained in this paper is how metaphor choice is, or ought, to be made. This requires further historical epistemological analysis that lies beyond the scope of the paper, but that would nevertheless shed light on why certain individuals have come to adopt some metaphors and others not.

Metaphors have had a continuous consistent definition since (as far I know) Aristotle who argued that metaphor gives a 'thing a name that belongs to something else; the transference being either from genus to species, or from species to genus, or from species to species, or on grounds of analogy' (Poetica, 1457b). Models, however, unlike metaphors, are more recent and still not well-understood. According to Heyck (2015), before 1950, only 7% of articles in the social sciences used the word 'model' or its variants. By the 2000s the range is between 70% and 90%, depending on the social science. But Morgan and Morrison (1999) remark, 'there remains a significant lacuna in the understanding of exactly how models function to give us information about the world' (p. 7) and 'we have very little sense of what a model is in itself and how it is able to function in an autonomous way' (p. 8). Accordingly, instead of labelling the large number of heterogeneous practices that fall under modelling, including metaphorical explanation, as models – it is strategic to avoid this broad categorisation which hides rather than reveals heterogeneity in scientific practices. Why label metaphors 'models', if metaphors have a clear and distinct definition? Modelling imperialism – seeing models everywhere – can, I suggest, add confusion instead of clarity. Metaphors are more primitive, simpler and sturdier explanatory entities.

While metaphors can be described as models, their distinct and simple internal structure would be lost in the large range of heterogeneous practices placed under modelling. In her most recent book on models, Morgan (2012, p. xv) argues that she no longer attempts to offer a definition of models because of the heterogeneity of objects that count as models. She contends that models are not easy to characterise and there are no easy answers as to what models are or how modelling works. In fact, 'there are lots of different kinds of things that legitimately count as models… and they often look and function very differently' (Morgan, 2012, p. xvi).

If some models have a simple internal structure, the different entities labelled as models are extremely broad. We could still of course label metaphors models. However, this adds an unnecessary conceptual layer: why refer to metaphors as models if we can directly refer to them as metaphors? It should be noted, nothing in my argument militates against the use of models except that no one has done that before (for this specific controversy at least) and I remain open to consider such an alternative.

The received view in economics, nevertheless, is that it is a modelling science – modelling is the principal tool in economics. Mäki (2002, p. 10) remarks that 'to do economics is to do modelling' and defines, following Robert Solow, model building as a 'fact oriented activity that takes as its objective to isolate key causal dependencies in reality' (p. 11). Earlier, Gibbard and Varian (1978) argued too, that economic theorising consists of investigating economic models. They claim that models are used 'whenever there is economic reasoning from exactly specified premises' (p. 666). Economic models, for Gibbard and Varian, have two elements: stories (that carry the interpretation) and an un-interpreted logical mathematical structure (the syntactical part). They argue that economic models pose counter-factual questions (of the following sort: what would happen if such and such was the case?) that are useful in generating explanations. Sugden (2002) too emphasises the explanatory power of models in economics seeing them as describing credible counterfactual worlds useful to warrant inferences from the model-world to the real world.

While Morgan (2002) supports Gibbard and Varian's claim that stories are necessary for economic modelling, she claims their account of how stories integrate deductive models is incomplete. McCloskey's account of the complementarity between metaphors (as models) and stories also fails to adequately describe models because models are not reducible to metaphors. For Morgan and Morrison (2000), models are *autonomous agents* partially independent from, but interacting with, theory and data. In subsequent work, Morgan (2002) identifies two aspects of models – the story/narrative and the structure/metaphor. The theoretical claims are embodied in the structure of the model which determines 'the relationships between the elements of the model' (Morgan, 2002, p. 195), connects the theory to the model and constrains the narrative. Morgan maps metaphors and (mathematical) structures on one side as elements of models arguing they need stories to produce knowledge about the world (p. 183). Stories are needed even when the mathematical elements that constitute the structure are interpreted (p. 189). The structure, specifically, contains mathematical equations that shape the story which, in turn, is not fully determined by structure (p.188). Models connect to the world in two distinct ways, firstly, in building the model, that is in the mathematics themselves involving the realism of assumptions and 'deeper questions about the nature of representation and denotation' (p.192), and secondly, in connecting the mathematics of the model to the world via story. The story here is 'a cognitive tool, a tool by which we explain something or come to understand something about the world' (p.193). Between theory and the world lies the model as an autonomous complex entity. The story becomes necessary when it is necessary to contextualise the model so that 'in telling stories with the model, we use it to explain the specifics of why coffee prices are high in 1976' (p. 194). On the relationship between structure and story Morgan claims that

> 'where to start the tale, which questions are interesting and relevant, and even the order of solving the model is somewhat open – the user has to make sensible choices in order to tell meaningful stories which are plausible and interesting about the world' (p. 195).

The gap in Morgan's account is the function of metaphors which are conflated with the mathematical structure. Instead, one function of metaphor is to bridge the gap between narrative and structure by generating a mapping. Stories connect the mathematical structure to the world using metaphors. I shall, in fact, argue that metaphors operate at two levels: at the highest level they sanction which stories can be told (the 'method of using metaphorical reasoning to construct historical narratives' Mirowski, 1994, p. 14) while they embed the story, at the lower level so to speak, with more specific metaphors. It is at the lower level that narrative integrates the mappings of metaphors to connect the mathematical structure of the model to the world.

Consider the following definitions of models and metaphors: 'the model as a metaphor' and 'the metaphor as a model' (Brown, 1977). Some authors define models in terms of metaphor (McCloskey, 1983; Bronk, 2008; Brown, 1977) while others see the model in terms of an analogy (Hesse, 1966; Klamer and Leonard, 1994), and yet for others, models are carriers of metaphors (Bicchieri, 1988a; Bouman, 1999; Morgan, 2002). As noted earlier, variegated opinions on what models are is to be expected, but that does not mean there is no agreement how metaphors work. The other side of the duality, 'metaphor is a model', requires explanation as well. This side appears redundant as noted earlier: an account of how metaphors work can be given without reference to models. As Maasen et al. (1995, p. 1) explain, in fact, interest lies not in duplicating and expounding 'fine grained terminological distinctions between metaphors, images, analogies, models, rhetoric, and systems of thought' but in how metaphors permit 'the transfer… of pieces of meaning from one delineable discourse to another'.

Another clarification needs to be made, namely, on the relationship between analogy and metaphor. Hesse's (1966) view is that while analogy resembles a relation of mathematical proportionality (the word for proportion in Greek is *analogia*), it is not constrained by it (pp. 66-7). Hesse also contends that Aristotle 'speaks of metaphor as being based on analogy' (p. 133) and, in her conception of theories as metaphors, she favours explanation as metaphoric re-description of the explanandum (the phenomena that needs explanation) against the orthodox deductive model of explanation. For McCloskey (1983), metaphors belong to the larger class of analogies whereas Mirowski (1988) uses both terms together and interchangeably. According to Hoffman et al. (1990, p. 213), metaphors arise before analogy which are '*post hoc* relative to the root metaphor'. Klamer and Leonard (1994, p. 34) are faithful to Aristotle's definition of analogy 'as a specie of metaphor' yet distinct. In a metaphor, Klamer and Leonard argue, there are attributes in common between the principal subject (mind, time, market) and the subsidiary domain (machine, money, game) whereas an analogy 'draws explicit parallel between subject and subsidiary domain' (p. 34).[4] An analogy is less than a full-blown metaphor since 'it does not capture all the associated commonplaces suggested by the metaphor' (p. 35) and requires no imaginative leap since it is based on the (Aristotelian) principle of proportionality. Klamer and Leonard agree with Hoffman et al. that analogy is an elaborated metaphor that focuses on certain relationships suggested in the metaphor. The metaphor is a heuristic which leads to the generation of analogies that form testable models. In their specification of metaphors, they note that

[4] As we shall see, contrary to what Leonard and Klamer claim, some metaphors are said to be heterologous in that the mapping is based on functions rather common properties in both domains.

> 'the mere coinage of a metaphor such as "human capital" does not make science. Science proceeds by taking a fertile metaphor and relentlessly articulating the nature of its subsidiary domains, probing the properties of that terrain, and testing the connections between that domain and the principal domain' (p.35).

An interesting implication is that, for Klamer and Leonard, models are not metaphors but an 'explicitly, most often formally articulated analogy' (p.35). Cohen (1994, p. 57), similarly, uses the Aristotelean definition of metaphors, defining analogies as denoting 'a similarity that centers on an equivalence or likeness of functions or relations or properties'. Lagueux (1999, p. 15) contends that analogies are methodologically more acceptable since metaphors violate 'what logic requires' whereas analogy is an explicit comparison in which the 'distinctiveness of the respective domains is explicitly preserved… [it is] perfectly suited to scientific analysis.' Given these variegated views, we can synthetise the following points: (i) metaphor and analogy are related and cannot be disentangled; (ii) metaphor uses analogy but is not reducible to it; (iii) analogy both contributes to the metaphorical mapping and can be the outcome (in the form of a testable model) of a further elaboration of the original metaphorical mapping.

3. Economic Metaphors: an Overview

Faced with a controversy in game theoretic modelling, how should an economist or a methodologist evaluate it? I will argue in this paper that some controversies could be explained in terms of the strategic choice of metaphors. The definition of metaphor as connecting, via a mapping, two distinct domains is all the tool kit needed to generate the explanation.

Economics contains a number of metaphorical expressions, including 'equilibrium', 'elasticity', 'human capital', 'accelerator', 'GNP is up', 'prices are inflated', 'liquid assets', 'price mechanism', and 'policy instrument' (as listed in Klamer and Leonard, 1994 and McCloskey, 1983). They are considered thought-changing, breaking the habit of thought as it were, by employing a deviation from the literal meaning to the figurative meaning: 'time is money', 'time flies', or 'mind as machine' (time is not money, time has no wings to fly, and the mind arises from an organic, not mechanic, organ). Metaphors, when not dead, are fluid and open to interpretation, being highly sensitive to the context in which they are used. The distinction between the literal and figurative meaning is also not necessarily dichotomous but continuous: expressions which begin as a metaphor harden, freeze, or die of overuse (Bicchieri, 1988a; Klamer and Leonard, 1994, p. 27). Their death in science, however, unlike poetry, signals a successful metaphor which has spread (Boyd, 1979 [1993]).

The metaphorical expressions listed above fill a gap in the economics lexicon and, while useful, they do not to elucidate how metaphors work. Such metaphorical expressions, it is said, are sanctioned by a higher-level metaphorical mapping between two domains (Lakoff, 1993, p. 209). I have found six different labels of these two domains.[5] In what follows, two accounts that reflect different perspectives on how metaphors work are discussed. The first view posits that metaphorical mappings generate a conceptual mapping of entities, properties, relations, and structure from a *source* to a *target domain* (Fernandez-Duque and

[5] Klamer and Leonard (1994) mention a few (i) subject / predicate, (ii) tenor / vehicle, (iii) target /import, and (iv) principal/subsidiary. Fernandez-Duque and Johnson use (v) source / target domain while Hesse (1966) and Bicchieri (1988a) (vi) secondary / primary subjects.

Johnson, 1999; Lakoff and Johnson, 1980; Lakoff, 1993). This mapping is illustrated with the 'the mind as machine' metaphor.

The 'Mind as Machine' Metaphor[6]

Source domain / subsidiary or secondary subject		*Target domain / principal or primary subject*
Machine	→	Mind
Functions within machine	→	Mental capacities
Products of the machine	→	Ideas
Automated machine functioning	→	Thinking
Normal machine functioning	→	Normal thought
Breakdown of machine	→	Inability to think

Each arrow above takes some entity or structure in the source domain and constructs a counterpart in the target domain. The mappings provide a 'fixed pattern of ontological correspondences across domains' that can be activated, not algorithms that mechanically take the source domain inputs and produce target domain outputs' (Lakoff, 1993, pp. 210-14).[7] These mappings also submit to the *invariance principle* which states that 'metaphorical mappings preserve the cognitive typology of the source domain, in ways consistent with the inherent structure of the target domain' (*ibid*.,). The invariance principle can be understood as a constraint on the correspondences that constitute the mapping where cognitive typology is the image-schematic structure of the domains. Similarly, the image-schematic structure of the target domain can in turn limit the possibilities of mappings from the source domain.

A more dynamic account of how metaphors function is given by Black (1962), and earlier Richards (1936), who contend that metaphorical meaning arises from the interaction between a *principal* and a *subsidiary subject* or, equivalently, between a *primary* and *secondary subject* (Black, 1979 [1993]).[8] Hesse's account of metaphors is that they work

> 'by transferring the associated ideas and implications of the secondary to the primary system. These select, emphasize, or suppress features of the primary; new slants on the primary are illuminated; the primary is seen through the frame of the secondary… it follows that the associated ideas of the primary are changed to some extent by the use of the metaphor… [T]he same applies to the secondary system, for its associations come to be affected by assimilation to the primary; the two systems are seen as more like each other; they seem to interact and adapt to one another…. Men are seen to be more like wolves after the wolf metaphor is used, and wolves seem to be more human' (Hesse, 1966, p. 163).

[6] From Fernandez-Duque and Johnson (1999, p. 85).

[7] More specifically, Lakoff (1993, p. 249) explains that the 'contemporary theory of metaphor is at odds with certain traditions in symbolic artificial intelligence and information processing psychology. Those fields assume that thought is a matter of algorithmic symbol manipulation, of the sort done by computer programs. This defining assumption is inconsistent with the contemporary theory of metaphor.'

[8] Aristotle identified four kinds of metaphors, and though he excluded them from logic (Klamer and Leonard, 1994, p. 24), his definition appears to be of the interactionist sort since it involves giving a 'thing a name that belongs to something else; the transference being either from genus to species, or from species to genus, or from species to species, or on grounds of analogy' (Poetica, 1457b)

Thus, while the (unidirectional) mapping gives the principal subject (man, time or mind) a name that belongs to the subsidiary subject (wolf, money or machine), the bi-directional interactionist account entails that the metaphor can modify the prevalent interpretation of both the principals and the subsidiaries as novel meanings, not reducible to or substitutable by a literal expression, arise. The interactionist view of metaphors allows us to clearly distinguish between analogy and metaphor, captures the *resonance* or *expansion* of a metaphor more adequately than the unidirectional view, and integrates performativity, for example, in the possibility that the market (as the principal subject that borrowed properties and relations from the subsidiary subject, the game) may in turn also modify our understanding of the game.

Distinctions between various kinds of metaphors have been made. Khalil (2000) identifies four kinds (the nominal, heterologous, homologous and unificational) while Klamer and Leonard (1994) identify three (the pedagogical, the heuristic and the constitutive).[9] Pedagogical metaphors are the simplest as they illuminate and clarify a complicated concept by providing mental images to help an audience. They are closest to Khalil's nominal metaphors which only use a superficial similarity between the principal and the subsidiary. They usually help answer the question 'what is the intuition?' and 'what is the story?' behind a mathematical model. It is pedagogical metaphors that scientists and economists usually have in mind when thinking of metaphors.

Nominal and pedagogical metaphors, however, are not the most influential since they may be omitted without affecting an argument. Heuristic metaphors are part and parcel of theories.[10] They cannot be paraphrased or substituted with a literal expression. Klamer and Leonard argue they are necessary to catalyse our thinking – they are thought-propelling. Accordingly, an example of a successful heuristic metaphor is that of 'human capital'. The 'human capital' metaphor signals the beginning of an inquiry and will, given the resonance it creates over time, generate additional developments including testable analogies.[11]

Constitutive metaphors frame our thinking, determine what makes sense, and work at the fundamental level of Kuhn's paradigm. They are spectacles necessary for the interpretation of our world and include 'those sets of assumptions, usually implicit, about what sort of things make up the world, how they act, how they hang together and, usually by implication, how they may be known… [they] constitute the ultimate presuppositions or frames of reference for discourse on the word or on any domain' (Brown, 1977, p. 125). Consider the following metaphors: 'we've hit a dead-end street', 'we can't turn back now', 'we may have to go our separate ways', 'look how far we've come', 'it's been a long, bumpy journey'. They all refer back to the same constitutive metaphor, namely, 'love as a journey' (Lakoff, 1993). The 'human capital' metaphor, similarly, is congruous with the constitutive metaphors of neoclassical economics (such as physical capital accumulation) and succeeded as a heuristic metaphor. Another example is the mechanistic world metaphor which generated concepts

[9] Maasen et al. (1995) identify three kinds of metaphors as well, the illustrative, heuristic and the constitutive. The definitions of the first two are the same as Klamer and Leonard but their definition of constitutive differs, however, since they see their function is to replace previous meaning by new ones.

[10] Boyd (1993, p. 486) refers to them as constitutive metaphors, but since this term is already employed by Klamer and Leonard to mean something else, it will be avoided.

[11] Human capital appeared roughly around the same time in Mincer (1958), Schultz (1961) and Becker (1964). Capital in economics classically refers to physical capital such as machinery and plant. Human capital, likewise, referred to education and skills as investments that generate returns for the owner. Human capital is also an input in the production process. The 'human capital' metaphor renders human capital interpretable as one of the inputs (alongside physical capital and technology) in a standard production function. Although an expenditure by individuals, human capital is distinct from the consumption of other goods since it provides a return in the future, like any other investment. In Schultz's case, a parallel between physical/capital investments and human skills investment is made explicit. In this way Schultz could use the existing capital terminology to explain large increases in national output. It is here, Klamer and Leonard argue, that the connection between thought in science and metaphor is strongest.

such as price mechanism, equilibrium and elasticities, among others (Brown, 1977; Hodgson, 1995). Hard to specify concretely, constitutive metaphors tend to operate below conscious awareness and 'can be exposed only by digging into or interpreting the relevant texts, both spoken and written' (Klamer and Leonard, 1994, p. 41). Constitutive metaphors answer the question of where our heuristic metaphors come from.

While the nominal distinction made by Khalil (2000) overlaps with Klamer and Leonard's (1994) pedagogical metaphor, the heuristic and constitutive metaphors, to the extent that they work for uncovering structures, processes and powers, may overlap as well with Khalil's heterologous, homologous and unificational metaphors. Khalil, however, is pessimistic about this overlap since he sees Klamer and Leonard, not as realists, but as sophists – more interested in persuasion than in uncovering real phenomenon. Heterologous metaphors exist when there is resemblance of analytical function, but the context or origins are not the same (for example, the wings of a bird and the wings of a bat, although both perform flying, emanate from a different context). As an example of a heterologous metaphor in economics, Khalil considers those on spontaneous order arising from climatic and ecological systems (subsidiary) and socio-political order (principal). Homologous metaphors exist when there is no common analytical function, but a similar scheme, context or common origin (thus, though the forelimbs of mice and bats have different functions they have the same origins and are homologous). Examples of homologous metaphors in economics include the evolution and entrenchment of habits and biological evolution; the division of labour within the firm and the differentiation of functions within organisms; and the autocrat of a chimpanzee troop with the modem state. Khalil identifies unificational metaphors as the strongest kind, in that the same law must be operating in the principal and subsidiary subject. Thus, the law of gravity is unificational because it is used to explain various physical and astronomical phenomena; the similarity of blood circulation in humans and chimpanzees is also considered a unificational metaphor. As for economics, Khalil claims that optimisation unifies disparate phenomena by drawing on the similarities between household and firm maximisation (utility and profits respectively).[12] Khalil, finally, explains that the use of any metaphor is appropriate if and only if it is classified in its appropriate category. He provides three levels of identification slips: from single degree (heterologous metaphor is used when the similarity is only nominal) to triple degree (unificational metaphor is used when the similarity is nominal) and offers, for each level of identification slip, an example from economics.

As this section has shown, metaphors can be more specifically defined than models. All types of metaphors generate a mapping from one domain to another. Equipped with this view of metaphors, the next section focuses on the backward induction controversy. In section 5 use will be made of the understanding gained in sections 3 and 4 to explain it.

4. The Backward Induction Controversy

'The economists do not know why they disagree' (McCloskey, 1990).

[12] Cohen (1994) adopts an overlapping but different classification from Khalil's, namely, analogy, homology, identity and metaphor. While Khalil considers all four kinds of relationships metaphorical, for Cohen, identity is similar to Khalil's unification metaphor while homology expresses similarity in form (not function) and is thus similar to Khalil's homologous metaphors. Aristotle has, in this regard, identified two types of analogies, analogies 'when there are properties in common' and (Platonic) analogies 'when there is similarity in the relation of the parts' (Hesse, 1966, p. 134; see also p.142). In his mature thought on analogy, Hesse claims that Aristotle combines both senses (p. 138).

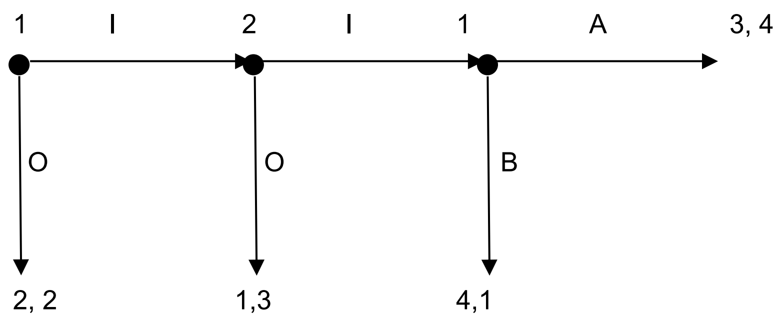
> '[O]ut-of-equilibrium play occurs with zero probability if the players are rational' (Grüne-Yanoff and Lehtinen 2012)

Backward induction (or BI) is a method to compute equilibria in finite, usually perfect information, extensive form games. It involves the analysis of games from back to front, proceeding by the elimination of dominated strategies.[13] However, the usefulness and epistemological function of BI led to a controversy to be elaborated upon in this section. The question is, in what terms can this controversy be explained?

BI led to the Nash refinement literature and to so-called subgame perfect equilibria (Selten, 1975). The term refinement is used since such equilibria involve additional criteria which take the form of eliminating non-credible threats. Subgames require an initial node and are self-contained games within a larger game (including the game itself). A node ***x*** initiates a subgame if neither ***x*** nor any of its successors are in an information set that contains nodes that are not successors of ***x***; the subgame is the tree structure defined by such a node ***x*** and its successors (Watson, 2013). A subgame perfect Nash equilibrium exists when there is a Nash equilibrium in every subgame of the larger game, and when a subgame is reached the players will play according to this equilibrium strategy. It follows that not all Nash equilibria are subgame perfect, but a subgame perfect equilibrium has to be a Nash equilibrium.[14] This classical account of playing games (to be further described below) is, via BI, considered 'the only possible pattern of play by rational players' (Bicchieri, 1988b, p. 383).

BI, combined with rationality and common knowledge of rationality (CKR), entails that in games such as the centipede, players choose at every decision node 'down' in Figure 1 until the first node is reached.

Figure 1 Rosenthal's (1981) extensive form centipede type game (player 1, player 2)



To understand why, note that in Figure 1 player 1 will opt for down (B) in the last node of the game because the payoff is highest (4>3). Knowing this player 2 plays down as well since this maximises payoff (3>1). Back to the first node of the game, player 1 knowing player 2 will play down if she had the chance will play O (2>1) ending the game before it starts. In repeated games, players play various rounds (of the same game). BI again implies that non-dominated

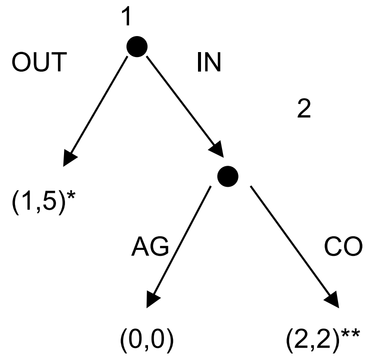
[13] The first explicit reference to BI is due to Kuhn (1953). Luce and Raiffa (1957, p. 68), though not the first, state that 'at a terminal choice point – we are assuming that all games have a stopping rule, and this enables us to work backward – the player whose move it is will naturally adopt the choice which suits him best. Thus, since the last choice is determinate, we may as well delete it and place the appropriate payoff directly at the terminal move position, if this is done for each terminal move, the penultimate moves now play role terminal moves, and so the process may be carried backward to the starting point.'

[14] This is because all non-credible threats Nash equilibria have been eliminated from the smaller set of subgame perfect Nash equilibria. Simulations have also shown that it is possible to derive Nash equilibria that are not subgame perfect (Binmore and Samuelson, 1996; Gale, Binmore and Samuelson, 1995).

choices are made: starting from the last game each player opts for the choice that maximises their own pay-off irrespective of what the other plays until the first game is reached. In the finitely repeated prisoner's dilemma, this means that players will defect on each and every round of the game. The repeated prisoner dilemma, although not a perfect information game, submits to a similar reasoning and both players are expected to defect. A third game where BI is applied is the chain-store game in Figure 2 (Selten, 1978). Here, an incumbent monopolist (player 2) holds a monopoly in *m* towns. The monopolist does not declare a price war when one of the *m* competitors, one in each town, decides to compete. In the chain store game, the monopolist's decision to play AG (for aggressive or price war) is seen as a non-credible threat since it is not in its interest to retaliate and it will play CO (for cooperate) (Gibbons, 1997). It should be noted here that this game has *two* Nash equilibria {IN, CO} and {OUT, AG} but only not declaring a price war {IN, CO} is subgame perfect.

Figure 2 Selten's (1978) Chain store game in extensive and normal form

1 \ 2	CO	AG
IN	2,2**	0,0
OUT	1,5	1,5*



*Nash Equilibrium
**Sub-Game Perfect Nash Equilibrium

Selten (1975, p. 35) early on identifies a difficulty with BI, noting that 'there cannot be mistakes if the players are absolutely rational'. Selten (1978) later observes that there is a paradox because, while it is more advantageous for the monopolist to cooperate in the short term (the BI and game theoretic decision), it is better to play aggressively in the long term (the deterrence, more convincing decision). Accordingly, the repeated chain store game is paradoxical because only BI is theoretically correct yet playing aggressively – and starting a price war – is much more convincing. The paradox is elegantly described by Hargreaves Heap and Varoufakis (1995) who argue that the Nash equilibrium (1,5) in figure 2 is eliminated by backward induction since it is not a credible threat. However, the subgame perfect Nash equilibrium (2,2) is singled out with the now out-of-equilibrium strategy (1,5). This creates a puzzle since while CKR is assumed, the rational strategy is identified by considering

> 'what would happen if what turns out to be an irrational move were to be made at some point… [an] equilibrium behaviour needs to be built on an analysis of out-of-equilibrium behaviour… we have to introduce the possibility of some lapse of rationality to explain what rationality demands' (pp. 87-88).

This analysis brings about two (difficult) questions for Hargreaves Heap and Varoufakis, namely, are such lapses from rationality consistent with CKR and how can one assume rational play in out-of-equilibrium play?

Selten, in this regard, argues that a satisfactory interpretation of equilibrium in extensive games seems to require that the possibility of mistakes is not completely excluded. Selten introduces irrational play assuming players are subject to rationality imperfections so that, at every information set *u,* there is a small probability ε_u for the breakdown of rationality. He sets the stage for the controversy, noting that there cannot be any unreached information set and that this is consistent with the definition of strategy profiles which inform what the player will do at every information set of the game, specifying behaviours even over unreached subgames.

Experimental results on the use of BI are ongoing but far from conclusive. In over 1000 experiments conducted since the late 1950s, cooperative choices were made about 30% of the time in the repeated prisoner dilemma (even as it emerged that experience raises the chances of defection or playing the BI equilibrium; see Andreoni and Miller, 1993; Colman, 1998, p. 356). Binmore, Shaked and Sutton (1985) find that in a bargaining game, individuals learn to play BI in round two when the roles are reversed. Balkenborg (1998) runs experiments over what he calls the basic (or stage) game and notes that 80% of the results support the outcome predicted by BI (with 13 sessions and 12 subjects, each playing the game 50 times, randomly varying with anonymous opponents). Johnson et al. (2002) test the extent to which deviation from the BI path is explained by 'limited cognition' or 'equilibrium social preferences'. They find that both contribute to explaining deviations from BI and suggest that individuals are not equipped to use BI without prior training. After parcelling the ultimatum game[15] into rationality, subgame consistency and truncation consistency, Binmore et al. (2002) find evidence against the use of BI. They thus back the long-held result of Güth et al. (1982). There are also experimental results against the use of BI in the ultimatum game (Henrich et al., 2005; Guala, 2008; Roth et al., 1991), the *p*-beauty contest (Camerer, 2003a), and the centipede game (McKelvey and Palfrey, 1992). Various explanations have been offered for these deviations from BI. Smith (2003), for instance, explains the deviations in the ultimatum, dictator and other trust games – in terms of reciprocity, instead of a preference for fairness in the utility function. More specifically, he posits a neurocogntive explanation based on the capacity of players to mind-read the other player's moves (McCabe et al., 2000).

A common critique against BI is that it is paradoxical (Basu, 1990; Bicchieri, 1989; Luce and Raiffa, 1957; Pettit and Sugden, 1989). The culprits are the conjunction of CKR and rationality. The argument is that if players are rational they may have to consider cooperation. Yet, and herein lies the paradox, BI renders cooperation non-utility maximising as it assumes (in the centipede game, for example) player 2 plays down if the second node, which is not meant to be reached, is reached. If node 2 is not meant to be reached according to the theory, how can it predict that player 2 plays down? Pettit and Sugden's (1989) argument against the BI equilibrium is innocuously simple and follows this train of thought, namely, that common belief in rationality breaks down – and the BI equilibrium does not prevail – when one of the players acts cooperatively in the repeated prisoner's dilemma. The paradox was demonstrated more rigorously in Reny's (1993) proof that rationality (as utility maximisation) and CKR are inconsistent in two-person perfect-information finite games. If either CKR or rationality is dropped, Reny further argues, BI is no longer the only type of rational play. Reny's own interpretation of his proof questions the plausibility of subgame perfection.

[15] In the ultimatum game, for example, player 1 makes an offer equal to a portion of a total sum to be shared with a receiver who then decides to accept or reject the offer. If the offer is rejected – as often happens when the offer is much under 50% of the total – no player gets any payoff. If the offer is accepted both players get the agreed sums. The subgame perfect equilibrium is for the receiver to accept any positive offer that is made since something is better than nothing. Rejection rates of positive non-trivial offers, however, are quite common across many cultures.

Until now the BI equilibrium has been criticised on at least two connected grounds: it is unreasonable for players not to cooperate when this will benefit all with enough rounds to play the game; and players must take hypothetical decisions in nodes they will never reach (CKR and rationality are inconsistent). The controversy, however, really begins after Aumann's (1995) formal mathematical proof that CKR and rationality in a perfect information game suffice to justify the BI equilibrium. More specifically, Aumann argues that when CKR and rationality prevail, no vertices off the BI path are reached (p. 18). Aumann sustains that the proof supports the intuition that common knowledge implies BI while relying on the usual meaning of concepts such as knowledge and rationality. Irony has its ways, since it was mostly as a response to Reny et al. that Aumann (1995) offered his proof. At this stage perplexity is allowed since we are left with a proof against a proof (Reny vs Aumann) on the one hand, or common sense and intuitions against the proof on the other (Aumann vs Pettit, Sugden, Bicchieri, Basu, Selten et al.). Even after Aumann (1995) published his proof, both supportive (Rabinowicz, 1998) and contesting counter-proofs (Binmore, 1997; Ben-Porath, 1997) were published.[16]

Binmore and Samuelson (1996) were quick to counter Aumann (1995), initially disputing the use of mathematical logic to establish the proof. Binmore (1997, p. 24) subsequently added that the proof reduces to 'inventing fancy formalisms… only to confuse matters'. Aumann's proof, it is claimed, brushes aside the question of how the players acquired rational and unambiguously-defined beliefs in the first place. For Binmore and Samuelson, equilibria should not be established via the static definition-axiom-theorem-proof format (that closes the mind) but via algorithms of players' reasoning and 'constructive' simulations of the equilibrating process.

According to Binmore and Samuelson, the problems of classic game theoretic rationality are compounded by CKR. Unless counterfactuals – such as rational players acting irrationally – are accounted for in the strategy profiles of players, rationality makes no sense. According to the critique, the BI path can only be justified with counterfactuals (of choices that could have been made but were not).[17] The possibility of hypothetical decisions of the sort 'what should player 2 do if player 1 does not play down but cross instead?' is now seen as requiring special attention. How would a player explain that the other player is not playing the BI path? What kind of mistake or irrationality brought us to node 50 out of 100 instead of ending the game at the first node? These questions were raised by Binmore and Samuelson (1996) and Binmore (1996; 1997) as a critique of Aumann (1995). While the critique acknowledges that rational play entails the first movers to play down (eg., Bicchieri, 1988b; Binmore, 1987, p. 196; 1997), the critique also states that rationality is not adequately modelled if players do not account for what they would have done if the other player does not follow the BI path.

Aumann's proof was meant to answer these critiques, which turned only more virulent, insisting his conception of rationality is mistaken as long as it does not specify what the players would play if they deviate from the path of BI. Aumann's (1996a, b) response is that his proof (i) does not necessarily imply that rational players will not deviate from the BI

[16] Without CKR, Rabinowicz defends BI for a class of BI-terminating games where rationality is a choice of moves, not strategies. Here BI-terminating games are games, such as the centipede games, where down ends the game, excluding the finitely repeated prisoner dilemma. Binmore uses a finite version of the centipede game to show that, even with CKR, the equilibrium of the game is a Nash mixed strategy equilibrium, not the BI equilibrium. Ben-Porath assumes CKR only at the first node, exploiting the distinction between certainty which allows surprises (playing cross with probability 0) and knowledge (which does not allow surprises), as well as the possibility of changing beliefs, Ben-Porath claims that the BI equilibrium is no longer the only justifiable equilibrium.

[17] Note that such arguments require observable behaviours by players or else there would not be a paradox, and indeed there is no such paradox in simultaneous-move games (Reny, 1993).

path; (ii) he insists that rational players may deviate at any point including the first move; and (iii) that the inductive choice used in the proof could be irrational. He acknowledges that his conception of rationality and strategy does account for player *i*'s knowledge of what the other player would do had *i* played across instead of down (in the centipede game). Aumann suggests his critics are confusing the assumption of rationality with CKR and shows that his theorem still holds if he adopts Binmore's stronger definition of rationality. It is relevant to add here that Aumann's interpretation has recently received the support of empirical game theorists (Gintis, 2009). In summary, Aumann (1996a, b) argues that as long as rationality is common knowledge, a player deciding across is absurd logically, since down is the only possible outcome at the first information set of the game.

By the looks of it, the debate has stalled, with both sides entrenched and unable to dig deeper for answers. The core of the controversy is what (if anything) can justify BI. For skeptics, when rationality and CKR do not contradict each other, something else beyond rationality and CKR, is needed, including maybe stories from outside the game. As tension emerges between game theoretic reasoning, on the one hand, and intuition or common sense, on the other hand, is it surprising that in the most formal modelling branch of economics – game theory – so much weight is given to intuition? If metaphors are as important as their proponents argue they are, this should not come as a surprise. The presence of intuitions and common sense may still be explained as an inchoate metaphorical choice that needs to be formally acknowledged. The literature on metaphors can step in to formally account for these choices.

5. The Strategic Use of Metaphor

How can metaphors be used to provide an explanation for the controversy on BI? Recall that Aumann assumes rationality and CKR sufficient to justify the use of BI in computing equilibria. This sufficiency relies on a mathematical proof that shows that irrational choices could not have been made (Aumann, 1996a, b). Binmore and Samuelson (1996), on the other hand, disagree and posit that his mathematical proof is irrelevant: something else is needed to justify BI. The nature of the controversy between Binmore and Samuelson, on the one side, and Aumann, on the other, remains puzzling and in need of explanation. Indeed, why is there a controversy if (i) the mathematical proof of BI with CKR is rigorous?[18] And (ii) if empirical tests unfavourable to BI could be – but are not – brandished to undermine its validity? Binmore and Samuelson (1996) argue that the controversy is not within mathematics: they (albeit not others) agree that the mathematical proof that CKR leads to the BI equilibrium is rigorous, yet they suggest that the proof has little value. The critics of Aumann are convinced that something is not right, and that the BI equilibrium cannot be justified in the way Aumann does (in fact his way is perceived as so fundamentally wrong by Binmore that it could prevent game theory from being taken seriously).

This section uses the definitions of metaphors provided earlier to explain the confusion around the status of BI. The first, if obvious, possibility is that Aumann is under the spell of a constitutive metaphor different from the one adopted by his critics. Keeping in mind that constitutive metaphors, as noted by Klamer and Leonard, are hard to specify concretely, we can posit that Aumann's constitutive metaphor casts the foundations of game theory in mathematical logic, which is the source domain of its theoretical results in terms of solution

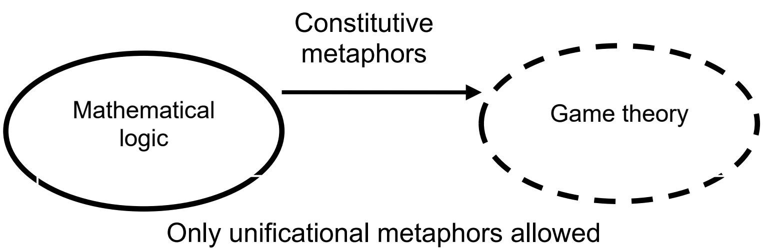
[18] Binmore (1987, p. 196), for example, claims that 'It is not disputed that the results of the play of this [centipede] game by rational players will be that I plays "down" at the first node' Binmore (1997) also provides his own proof.

concepts and equilibria. Within such a constitutive metaphor, there is little sensitivity to the context or type of game. Indeed, most of Aumann's results can be applied to a range of games, not one. Here there is less concern for the structure and context of the game than his critics would like. In mainstream economics there could be other competing and conflicting constitutive metaphors behind the critique's claim that game theory may become irrelevant if it adopts too abstract or idealised foundations à la Aumann. The critique operates under at least one, possibly various, constitutive metaphors that do not use the same source domain as the foundation of game theory.

If this controversy is indeed about the adoption of different constitutive metaphors, it should be acknowledged as such. The confrontation, accordingly, is not over who has the best mathematical proof or which theory is more supported empirically – but which constitutive metaphor is more adequate for the foundations of game theory. Focusing on constitutive metaphors also forces questions on explaining why different constitutive metaphors are employed by different individuals. While helpful in some ways, framing the controversy simply as one of constitutive metaphors fails to provide a comprehensive explanation for the controversy. For a start, it would force us to place Binmore, Samuelson or Bicchieri under the spell of one constitutive metaphor, a highly doubtful claim. Indeed, there could be various constitutive metaphors uniting the opponents of Aumann on what the foundations of game theory ought to be based. Secondly, the conflict of constitutive metaphors does not rule out the possibility that some of Aumann's critics adopt the same constitutive metaphor of mathematical logic (eg., Reny, 1993). And thirdly, among those who defend the equilibrium of BI, not all use mathematical logic 'that closes the mind' but stories and common-sense arguments (eg., Broome and Rabinowicz, 1999; Sobel, 1993). Unless there is a yet unidentified constitutive metaphor, which could tightly explain by regrouping those for and against the justification of BI via common knowledge of rationality, constitutive metaphors, for the time, being are not sufficient to explain the controversy away.

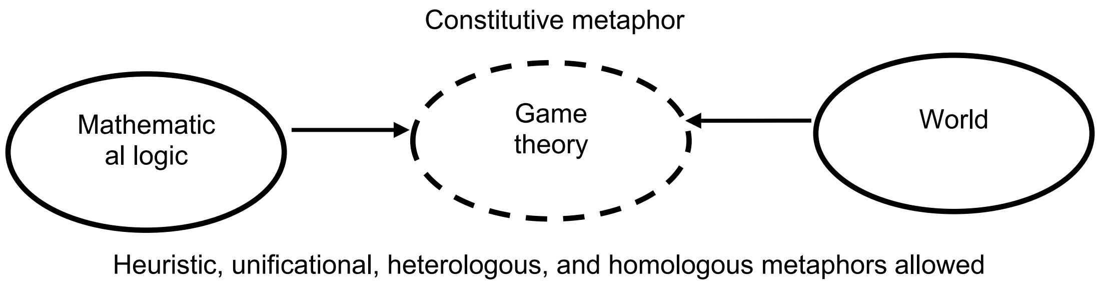
This does not mean constitutive metaphors play no role in the controversy. Inspired by Popper's (1972) three 'worlds', a finer possibility is to frame the controversy as one of three metaphorically interconnected, but relatively autonomous, domains namely, mathematical logic, the game and the economic world (see Figures 3-5). A similar division is used by Grüne-Yanoff and Schweinzer (2008) (and more recently Grüne-Yanoff and Lehtinen 2012) to describe the architecture of game theory. Their view is, as ours, but lacks the metaphorical theoretical justification that explains the triptychal architecture. Game theory connects or bridges the gap between the domains of mathematics and the world. It is a semi-autonomous model between mathematical logic and the complex economy (Morrison and Morgan, 1999). Each source domain maps its properties into game theory. Accordingly, mathematical logic and the world map their properties into decision theory, which then generates game-theoretic models. The metaphors on the right hand side that connect the model to the economy are much wider than the unificational metaphor that connects mathematical logic to the model – because the model, and more so the world, are highly complex entities.

Figure 3 Aumann's constitutive metaphor



Aumann's foundation of game theory is limited to the constitutive metaphor in Figure 3. His metaphor allows only unificational mappings that connect mathematical logic to game theory.[19] This is where Aumann is mostly active (in the controversy at least): he maps mathematical, logical properties into game theory, constructing different models that say nothing specific about the right-hand side – the economic world. In a wider constitutive metaphor – various kinds of heuristic, unificational, heterologous and homologous metaphors operate between the game and the world on the right-hand side. Binmore, Samuelson and others partake in the controversy using metaphors from both ends (Figure 4). Accordingly, they do not allow metaphors from the left-hand side only. While they use some left-hand side mapping, they also focus on metaphors that hook game theory to the world.

Figure 4 Binmore and Samuelson's metaphorical mappings



The changing metaphors across the triptych can lead to identificational slips of the form identified by Khalil, or to the breaches of invariance identified by Lakoff. Such slips and breaches are explained by the prevalent understanding of what games are and what they are attempting to do. They may be strategically used to undermine the work of an opponent. Binmore (1997), for example, freely uses proofs – and even has his own version of the BI equilibrium proof – yet he at the same time criticises Aumann's use of mathematical logic to prove the BI equilibrium. Binmore and Samuelson's slip is to project heterologous or homologous metaphors – that operate between the game and the economic world (via simulations and computational economics) – onto the unificational relationship that connects the game to mathematical logic. In the language of the Lakoffian invariance principle, their metaphorical mapping violates the image-schematic structure that mathematical logic maps into the game, its target. The acceptance of this identificational slip, or breach of invariance, is of course contingent on accepting our description of the controversy as a metaphorical triptych between mathematics, the game and the world.

Within each of the constitutive metaphors, the participants in the controversy employ distinct heuristic, heterologous, homologous or unificational metaphors. Since Aumann (1995) is working on the left side of the triptych, his proof is consistent with the constitutive metaphors that unify game theory and mathematical logic. Such unificational metaphors have the principal as the game and mathematics as subsidiary. The connection is unificational because the same principles of mathematical logic exist in the principal and the subsidiary subjects. Without exception, the properties of the mathematical proofs are mapped as possibilities of solution concepts in the game (they constrain the theoretical form of the game in the same way as the structure of the game constrains the metaphors that connect it to the economic world). One such proof uses CKR and rationality to explain the game, or the equilibrium of the game, via BI. When Aumann notes that he wants to keep the proof as

[19] Hesse (1966, p. 137) claims that Aristotle 'speaks in several places of the basic truths of logic and mathematics as being "one by analogy" when they apply in different fields.'

transparent and as simple as possible he echoes Bicchieri – that familiarity and manageability are two parameters that guide the selection of metaphor. Aumann is comfortable with the principles of mathematical logic (his subsidiary subject is both familiar and manageable) to prove that the equilibrium entails that the first mover plays down. Furthermore, Aumann (1995, p. 6) suggests, this proof is expansionary since there is a modern refinement literature supporting it and his own work extends a number of recent papers on similar fundamental notions of non-cooperative game theory.

Binmore and Samuelson are more sensitive to the fact that they are dealing with metaphorical choices and acknowledge that (i) context matters especially when nodes not meant to be reached by rational players, *are* reached (Binmore, 1987, p. 196); and that (ii) a preliminary informal classification of different equilibrating processes, through the choice of interesting environments, in which games are played should be made (Binmore, 1987, p. 183). As noted in section 2, context is critical for metaphors to be rigorously interpreted. Aumann, however, suggests that his proof is context-free and works as an ideal gas whose implications are not affected by what happens in practice. Statements by Binmore (1997, p. 28) that 'it is at the interpretive level that the importance of common knowledge assumptions needs to be acknowledged' are significant for choices of metaphors. Binmore and Samuelson are explicit that Aumann's subsidiary domain (mathematical logic) is inadequate to justify the computation of equilibria. They claim that the current sequence of 'axiom-definition-theorem-proof' does not just close the mind to irrelevancies (a good thing), but that it also closes the mind to issues it is perilous to neglect. They suggest instead that the proper way to identify potential equilibria is by using simulations and stories to interpret counterfactuals. Such modes of reasoning are sanctioned by the broader, constitutive metaphor that connects the game to the world.

Binmore (1987) presents his approach to game theory as fundamentally different from Aumann's. He, firstly, posits an algorithmic, 'machine programmable' definition of rationality.[20] Secondly, Binmore and Samuelson (1996, p. 114) proceed to search for stories that could explain deviation from the rational BI path.[21] Binmore and Samuelson list stories to overcome Aumann's problem, which is seen as the traditional approach to game theory. Their approach identifies mechanisms that explain deviation from rational play (trembling hand, irrational mistakes, defective reasoning). What is the relationship between stories and metaphors? The stories use metaphors to form subsidiary subjects whose properties are mapped onto the principal subject – the solution concept of the game that needs explanation. As Morgan (2007, p. 169 emphasis added) states for the prisoner's dilemma, 'the narratives translate the prisoners' situation into the economic situation – *they link particulars to particulars* – and 'explain' how it is, for example, that two large firms can end up doing damage to each other just as the prisoners end up with the double-defect outcome'. The constitutive metaphor sanctions a particular type of story which uses a particular type of heuristic, unificational, heterologous or homologous metaphor to connect distinct domains. The metaphor and the story thus complement each other.

---

[20] For Binmore (1987, p. 181), a rational decision process refers to the 'entire reasoning activity that intervenes between the receipt of a decision stimulus and the ultimate decision, including the manner in which the decision-maker forms the beliefs on which the decision is made. In particular, to be rational will not be taken to exclude the possible use of the scientific method.'

[21] Their alternative interpretation of the centipede game involves a husband who, after missing his mortgage, explains to his (furious) wife that he would not have lost the repayment had he been dealt the ace of diamonds rather than the queen of spades in last night's poker game. They point out such counterfactual stories are not obtained from abstract mathematical contemplation, but as stories from the world. In the centipede game Binmore and Samuelson discuss two possible competing interpretations of counter-factuals. Both stories provide an explanation of irrational play, or play that plays across not down or out.

Two of the stories identified by Binmore and Samuelson provide counterfactuals without which rationality and deviation cannot be accounted for. Selten (1978), on the one hand, posits deviations from the BI path as a trembling hand which makes mistakes. Kreps et al., (1982), on the other hand, focus on modelling incomplete information into the finitely repeated prisoner dilemma. Kreps et al., posit that when a player is not sure what the other will play, and there is a small positive probability ($\delta>0$) she may cooperate, tit-for-tat cooperation will be the equilibrium outcome for long periods in finite games, including the prisoner's dilemma. A third possibility, not discussed in Binmore and Samuelson, but suggested by Coleman (1998), is to substitute rationality (and therefore modify CKR) with non-monotonic reasoning that reflects common-sense, everyday reasoning. This modification, Coleman argues, solves the BI paradox by offering theoretical options for players that face the unexpected across play in the centipede game, cooperation in the repeated prisoner's dilemma, or the declaration of a price war to prevent further entry in the chain-store game. All three stories provide entry points for distinct metaphors.

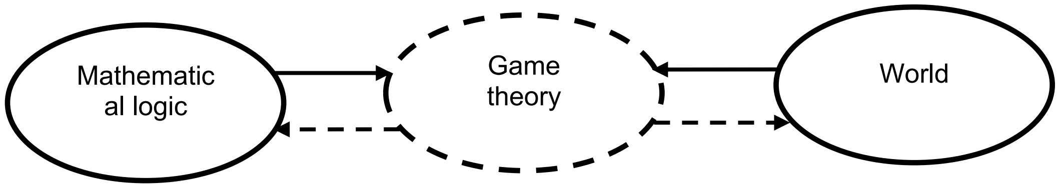
Who has the better story-metaphor combination? Binmore and Samuelson consider this is a wrongheaded question, which has no absolute answer, since it is always necessary 'to look at the context in which the game is played for inspiration on this score. But this context is exactly what is abstracted away when one adopts the conventional mathematical formalism' (p. 115). Thus, the trembling hand story is not applicable to chess – it provides a poor mapping from one domain to another (a chess player is unlikely to consistently make the same mistake due to a trembling hand). Binmore (1987), accordingly, states that irrational play in games such as chess should be modelled – not as trembling hand mistakes – but as defective reasoning. Which metaphor provides the most adequate mapping depends on the type of game and the context of application.

Understanding the dynamic interdependence-autonomy between the three domains of game theory can shed some light on its evolution. Whether the constitutive metaphors are fixed or not, whether one can take over from the other, is an open question I further elaborate upon in the conclusion. The triptych also depicts how game theory can performatively change both mathematical logic and the world (dashed arrows from game theory to mathematical logic and the world; Figure 5). An example of this possibility is provided by Morgan's analysis of a World War II text by game theoretician Rapoport (see also Rubinstein, 2006)

> 'While Rapoport suggests game theory was taken because of the "civilization" of war, it seems equally part of the process that war became acceptable because it was reinterpreted in game theory terms … the cold war came to be seen as a set of game situations… it comes to the point at which we understand and interpret that [nuclear arms] race as a prisoner's dilemma game' (Morgan, 2007, p. 159).

Properties of the model are mapped onto the world via game theory – the model changes the way the world is perceived ('the cold war came to be seen as a set of game situations'). But there is another metaphorical loop that maps properties of the model into mathematics, pushing for novel interpretations in mathematical logic and leading to developments in mathematics (mathematics as an applied science that can experience empirical discoveries and novel interpretations). Accordingly, the generation of new game theoretic models can lead to the development of novel mathematical objects, theories and techniques.

Figure 5 Performativity and game theory



The last point to be made here refers back to the BI micro-industry mentioned in the introduction. I distinguish between pre-Aumann (Basu, 1977; 1990; Bicchieri, 1988b; 1989; Binmore, 1987; Pettit and Sugden, 1989; Reny, 1993; Selten, 1978; Sobe,l 1993; Sugden, 1992) and post-Aumann (1995) publications (Ben-Porath, 1997; Broome and Rabinowicz, 1999; Aumann, 1996a,b; 1998; Binmore, 1996; 1997; Binmore and Samuelson, 1996; Rabinowicz, 1998; Stalnaker, 1996). Morgan proves helpful here again with her contention that game theory has grown

> 'from the narratives, which… go through a process of matching the economic situation with the game situation and then exploring how and why it does not fit. When it does not fit, a new version of the game is developed with slight changes in the rules, payoffs, or information arrangements' (Morgan, 2007, p. 176).

A review of only a few specimens, from post-Aumann (1995), supports Morgan's diagnosis. Stalnaker (1996), for example, posits common *beliefs* of rationality instead of CKR to defend the BI equilibrium (while Sugden (1992), pre-1995, does the opposite, namely, he uses so-called entrenched common beliefs to overcome the paradox of CKR and BI); Rabinowicz (1998) defends BI for a class of BI-terminating games where rationality is a choice of moves not strategies; Ben-Porath (1997) assumes CKR only at the first node, exploiting the distinction between certainty, which allows surprises (playing cross with probability 0), and knowledge which does not; Aumann (1998) distinguishes between *ex ante* and *ex post* knowledge operators of rationality and argues that the proof of the BI equilibrium in the centipede game via (a less intuitive) *ex ante* definition of rationality subsumes (a more intuitive) *ex post* definition of rationality. And so on and so forth. With some exceptions (cf., Camerer, 2003b and Colman, 1998), the publishing micro-industry on the BI controversy does not empirically confront BI, CKR or rationality. Instead, it creates interminable new taxonomies based on changing assumptions, introducing new definitions, logical proofs, lemmas and theorems. It is not clear to me where lies the epistemological contribution to social scientific knowledge of these additions to the BI controversy. I will further comment on this in the conclusion.

6. Conclusion

In this paper the cognitive efficacy of metaphors to explain a controversy in game theory was considered. It suggests that, appropriately employed, metaphors – as theoretical descriptions of the explanandum – can shed light on the source of the controversy around BI among game theorists. A metaphorical account casts the disagreement as one primarily due to protagonists operating under a different metaphorical spell. Those involved in this controversy published past each other over a period of a few years, possibly because the source of their misunderstanding – and that which would have aligned the discussion plane – is an acknowledgement of the strategic use of metaphors. If economists trade in model building, model-based reasoning has not led to the breakthrough that would have reconciled their differences. The metaphor was introduced as a simpler, sturdier and possibly, looser mode of reasoning to explain why there is a controversy. The metaphorical explanation – if adopted – would also move the controversy to a common plane, focusing the discussion on the context and source of disagreement. Even if the disagreement does not vanish and protagonists stick to their guns, as it were, they would have at least narrowed their disagreement to the choice of metaphor.

Binmore and Samuleson's critique of Aumann drove us to consider various uses of metaphors. At the highest level, the level of the paradigm, the confrontation is over which constitutive metaphor is more adequate, not which proof is more convincing or the extent to which proofs are weakened by empirical evidence or common sense play. Other metaphors within the theme (or allowed mappings) of the constitutive metaphors were identified.

Our metaphorical account of the controversy entails that any potential change is a complex and negotiated endeavour over constitutive metaphors. Any change must be sanctioned by the constitutive metaphor which acts as a filter. Unless the change is consistent with the mapping of the constitutive metaphor, it will be rejected for committing identification slips and breaches of the invariance principle. Binmore and Samuelson's attack on the 'axiom-definition-theorem-proof' sequence could be an example of an identificational slip that challenges the prevalent constitutive metaphor that posits a unificational relationship between game theory and mathematical logic.

The benchmark of a successful constitutive metaphor was linked to its uptake and expansion in the discipline it operates in. A metaphor's success is partly a function of the new perspectives, interpretations and explanations it generates. Metaphor, as pointed by Hesse (1966, p. 177), forms an essential element in 'the continuous adaptation of our language to our continually expanding world'. The metaphorical mapping that projects mathematical properties and relationships onto strategic decisions giving rise to game theoretic modelling and under which some of our protagonists operate, is now commonly accepted. This metaphoric mapping expanded the analysis of strategic decisions into new directions. It is, however, now an established metaphor and may be categorised as dead, having achieved success.

We cannot rule out that, in the future, new metaphors will project novel mapping onto the strategic analysis of decisions, transforming, weakening or strengthening the grip of the now dead mathematical metaphor. Nevertheless, uptake and expansion remain insufficient success criteria to adjudicate whether a successful metaphorical choice has been made for a realist to whom scientific theories need to explain by referring to the world. Hesse identifies a difficulty, which turns out to be a strength, in the interactionist view of metaphors and their referents. Rather than prove that metaphors in science refer in the orthodox understanding of refer, she argues that the orthodox view of explanation fails to refer whereas the metaphoric view refers to the mappings of properties and functions from one domain to another. In this way the understanding of 'referring to' is no longer the static application of a covering law and correspondence rules that connect theory to observation, as in the orthodox deductive view of explanation. Instead,

> 'the process of metaphoric description is such as to cast doubt on any simple identification of the metaphor's reference with the primary system. It is claimed in the interaction view that a metaphor causes us to "see" the primary system differently and causes the meanings of the terms originally literal in the primary system to shift toward the metaphor' (Hesse, 1966, p. 167).

Hesse then asks,

> 'how can initial similarities… justify such changes in the meanings of words and even, apparently, in the things themselves? Man does not in fact change because someone uses the wolf metaphor. How then can we be justified in identifying what we see through the framework of the metaphor with the primary system itself? It seems that we cannot be entitled to say that men *are* wolves, sound *is* wave motion' (Hesse, 1966, p. 167).

Ultimately, Hesse, like Khalil, wants to argue that metaphors in science are consistent with realism – they explain by referring to the world. However, how this can be achieved if 'the interaction view implies that the meaning of the original literal language of the primary system is changed by the adoption of the metaphor' (p.169)?

Hesse contends that the use of metaphors discards deductive literal descriptions that are inadequate and, without abandoning deduction, the metaphoric view

> 'focuses attention on the interaction between metaphor and primary system, and on the criteria of acceptability of metaphoric descriptions of the primary system, and hence not so much upon the deductive relations that appear in this account as comparatively uninteresting pieces of logical machinery' (Hesse, 1966, p.174).

Unlike the deductive view of explanation, which makes use of correspondence rules that fail to refer because the meaning of explanandum is shifting with the introduction of new theoretical terms in the explanans, in a metaphoric explanation, 'there is no problem about connecting explanans and explanandum other than the general problem of understanding how metaphors are introduced and applied in their primary systems' (p. 175). Metaphors refer (in the strong sense of prediction)

> 'since the domain of the explanandum is redescribed in terminology transferred from the secondary system, the original observation language will both be shifted in meaning and extended in vocabulary, and hence that prediction in the strong sense will become possible. They may, of course, turn out not to be true, but that is an occupational hazard of an explanation or prediction' (p. 176).

Accordingly, metaphoric explanation is consistent with realism, not in the orthodox sense where theories refer to observation, but by referring to the properties and functions transferred from one domain to another. Metaphoric explanation is thus consistent with an interactionist relationship between primary and secondary subjects in which the perception of the primary subject is consistently shifting with novel metaphorical mappings, offering a view of scientific explanation consistent with a continuously expanding meaning of the explanandum, one that is not engraved in stone, but is as malleable as the social world.

Acknowledgements

I am grateful to Johanna Thoma and Nora Schwartz for helpful and insightful comments provided on the *Economic Thought* Open Peer Discussion Forum. Tamer Amin provided helpful commentaries on earlier versions. I also wish to thank the participants at the TINT seminar series (June 2014) and the AUB junior research leave funding under which most of this research was undertaken.

References

Andreoni, J., & Miller, J. (1993) 'Rational cooperation in the finitely repeated prisoner dilemma: Experimental evidence.' *Economic Journal*, 103, pp. 570-585.

Aumann, R. (1995) 'Backward induction & common knowledge of rationality.' *Games & Economic Behaviour*, 8, pp. 6-19.

Aumann, R. (1996a) 'Rejoinder.' In *The Rational Foundations of Economic Behaviour*, Proceedings of the IEA Conference held in Turin, Italy, E. Colombatto, M. Perlman, & C. Schmidt (eds). Palgrave Macmillan.

Aumann, R. (1996b) 'Reply to Binmore.' *Games & Economic Behaviour*, 17, pp. 138-146.

Aumann, R. (1998) 'A note on the centipede game.' *Games & Economic Behaviour*, 23, pp. 97-105.

Balkenborg, D. (1998) 'An experiment on forward versus backward induction.' Working Paper.

Basu, K. (1977) 'Information & strategy in iterated prisoner's dilemma.' *Theory & Decision*, 8, pp. 293-198.

Basu, K. (1990) 'On the non-existence of a rationality definition for extensive games.' *International Journal of Game Theory*, 19, pp. 33-44.

Becker, G. (1964) *Human Capital. A Theoretical & Empirical Analysis with Special Reference to Education*. Chicago: University of Chicago Press.

Ben-Porath, E. (1997) 'Rationality, Nash equilibrium, & backwards induction in perfect information games.' *The Review of Economic Studies*, 64, pp. 23-46.

Bicchieri, C. (1988a) 'Should a scientist abstain from metaphor?' In *The Consequences of Economic Rhetoric*, A. Klamer, D. McCloskey, & R. Solow (eds). Cambridge, New York, Melbourne: Cambridge University Press.

Bicchieri, C. (1988b) 'Common knowledge & backward induction: A solution to the paradox.' Behavioural Ethics Lab, University of Pennsylvania.

Bicchieri, C. (1989) 'Self-refuting theories of strategic interaction: A paradox of common knowledge.' *Erkenntnis*, 30, pp. 69-85.

Binmore, K. (1987) 'Modelling rational players I.' *Economics & Philosophy* 3, pp. 179-214.

Binmore, K. (1996) 'A note on backward induction.' *Games & Economic Behaviour*, 17, pp. 135-137.

Binmore, K. (1997) 'Rationality & backward induction.' *Journal of Economic Methodology*, 4, pp.23-41.

Binmore, K., & Samuelson, L. (1996) 'Rationalizing backward induction.' In *The Rational Foundations of Economic Behaviour*, Proceedings of the IEA Conference held in Turin, Italy, E. Colombatto, M. Perlman, C. Schmidt (eds). Palgrave Macmillan.

Binmore, K., Shaked, A., & Wolinsky, A. (1986) 'Testing non-cooperative bargaining theory: A preliminary study.' *American Economic Review*, 75, pp. 1178-80.

Binmore, K., McCarthy, J., Ponti, G., Samuelson, L., & Shaked, A. (2002) 'A backward induction experiment.' *Journal of Economic Theory*, 104, pp. 48-88.

Black, M. (1962) *Models & Metaphors*. Ithaca, NY: Cornell University Press.

Black, M. (1979 [1993]) 'More about metaphor.' In *Metaphor & Thought*, A. Ortony (ed). Cambridge: Cambridge University Press.

Boumans, M. (1999) 'Built in justification.' In *Models as Mediators*, M. Morgan & M. Morrison (eds). Cambridge: Cambridge University Press.

Boyd, R. (1993) 'Metaphor & theory change.' In *Metaphor & Thought*, A. Ortony (ed). Cambridge: Cambridge University Press.

Bronk, R. (2008) *The Romantic Economist. Imagination in Economics*. New York: Cambridge University Press.

Broome, J., & Rabinowicz, W. (1999) 'Backward induction in the centipede game.' *Analysis*, 59, pp. 237-242.

Brown, R. (1977) *A Poetic of Sociology. Toward a Logic of Discovery for the Human Sciences*. Chicago: University of Chicago Press.

Camerer, C. (2003a) 'Behavioural studies of strategic thinking in games.' *Trends in Cognitive Sciences*, 7, pp. 225-231.

Camerer, C. (2003b). *Behavioural Game Theory*. Russell Sage Foundation & Princeton University Press.

Clarke, J., Holt, R., & Blundel, R. (2014) 'Re-imagining the growth process: (co) evolving metaphorical representations of entrepreneurial growth.' *Entrepreneurship & Regional Development, An International Journal*, 26, pp. 234-256.

Cohen, I. (1994) 'Newton & the social sciences, with special reference to economics, or, the case of missing paradigm.' In *Natural Images in Economics*, P. Mirowski (ed). New York: Cambridge University Press.

Coleman, A. (1998) 'Rationality assumptions of game theory & the backward induction paradox.' In *Rational Models of Cognition*, M. Oaksford & N. Chater (eds). Oxford: Oxford University Press.

diSessa, A. (1993) 'Towards an epistemology of physics.' *Cognition & Instruction*, 10, pp. 105-225.

Emmeche, C., & Hoffmeyer, J. (1991) 'From language to nature—the semiotic metaphor in biology.' *Semiotica*, 84, pp. 1-42.

Fernandez-Duque, D., & Johnson, M. (1999) 'Attention metaphors: How metaphors guide the cognitive psychology attention.' *Cognitive Science*, 23, pp. 83-116.

Gale, J., Binmore, K., & Samuleson, L. (1995) 'Learning to be imperfect: The ultimatum game.' *Games & Economic Behaviour*, 8, pp. 56-90.

Gibbard, A., & Varian, V. (1978) 'Economic models.' *The Journal of Philosophy*, 75, pp. 664-677.

Gibbons, R. (1997) 'An introduction to applicable game theory.' *Journal of Economic Perspectives*, 11, pp. 127-149.

Gintis, H. (2009) *The Bounds of Reason*. Princeton: Princeton University Press.

Grüne-Yanoff, T., and Lehtinen, A. (2012) Philosophy of game theory. In *Handbook of Philosophy of Science*, U. Mäki (Volume editor). Elsevier.

Grüne-Yanoff, T., & Scheinzer, P. (2008) 'The role of stories in applying game theory.' *Journal of Economic Methodology*, 15, pp. 131-146.

Guala, F. (2008) 'Paradigmatic experiments: The ultimatum game from testing to measurement.' *Philosophy of Science*, 75, pp. 658-669.

Güth, W., Schmittberger, R., & Schwarze, B. (1982) 'An experimental analysis of the ultimatum game.' *Journal of Economic Behaviour & Organization*, 3, pp. 367-368.

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

Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., Gintis, H., McElreath, R., Alvard, M., Barr, A., Ensminger, J., Henrich, N., Hill, K., Gil-White, F., Gurven, M., Marlowe, F., Patton, J., Tracer, D. (2005) '"Economic man" in cross-cultural perspective: Behavioural experiments in 15 small-scale societies.' *Behavioural & Brain Sciences*, 28, pp. 795-855.

Hesse, M (1966) *Models & Analogies in Science*. Notre Dame: Notre Dame University Press.

Heyck, H. (2015) *Age of System: Understanding the Development of Social Science*. John Hopkins University Press.

Hodgson, G. (1995) 'Biological & physical metaphors in economics.' In *Biology as Society, Society as Biology: Metaphors*, S. Maasen, E. Mendelsohn, & P. Weingart, P. Kluwer Academic Publishers.

Hodgson, G. (1996) *Economics & Evolution: Bringing Life Back into Economics*. University of Michigan Press.

Hodgson, G. (2005) 'Decomposition & growth: Biological metaphors in economics from the 1880s to the 1980s.' In *The Evolutionary Foundations of Economics*, K. Dopfer (ed.). New York: Cambridge University Press.

Hoffman, R., Cochran, E., & Nead, J. (1990) 'Cognitive metaphors.' In *Metaphors in the History of Psychology*, D. Leary (ed). Cambridge, New York, Victoria: Cambridge University Press.

Johnson, E., Camerer, C., Sen, S., & Rymon, T. (2002) 'Detecting failure of backward induction: Monitoring information search in sequential bargaining.' *Journal of Economic Theory*, 104, pp. 16-47.

Kagel, J., & Ross, A. (1995). 'Introduction.' In *Handbook of Experimental Economics*, J. Kagel & A. Roth (eds). Princeton: Princeton University Press.

Keller, E. (2002) *Making Sense of Life*. Cambridge: Harvard University Press.

Khalil. E. (2000) 'Types of metaphors & identificational slips in economic discourse.' *Research in the History of Economic Thought & Methodology: A Research Annual*, 18, pp. 83-105.

King, J. (2012) *The Microfoundations Delusion. Metaphors & Dogma in the History of Macroeconomics*. Cheltenham: Edward Elgar.

Klamer, H., & Leonard, T. (1994) 'So what's an economic metaphor?' In *Natural Images in Economics*, P. Mirowski (ed). New York: Cambridge University Press.

Kreps, D., Milgrom, P., Roberts, J., & Wilson, R. (1982) 'Rational cooperation in the finitely repeated prisoners' dilemma.' *Journal of Economic Theory*, 27, pp. 253-279.

Kuhn, H. (1953) 'Extensive games & the problems of information.' In H. W. Kuhn & A. W. Tucker (eds). *Contribution of the Theory of Games*. Princeton: Princeton University Press.

Kuhn, T. (1979 [1993]) 'Metaphor in science.' In *Metaphor & Thought*, A. Ortony (ed). Cambridge: Cambridge University Press.

Lakoff, G. (1993) 'Contemporary theory of metaphor.' In *Metaphor & Thought*, A. Ortony (ed). Cambridge: Cambridge University Press.

Lakoff, G., and Johnson, M. (1980) *Metaphors We Live By*. Chicago: Chicago University Press.

Leary, D. (1990) *Metaphors in the History of Psychology*. David Leary (ed). Cambridge University Press.

Lakoff, G., & Núnez, R. (1996) 'The metaphorical structure of mathematics: Sketching the cognitive foundations for a mind-based mathematics.' In *Mathematical Reasoning: Analogies, Metaphors, & Images*, L English (ed). Hillsdale, NJ: Erlbaum.

Luce, D., & Raiffa, H. (1957) *Games & Decisions. Introduction & Critical Survey*. John Wiley & Sons, New York.

Mäki, U. (2002) 'The dismal queen of the social sciences.' In *Fact & Fiction in Economics*, Uskali Mäki (ed). Cambridge, Cambridge University Press.

Maasen, S., Mendelsohn, E., & Weingart, P. (1995) *Biology as Society, Society as Biology: Metaphors*. Kluwer Academic Publishers.

McCabe, K., Kevin, A., Smith, V., & Lepore, M. (2000) 'Intentionality & mind-reading: Why does game form matter?' *Proceedings of National Academy of Science*, 97, pp. 4404-4409.

McCloskey, D. (1983) 'The rhetoric of economics.' *Journal of Economic Literature*, XXI, pp. 481-517.

McCloskey, D. (1990) 'Storytelling in economics.' In *Economics & Hermeneutics*, D. Lavoie (ed). New York: Routledge.

McCloskey, D. (1994) *Knowledge & Persuasion in Economics*. Cambridge: Cambridge University Press.

McKelvey, R., & Palfrey, T. (1992) 'An experimental study of the centipede game.' *Econometrica*, 60, pp. 803-836.

Mincer, J. (1958) 'Investment in human capital & personal income distribution.' *Journal of Political Economy*, 66, pp. 281-302.

Mirowski, P. (1988) 'Rhetoric, mathematics, & neoclassical theory.' In *The Consequences of Economic Rhetoric*, A. Klamer, D. McCloskey, & R. Solow (eds). Cambridge, New York, Melbourne: Cambridge University Press.

Mirowski, P. (1994) 'Doing what comes naturally: Four metanarratives on what metaphors are for.' In *Natural Images in Economic Thought*, P. Mirowski (ed). New York: Cambridge University Press.

Morgan, M. (1999) 'Learning from models.' In *Models as Mediators*, M. Morgan & M. Morrison (eds). Cambridge University Press.

Morgan, M. (2002) 'Models, stories, & the economic world.' In *Fact & Fiction in Economics*, U. Mäki (ed). Cambridge: Cambridge University Press.

Morgan, M. (2007) 'The curious case of the prisoner dilemma: Model situation? Exemplary narrative?' In *Science Without Laws*, A. Creager, E. Lunbeck, & M. Norton Wise (eds). Durham: Duke University Press.

Morgan, M. (2012) *The World in a Model: How Economists Work & Think*. Cambridge: Cambridge University Press.

Morrison, M., & Morgan, M. (1999) 'Introduction.' In *Models as Mediators*, M. Morgan & M. Morrison (eds). Cambridge: Cambridge University Press.

Pettit, P., & Sugden, R. (1989) 'The backward induction paradox.' *The Journal of Philosophy*, 86, pp. 169-182.

Popper, K. (1972) *Objective Knowledge. An Evolutionary Approach*. Oxford: Clarendon Press.

Pulaczewska, H. (1999) *Aspects of Metaphors in Physics. Examples & Case Studies*. Walter de Gruyter.

Rabinowicz, W. (1998) 'Grappling with the centipede game: Defence of backward induction for backward induction-terminating games.' *Economics & Philosophy*, 14, pp. 95-126.

Richards, A. (1936) *The Philosophy of Rhetoric*. Oxford: Oxford University Press.

Reny, P. (1993) 'Common belief & the theory of games with perfect information.' *Journal of Economic Theory*, 59, 257-274.

Rosenthal, R. (1981) 'Games of perfect information, predatory pricing & the chain-store paradox.' *Journal of Economic Theory*, 25, pp. 92-100.

Roth, A., Vesna, P., Masahiro, O-F., & Shmuel, Z. (1991) 'Bargaining & market behaviour in Jerusalem, Lubljana, Pittsburgh, & Tokyo: An experimental study.' *American Economic Review*, 81, pp. 1068–1095.

Rubinstein, A. (2006) 'Dilemmas of an economic theorist.' *Econometrica*, 74, pp. 865-883.

Schultz, T. (1961) 'Investment in human capital.' *The American Economic Review*, 51, pp. 1-17.

Selten, R. (1975) 'Reexamination of the perfectness concept for equilibrium points in extensive games.' *International Journal of Game Theory*, 4, pp. 25-55.

Selten, R. (1978) 'The chain store paradox.' *Theory & Decision*, 9, pp. 127-159.

Smith, V, (2003) 'Constructivist versus ecological rationality in economics.' *American Economic Review*, 93, pp. 465-508.

Sobel, J. (1993) 'Backward-induction arguments: A paradox regained.' *Philosophy of Science*, 60, pp. 114-133.

Stalnaker, R. (1996) 'Knowledge, belief & counterfactual reasoning.' *Economics & Philosophy*, 12, pp. 133-163.

Sugden, R. (1992) 'Inductive reasoning in repeated games.' In *Rational Interaction*, Reinhard Selten (ed). Springer.

Sugden, R. (2002) 'The status of theoretical models in theoretical economics.' In *Fact & Fiction in Economics*, U. Mäki (ed). Cambridge: Cambridge University Press.

Watson, J. (2013) *Strategy: An Introduction to Game Theory*. W. W. Norton & Company.

SUGGESTED CITATION:

Mabsout, R. (2018) 'The Backward Induction Controversy as a Metaphorical Problem.' *Economic Thought*, 7.1, pp. 24-49. http://www.worldeconomicsassociation.org/files/journals/economicthought/WEA-ET-7-1-Mabsout.pdf