

A model intervenes: the many faces of moral hazard

Paper forthcoming in the Cambridge Journal of Economics

1. Introduction

The growing literature on social ontology and its relevance to economics has focussed on two distinct, but inter-related areas since it began in the early 1990s (Lawson, 1989; Mäki, 1992). First, there has been a lively debate about the links between ontology and methodology. Tony Lawson's methodological critique states that there is a tension between the employment of mathematical methods to analyse the economy and the nature of the social relations that they are used to explain (Lawson, 1997). Lawson and other critical realists have argued that the deductivist method favoured by mainstream economics presupposes the existence of a closed system made up of stable and predictable event-regularities (Fleetwood, 2002; Lawson, 2003; 2012; Lewis, 2004). This assumption of a closed system is deemed to be inapplicable in the social sciences and subsequent research has focussed critical attention on mainstream methods and the role that they play in delimiting and constraining academic research (Downward and Mearman, 2007; Lewis, 2006; Pratten, 2004).

The methodological critique is bound up with a second, more constructive, project involving the articulation and defence of a social ontology that treats the social world as an open system. Following Roy Bhaskar (1975 [2008]; 1979 [1998]), contributors to this literature explain the difference between the social and natural sciences by referring to the

fact that social systems are made up of intentional human actors with the power to make choices. As Lawson puts it, the very idea of human choice presupposes that actual patterns of events might have turned out otherwise (Lawson, 1997, 30). This observation also lies at the heart of Bhaskar's *transformational model of social activity* (Bhaskar, 1979 [1998], 27-70), in which the vehicle for social transformation is the intentional activities of individual agents who draw on, but are not fully determined by, the structural circumstances in which they find themselves (see also Archer and Bhaskar, 1998).

Thus, the first strand of ontological research points to the mathematical methods of economists and the potential costs of the increasing primacy of modelling in mainstream economics. The second strand points to the need for (one or several) ontologically grounded alternatives to the mainstream. But, whilst both of these strands remain fertile research areas in their own right, they raise unanswered questions. One area where such questions arise is the potential connections between mainstream modelling and the real world. Most ontologists place a significant emphasis on the transformative power of human ideas and activities in general, but say little about the transformative power of the economic ideas that they criticise. Given the increasing importance of technical ideas and concepts imported from mainstream economics in shaping policy and public discourse, this relative silence is a significant omission¹.

In this paper we build on the research in social ontology in order to address this omission. In so doing, we propose a novel understanding of how economic ideas affect reality. Earlier research on this question does not make use of the social ontology that we have outlined above. Instead, some see the use of economic ideas as part of a clash of competing ideologies whilst implicitly accepting the representative role of economic

models². Others – influenced by social constructionism – have defended the view that models define economic reality (Callon, 1998). Both of these approaches share an emphasis on the idea of economics as intervening in the real world rather than seeking to represent or describe it, but neither provides a systematic framework connecting this insight to the distinctive epistemic features of modern economics research.

The purpose of this paper is to devise a set of theoretical tools that explores how economic ideas interact with reality without committing to the thesis that these interventions are constitutive of it. In Section 2 we discuss the reliance of modern mainstream economics on modelling and explain its implications for the production of economic knowledge. In Section 3 we connect our analysis of mainstream modelling to the question of whether and how economic models represent economic reality. Our analysis emphasises the detachment of models from grand socio-economic theories and from real world events and problems. We identify three epistemic features – *specificity*, *portability*, and *formal precision* – that are essential to the production of analytical findings in economics. These features are used to explain how the analytical findings of a model are operationalized in policy. In Sections 4 and 5 we illustrate our framework through an investigation of the definition and application of the concept of *moral hazard* by economists. The deeper purpose of our illustration is to assess the usefulness of our interpretative scheme as a vehicle for analysing the influence of models on economic reality.

2) Modelling in Mainstream Economics

Philosopher Allan Gibbard and economist Hal Varian define a model as a composite of two distinct elements: a 'story' and a 'structure' (Gibbard and Varian, 1978, 666). The 'structure' element refers to the (uninterpreted) formal system and the 'story'³ refers to the specific economic interpretation of the logical machinery embodied in the structure. Philosophers and historians have been quick to point out that modern mainstream economics is defined by its use of models of this type (Fine and Milonakis, 2009; Lawson, 1997).

Economist's work was not always defined by the use of models in the modern sense. Econometric research produced early in the 20th century first introduced the term 'model' into economics (Morgan, 2012, 1-38). Initially the 'stories' told by modellers were an intermediate step between non (or partially) formalised economic theories (e.g. of the business cycle) and the phenomena that they described. Models were designed for "bridging the gap between theories of the business cycle and specific (time and place) statistical data of the cycle" (Morgan, 2008, 2). Subsequently, however, modelling took on a much broader role in modern economics. Rather than simply connecting general theoretical schemes and statistical data, models came to play a dominant theoretical role as well: "economists build models to depict some particular phenomenon or to figure out some puzzle or problem about a set of relations in the economic world" (Morgan, 2012, 387).

In this theoretical context, models are now the principal method by which economists hope to make analytical 'discoveries' such as the exposition of logical fallacies or the deduction of new principles from old premises (Hausman, 1984, 13). Mainstream modellers can justifiably focus on internal, mathematical characteristics of the model as a form of 'conceptual exploration' that is independent both of general theoretical schemes and actual economies (Hausman, 1992). Indeed, 'new style' economic theorists engage almost exclusively in modelling and see no reason to invoke or defend a viable distinction

between model and theory, regularly using the terms interchangeably (Leijonhufvud, 1997, 193). George Akerlof (1970) seen by many as the pioneer of the new style of economic theory, breaks new ground by defining his approach to model-building as one which eschews the study of deviations from a standard general theoretical scheme. Instead, he favours the elaboration of self-supporting models that are "... customised to describe the salient features of reality that describe the special problem under consideration" (Akerlof, 2003, 1). In this context models are freestanding: their 'structure' gives form to their argument and they relate to real world phenomena not through their connection to theories, but directly through their 'stories'.

As this model-based approach has taken hold, economists have tended to cluster around 'fields' that are delimited almost exclusively by the analytical tools that are admissible and the structural form of the models that they are composed of. Understanding the way in which these fields function and how they relate to older conceptions of theory is a matter for historical and empirical investigation which goes beyond this paper (see Fine and Milonakis, 2009; Milonakis and Fine, 2009; Walsh, 1987). However, Morgan and Knuuttila (2012) provide an illuminating illustration drawn from recent research in the field of industrial relations. Here a complex typology of games has developed in which,

"the proliferation of cases and the labelling activity [of the different types of games] suggests that we think of both these fields [referring to industrial relations and game theory] not as consisting of one general theory (of industry or of games) accompanied by an additional set of special cases, but as theoretical fields in which the main material consists of a carefully classified set of well defined models" (Morgan and Knuuttila, 2012, 77).

Morgan and Knuttila's emphasis on the taxonomic and puzzle-centric characteristics of theoretical fields is consistent with the conception of modelling advanced and defended by Akerlof.

Thus, models are not expected to appeal to an overarching theory that embodies a range of well-defined and standardised assumptions⁴. Instead, the model is identified with its logical and technical machinery (its structure), which may often be imported from mathematics or the natural sciences (Davis, 2006, 6-11). The structure imposes internal consistency on the model and defines its location within a relevant field. The field itself is a patchwork of models connected to each other principally by their shared methodological standpoint, but also (in mature fields) by path dependent processes of research citation and embedding into the teaching canon.

Fields may include models that reproduce relevantly similar stories, or target a particular empirical phenomenon, but the key to holding a field together is the acceptance and usage of a finite set of recognised 'analytical findings'. These analytical findings must have two basic characteristics: (i) they must be the conclusions of a deductive argument, thus ensuring their epistemic status; (ii) they must relate to an empirical or theoretical puzzle that is agreed to be relevant by participants in the field, thus ensuring their location in the patchwork. An example of what we mean by an analytical finding is the construction and use of moral hazard by mainstream economists. The genesis of this analytical finding in economics and the different types of models which have been constructed to exemplify 'moral hazard' problems is discussed in sections 4 and 5. Our analysis of moral hazard also shows that a model's analytical findings are first and foremost defined within the model framework that produced them. However, they re-appear when newer models develop, extend, build on, or further apply them to other scenarios. Armed with a clearer

understanding of what models are and how they are used we now turn our attention to their contested relationship with the phenomena that they are supposed to illuminate, analyse, explain or predict.

3) From representing to intervening

The existence of economic models outside overarching theoretical frameworks is not the only intellectual development that characterises this new style economics. There has also been a marked decline in the representative ambitions of economic modellers which has been the subject of heated debate amongst philosophers of economics.

Most commentators agree that the assumptions of a theoretical argument in economics (whether it's a model or a theory) need not be accurate representations of the phenomena under investigation (see Mäki, 2002; Morgan and Morrison, 1999). However, there is considerable disagreement over the extent to which the link between assumptions and reality can and should be severed. The most widely cited position in the debate is Milton Friedman's instrumentalist approach⁵, though many philosophers and methodologists have rejected Friedman's analysis. One plausible critique is that Friedman appears to conflate the reasonable claim that all assumptions must engage in some form of abstraction or simplification of reality, with the far less reasonable claim that all assumptions can or should be demonstrably false as descriptions of reality (for a good overview see Hausman, 1992).

Having demonstrated the weaknesses of Friedman's argument, however, many philosophers and methodologists have settled on a position that nevertheless takes the

representative ambitions of models as less crucial than was originally thought. Gibbard and Varian provide an early and illuminating example of this perspective:

"Within the class of theoretical models, we can distinguish between "descriptive" and "ideal" models. Descriptive models attempt to describe, *in some sense*, economic reality. Ideal models, on the other hand, are concerned with the description of some ideal case which is interesting either in its own right or by comparison to reality...." (Gibbard and Varian, 1978, 665)

This passage is revelatory because of the distinction that the authors draw between types of models and particularly their elaboration of a category of 'ideal models' that are unconnected to real world economic phenomena.

It is our contention that Gibbard and Varian's conception of ideal models challenges the commonsense notion of representation in the social sciences in favour of a method that relies on the construction of artificial 'realities' and the investigation of their mathematical characteristics. This is a view that was subsequently famously endorsed by the Nobel Laureate Robert Lucas:

"A 'theory' is not a collection of assertions about the behaviour of the actual economy but rather an explicit set of instructions for building a parallel or analogue system - a mechanical, imitation economy. A 'good' model, from this point of view, will not be exactly more 'real' than a poor one, but will provide better imitations." (Lucas, 1980, 697)

Similar positions have also received qualified support from influential philosophers of economics, whose analysis of modelling has tended to focus on rendering the opaque connections between imitations and reality clearer and more precise rather than questioning whether the construction of artificial realities is justifiable in the first place. Thus Gibbard and Varian (1978) suggest some models should be seen as ‘caricatures’, McCloskey (1983) develops the notion of models as ‘metaphors’, whilst Mäki (1992) sees them as ‘laboratories’, and Morgan (2001) elaborates on the suggestion that models are essentially ‘stories’.

Though they are different in content, these approaches (except perhaps McCloskey’s) seek to defend some ultimate, though often subtle and complex, connection between models and economic reality. Thus models can be seen as simplifying, isolating or emphasising certain aspects of reality in order to understand its underlying functioning. Crucially, though, a purified conception of the mechanisms at work can be constructed through a model that does not seek to represent at all. Instead a ‘credible’ explanation of the functioning of some part of economic reality can illuminate by describing an artificial world with relevantly similar characteristics (Sugden, 2002, 132-3).

There is, however, a minority position, to which we subscribe, which rejects the legitimacy of ideal models within economics. The critical realists who developed this approach (Lawson, 1997; Sayer, 1981) agree with the conventional view that a successful economic theory cannot engage in exhaustive description. They accept that economists must simplify by excluding features that fallible background assumptions suggest are irrelevant to the analysis of the problem at hand. They do not, however, accept that this is what economists do when they construct ideal models. Instead, they refer to the construction of economic models as (false) *idealisation* or the adoption of ‘convenient

fictions' (Lawson, 1997, 232-7). They further argue that, unlike abstraction, idealisation involves the introduction of assumptions that are known to be false by the investigator in order to facilitate the modelling process rather than to simplify a complex and messy reality.

It may help to consider the use of another abstract instrument we use in everyday life in order to grasp this argument. A good map will never be a complete description of the geographical area that it represents. It will exclude bushes, trees, hedgerows, park benches, and many other things besides, though they are relatively enduring features of the landscape. However, for a map to be useful in giving directions, it should not *misrepresent* the layout of the roads, or the location of parks, hospitals and airports. The key to this analogy is that the maps or models can be permitted to exclude elements that can be justified by either background theory or evidence as irrelevant, but they should not willingly exclude elements or invent fictitious ones in order to facilitate theorisation or tractability. According to critical realists, abstraction is the key to theorising in the social sciences, but idealisation is the hallmark of recent economic modelling.

To recap our argument thus far, we have claimed that the rise of modelling in economics has coincided with two related processes. First, the decline of general theories in favour of fields populated by patchworks of models that share a family resemblance and commitment to the centrality of a restricted set of analytical findings. Second, the gradual retreat from the idea of models as (partial or modified) representations of economic reality. As a result mainstream modelling, which combines these two trends, is accurately described by Lucas as comprising an "analogue system - a mechanical, imitation economy" (Lucas, 1980, 697).

This updated picture of modern mainstream models suggests that representing the economy is no longer their main objective. Consistent with this observation, economic

sociologists have more recently been suggesting that thinking of models as interventions in (rather than descriptions of) the economy is more promising (Callon, 1998; MacKenzie, 2006; MacKenzie, Muniesa and Siu, 2007; Santos and Rodrigues, 2009). There is considerable diversity under this broad rubric: some see models as self-referential speech acts (Callon, 1998), others see them as akin to technical innovations (Faulhaber and Baumol, 1988; MacKenzie, 2006; 2008), still others see them as specific policy implementations (Guala, 2001). However, the key idea implicit in all these approaches is that, by providing a tightly defined framework from which to interpret a given social situation, economic models do not describe their object, but they may transform it.

We contend that the basic insight of this emerging literature on models as interventions has significant merit. As we noted above, when models distance themselves from their representative function, they become more detachable from theoretical contexts and thus potentially more malleable as tools that can be used to intervene (either intentionally or unintentionally) in economic arrangements, as our discussion of moral hazard will show. This approach remains under-developed by academic economists, but two different strands of the literature are following this line of enquiry. One approach within the history of economic thought literature has been to investigate how recent economic ideas can inform and influence public opinion through popular journalism (Barber, 2011; Bennett, 2011; Mata, 2011; Uchitelle, 2011). Another, usually referred to under the heading of performativity, has been to investigate the socio-technical arrangements that permit models to produce and transform the institutional settings in which their analytical findings are applied (Callon, 1998; MacKenzie, 2006; 2008).

Possibly due to the prevalence of historical and ethnographic case studies in these literatures, neither of these strands of research has attempted to provide a general

theoretical lens through which to interpret the interventionist character of economic models. In the case of the history of economic thought literature, the tendency has been to see the problem as purely an issue of ideology, where economic journalism is seen as providing a distorted representation of reality that uses the purported objectivity and scientific status of economic findings to quell debate. In the case of performativity research, the situation has been reversed as researchers have argued that the distance between distorted representation and reality has narrowed to the extent that what we would call idealisations have become abstractions.

In this paper we develop a more theoretically grounded interpretation of the interventionist role for models, focussing on three epistemic features that make them both peculiarly attractive and dangerous as tools for shaping social arrangements: *specificity*, *portability*, and *formal precision*.

Specificity refers to the origins of the modelling exercise and its detachment from practical and empirical contexts. Once the assumptions of a model have been set, the model's technical apparatus delivers results detached from any real world problem context as there is no requirement for a deep understanding of the social context of the perceived problem. Unlike in the experimental sciences where a model's viability is connected to its applicability to empirical contexts, it is the distance of economic models from empirical contexts that ensures their applicability (e.g. game theory can be used to analyse telecoms auctions (see Guala, 2001) and the Israel-Palestine conflict).

This detachment is distinct from but related to the idea of portability. Portability refers to the idea that, given a relatively precise technical specification and puzzle structure, the same set of analytical findings can be transported across a range of theoretical and

empirical contexts, which can then be presented as manifestations of the same basic problem.

Finally, by 'formal precision' we mean the set of precise outcomes that are directly deduced from model environments in which all technical aspects are completely defined. Formal precision flows from the technical machinery of the model that gives both form and structure to the modelling exercise. This epistemic feature enables modellers to use different models in different ways, including: (1) as vehicles of discovery for analytical findings; (2) as tools for numerical analysis in these constructed 'mechanical worlds', which can then be compared with real world data or for practical or policy-related reasons; (3) for didactic purposes, as the formal language and deductive logic of the modelling exercise demarcates the limits of the discipline and becomes the only way to do economics.

In the remainder of this paper these theoretical proposals will be further articulated and put to use through a case study on the economic definition and use of moral hazard across a number of different real world cases.

4) Defining moral hazard

Moral hazard is a widely used concept within economics and has appeared in models in a variety of subfields. It can be defined as the "actions of economic agents in maximising their own utility to the detriment of others, in situations where they do not bear the full consequences or, equivalently, do not enjoy the full benefits of their actions *due to uncertainty and incomplete or restricted contracts* which prevent the assignments of *full* damages (benefits) to the agent responsible" (stress in original Kotowitz, 1987, 549). The genealogy of the term has also been the subject of a number of studies (Baker, 1996; Dembe and Boden, 2000; Rowell and Connelly, 2012).

The origins of the term lie in the insurance literature where it was used to describe the tendency of insured parties towards negligence once they are adequately covered in the event of loss. The insurance literature considers the following two terms: *moral hazard*, which is "the increase in probability of loss associated which results from evil tendencies in the character of the insured person...." (Rowell and Connelly, 2012, 18) and *morale hazard* which "results from the insured person's careless attitude towards the occurrence of losses" (Rowell and Connelly, 2012, 18). This distinction has not survived the transposition to mainstream economics. Mainstream economists have been keen to stress that their discipline cannot accommodate moral terminology and thus that the distinction between 'evil' tendencies and 'carelessness' cannot be defined in their analytical framework (Dow, 2010; 2012b). Appropriately isolated and atomised agents would simply try to maximise their expected utility given the model environment they are in.⁶ Economists are concerned with problems in which agents try to use privileged information to achieve advantageous market outcomes, and how this changes the structure of the market. In cases when market outcomes are suboptimal, economists may devise new types of contracts or novel ways of signalling and/or screening to solve the perceived problem. The key point here is that the moral quandary faced by insurers has been seemingly emptied of its moral content by economists.

Thus, in economic terms, moral hazard is seen as a problem of incentives between a principal and an agent. Nevertheless, each model explains the existence of moral hazard differently depending on the model's particular structure.⁷ This is because the behavioural responses associated with moral hazard are derived from the model environment that produces them. The fact that these responses are derivations, in our terms analytical

findings of the model, is significant because it underwrites their removal from the moral discourse of the insurance industry. Economists define agents as rational and compute their optimal courses of action as a consequence of the constructed model environment, not of their individual or collective value judgements.

This is typical of how analytical findings are created in economics. The insurance literature, with its morally loaded language, is explicit in its understanding of behavioural traits ('evil' character or 'careless' behaviour) and considers how different people may respond in their own way to existing insurance policies. The economists' reduction of personal traits to those of an isolated, atomised and calculative individual maximising a utility function recasts behaviour as the only rational response of the agent to his environment. This response is determined by the specific details of each model environment, which means that specific outcomes are valid or intelligible only within that model environment and are a direct result of the particular aspects of the mechanical world that it describes.

The way ideas are transported into economics and subsequently transformed is directly related to the first of our three epistemic features, *specificity*. Moral hazard is a good example of this as its transportation from the insurance literature required economists to empty the concept of all social context. Therefore, agent interaction is interpreted solely as a response to the specific incentive structure of a closed model environment. This de-contextualisation is a central feature of deriving analytical findings in economics. Specificity ensures that this new analytical understanding of how agents interact follows automatically, since agents with tightly specified capacities and objectives are driven towards a 'single exit' (Latsis, 1976). The result of this academic exercise is the creation of a family of models

which have an analytical affinity in that they are exploring similar model-worlds and can be said to be investigating problems of 'moral hazard'. How varied the application of these models can be in the real world, or, to put it another way, how broad the range of application of moral hazard in different real world situations is what we now turn to.

5) A model intervenes: moral hazard and policy

Economists have investigated incentive structures that give rise to moral hazard problems across a number of empirical fields and in a variety of stylized situations. In this section we will concentrate on three different situations where moral hazard models have been deemed to be very effective in framing the perceived real world problem.⁸

Perhaps the most interesting study to date of how economists' treatment of moral hazard may intervene in and transform social arrangements can be found in a study of workers' compensation schemes carried out by scholars in public health (see Dembe and Boden, 2000). The focus of the literature on workers' compensation is on how injured employees may overstate their injuries or file fraudulent claims. According to the authors, this academic debate within economics has had a demonstrable impact on policy by undermining public support for workers compensation programs (Dembe and Boden, 2000, 266). Dembe and Boden note that a significant part of the problem emanates from the manner in which economists analyse the problem of workers' compensation schemes. What actually determines the employees' decision to claim benefits is quite complex and may involve "a variety of economic, legal, psychosocial, and medical factors" (Dembe and Boden,

2000, 273). However, they claim that, "Economists' focus on increasing benefits as a supposed cause of employee claims reporting obscures this complexity and distracts us from gaining a fuller understanding of the workers' compensation program" (Dembe and Boden, 2000, 273).

The construction of convenient fictions to facilitate modelling does not only affect the results of the models, it also leads economists to ignore (or assume away) a host of other issues that affect compensation schemes, therefore reshaping the policy debate to fit their understanding of the situation.⁹ In this case this modelling framework ignores a host of issues that include workers' family situations, prevalent social norms, beliefs, and the physical or psychological effects of work and on-the-job injuries. Furthermore, the framing of worker compensation as a moral hazard problem has also crowded out other areas of data collection, for example how often injured workers do not file claims because they feel intimidated or are afraid of losing their jobs. Thus, the initial abstract model environment is supplemented by empirical data collected with that specific model in mind. The resulting narrative is then used to frame the public discussion and drive policy-making.

One of the defining features of the problem of moral hazard in economics is how *portable* it has been. It is not perceived to be a problem found only in one type of market, i.e. a problem of worker's compensation, therefore a specific issue of labour markets. Nor is it seen as a problem in the insurance market, for example when an insurance policy is drawn between a principal and an agent. As section 4 shows, any principal-agent interaction can exhibit this problem. However, the concept is more portable than even these broad limits allow. This is because what counts as an 'agent' and the specific type of interaction these 'agents' can have is equally broadly conceived. Therefore, principals and agents do not have to be individuals, but can be corporations or institutions. Furthermore, the actual

environment of this interaction can be almost anything, from the implicitly hierarchical nature of an employer-employee environment, to the client relationship between a purchaser and a seller of insurance, or the legal framework that guides the interaction of commercial banks with national governments. The ease by which analytical findings can be transplanted between scenarios has been central to its use in public debates in the current financial crisis, to which we now turn.

Moral hazard has been considered to be a key feature of the current economic crisis by the majority of mainstream economists. The crisis has been seen to consist of two separate but inter-linked stages. In the beginning there was the banking crisis, starting in 2008 in the US property and financial sector, that later led to a sovereign debt crisis in Europe as a number of peripheral euro area countries lost access to international financial markets and could not re-finance their debt. While the causes of the crisis remain contested within the academic community, a number of economists have identified moral hazard as a major component in both stages of the crisis (for a critical account see De Grauwe, January 2011).

This conventional wisdom states that the belief that commercial banks are too big to fail led bankers to take too much risk and overextend lending, believing that if these loans are in arrears national governments will have to bail them out. The same device can then be used to analyse the behaviour of national governments in monetary unions, where some governments may issue too much debt expecting to be bailed out by the solvent countries in the union if they cannot refinance it. Thus, in both scenarios, issues of moral hazard arise as the agents (bankers or insolvent states) have incentives to take actions that have adverse consequences for the other agents in the 'game' (national governments or solvent states).

A concrete example of how this academic discussion on ‘moral hazard’ entered the public arena and eventually formed the policy response to the crisis can be seen in the official documentation of the European Commission. In a current working document of the commission that intends to establish a framework for the recovery and resolution of credit institutions in the EU, the commission’s objective is “to help mitigate the disruption of financial failures and reduce moral hazard in the future”(European Commission, 2012, 4). Moral hazard is mentioned 27 times in various parts of the document. In this context it is defined as “the management, senior executives and shareholders of systemic institutions could on the basis of past evidence reasonable expect that while they stand to gain from the upside risk (profits) of their actions, society would have to cover the downside risk (losses)” (European Commission, 2012, 10). The authors of the working document further note that “research shows that such a skewed incentive structure within financial institutions is not only a theoretical proposition, but over time has a material impact” (European Commission, 2012, 10).

The problems that arise from reducing the banking crisis to a moral hazard problem have been addressed in detail by Sheila Dow (2010; 2012a). Dow notes that an implication of the aforementioned analysis is that regulation and banking practice should change so that commercial banks do not enjoy the guaranteed support of the state in the form of deposit insurance or as the lender of last resort, which is the aim of the EU commission proposal. Therefore, the banking crisis is perceived to emanate from a problem with the incentive structures that frame the relationship between clients, commercial banks and the public sector. However, Dow disputes the (modelling) assumption that all inter-personal interaction can be reduced to a specific incentive structure. Furthermore, she considers how replacing this with a framework of analysis that includes relationships of trust as the key to a

functioning financial system, completely transforms our understanding of the situation. She therefore proposes an alternative understanding of the situation and contrasts it to the mainstream approach:

"While mainstream theory would analyse trust effectively as the exercise of rational self-interest, the open system approach analyses it in terms of conventional understandings about the exercise of agency.... Moral hazard [in Dow's framework] is the danger that these understandings break down, eroding trust...." (Dow, 2010, 2-3).

Dow is not arguing in favour of a different modelling framework in which issues of trust and reciprocity arise due to repeated interaction by rational agents. More fundamentally, she argues that radical uncertainty is the core problem faced by market participants, and both lenders and borrowers need to build up a relationship of trust for the market system to work. This implies that the state is not merely a provider of incentive frameworks that would deliver favourable market outcomes, but must also ensure the necessary conditions for any type of market interaction to exist.

Current research on trust between individuals and institutions shows why trust is a complex issue that cannot be boiled down to a game theoretic framework between atomised agents (see Bachmann, 2001; 2011; Granovetter, 1985). Reinhard Bachmann distinguishes between interaction-based trust, which is trust that develops between repeatedly interacting agents, and institutional-based trust, which is trust that "*constitutively* builds on institutional arrangements"(Bachmann, 2011, 206). Following the work of (Luhmann, 1979) on trust, he argues that it is "trust *in* institutions"(Bachmann, 2011, 209) that is central in providing structure to the agent's world and reducing

uncertainty. He notes that this crisis “was not induced by individuals’ fraudulent ambitions but by inappropriate collective strategies and systemic effects”(Bachmann, 2011, 209). Therefore what is needed is to strengthen the trust market participants and the public has in institutions. This can be done by making institutions effective and fair in how they approach and resolve crises, rather than trying to find the incentive structure that will make atomised actors mechanically produce optimal outcomes.

Thus, casting the banking crisis as a moral hazard problem that boils down to misaligned incentives does not constitute a partial, abstracted or approximate picture of the social world, but a different picture altogether. Modelling itself is the lynchpin of this transformation. In order to interpret the banking crisis as a moral hazard problem economists must comprehensively define the model world in terms of risk-weighted potential outcomes within a structure of skewed incentives, whilst re-describing actors (individual and institutional) as rational maximisers who respond in a determinate way to specific stimuli. The creation of alternative institutions or agency structures with other capabilities that cannot be readily fitted to this incentive framework is not admissible within the confines of the model.

Nevertheless, economists have considered how manipulating existing institutional arrangements can solve the perceived 'core' moral hazard problem. Alternative institutional structures resurface regularly in the recent debates about the European sovereign debt crisis. These models have come to the fore since the 1990's, partly due to moves towards greater financial integration in Europe (in response to and feeding into policy debates). One well-known example is (Persson and Tabellini, 1996) that considers how risk sharing in a union can create problems of moral hazard. The authors find that in an institutional setting similar to that of the EU, intra-state transfers which create Union-wide risk-sharing schemes

lead to moral hazard, as states do not internalise how their behaviour on risk influences other states' welfare. This leads to under-provision in investment by governments. If governments did not participate in a risk-sharing scheme then they would undertake investment that would reduce the risk of an adverse shock in the economy, as they could not rely on financing from the other state. Thus risk-sharing schemes lead to equilibria that are inefficient in comparison to the optimal response that would arise if country governments fully internalised the effect a negative shock would have to the other country's finances.

This model environment develops a typical moral hazard narrative on how the merits of a risk-sharing union are undermined by the perverse incentives that the union creates for national governments and their voters. In this context, the exact institutional setting becomes important for the incentives that the different players have when they try to optimise their behaviour. It is also important because it demarcates what the different agents can and cannot do. This leads Persson and Tabellini to consider "which federal institutions are more desirable" (Persson and Tabellini, 1996, 624) within the modelling framework. Thus different systems of governance would lead to different economic outcomes, as they would provide the agents – in this case voters– with different incentives, leading to efficient or inefficient results. The economist's problem is how to find the system of federal governance that reduces incentive problems to a minimum. The latter institutional structure is then deemed to be optimal (see Persson and Tabellini, 1996, 644).

Current policy analysis of the Eurocrisis tends to mirror this line of argumentation. For example, John Muellbauer notes in a recent policy paper that the "eurozone is in an existential crisis" and summarises the problem as perceived by the economist, by noting that the Eurozone "needs to create the right incentives through a mixture of carrots and

sticks to enable the poorly performing economies to return to economic growth and to avoid a future existential crisis. It needs to discourage moral hazard and arrive at a fair distribution of burden sharing between taxpayers in different countries and holders of sovereign and bank bonds" (Muellbauer, October 2011, 1). Muellbauer translates the theoretical model into policy language, affirming the model's assumption that actual institutional arrangements are a mere proxy for the incentive structure of the relevant agents in a moral hazard setting. From a mechanical world without representative ambitions (the theoretical model), the economist has seamlessly produced a vehicle for assessing and transforming existing institutions.

In the three policy situations considered the economist's use of moral hazard reduces the complicated interactions between agents to a simple incentive problem. Any institutional process, from the labour market, to the credit market, or even intra-governmental institutional interaction or electoral systems, can be analysed as an expected utility maximising problem in which moral hazard can be shown to be present. In these three case studies we have seen three different forms of idealisation. With worker compensation schemes the complex relationship between employer and employee situated in a social environment is reduced to an incentive problem for filing claims. With the banking crisis complex issues of trust and fundamental uncertainty are reduced to rational responses of agents taking risky decisions. Finally, in the case of the sovereign debt crisis, entire institutions and their intricate constraining and enabling effects on individual action, are reduced to a game in which institutions perform very specific and pre-determined functions.

It is important to note how limited this view of institutions is. Institutions are not perceived to be able to change or evolve in response to a crisis as unforeseen situations

develop, and instead the focus when assessing institutional effectiveness or setting up new institutions is to make the institution's purpose and charter transparent, so that its scope for action is clearly demarcated. By making institutional responses rigid and predetermined, then, the model's perceived structure is replicated in the real world, with little recognition for the distance between the real world and that of the model.

Through this discussion we have attempted to demonstrate how the model world can intervene in a tangible fashion and influence real world phenomena not only by interpreting, but also by assessing and setting the public policy agenda. Our analysis and that of other critical scholars brings out the three epistemic features of models introduced in section 3. We have shown that specificity is central to the construction of moral hazard as an analytical finding. However, it does much more than that. By decontextualizing the finding and setting it as an abstract problem detached from reality it places the analytical finding effectively beyond empirical falsification. Any question as to why the results found in the model world do not correspond with reality can be dealt with by resorting to a discussion of the distance between the relevant modelling assumptions and the real world situation. Thus the general analytical finding, in this case the existence of moral hazard, is never seriously in dispute. Only the details of the model structure – and perhaps the specific type of moral hazard which may be relevant – are open to debate. What remains is that the real world problem is seen as some type of moral hazard problem. This fundamental insight frames the discussion, obliging the various participants to use the modelling terminology and apply the general structure of argument and approach that gives rise to this type of discourse.

An example of this can be seen in the proposal on financial stability prepared by the European Commission. The commission report, motivated by a wish to reduce 'moral

hazard', proposed bank resolution as a disciplining device against excessive risk taking by commercial banks. One of the issues considered has to do with raising the funds necessary for banks to be resolved in a way that financial stability is maintained. The finance industry in the public consultation of this proposal argued against a plan by the commission to raise funds for this resolution program from the finance sector, suggesting that existing funds at national level are adequate and there was no need for an EU wide fund. The industry also added that such a measure will "increase moral hazard" (European Commission, 2012, 59 and 232) as there would be "less incentive for market participants to police the system"(European Commission, 2012, 232). This brings to the fore an interesting contradiction: the new financial framework that should reduce one type of moral hazard, may actually be increasing another type. Beyond these disagreements fuelled by conflicting interests, the example demonstrates that the whole public discussion must fit within the broader modelling framework. It is this broader framework that delineates what can and cannot happen in the policy arena, and which arguments are admissible in policy discussions.

The story of specificity and the family of moral hazard problems that economists 'discover' in the real world, is closely related to the ability to transport 'moral hazard' between situational scenarios. This *portability* has two distinct uses in the policy arena (1) extending the domain of abstract theory to a host of real world situations and (2) giving a sense of continuity in the economist's view of the world. Since moral hazard is a problem of economics, any situational scenario that can fit into this framework is within the domain of economic analysis. Furthermore, economists can ex cathedra say something about this real world problem, propose solutions or sketch policy interventions, and even import

experience gained in another field of study where the same ‘perceived’ core problem was present.

This ability to transplant analysis between disparate real world problems is related to our third epistemic criterion of *formal precision*. What we have seen in our analysis of moral hazard is that the term combines a generic description of the real world situation with the ability to effectively transform the public debate to fit its framework. Moral hazard models use deductive logic and technical machinery of some sophistication, making them appear to share epistemic features with the natural sciences. The fact that the economist can show precise modelling outcomes for the abstract problem that he has set himself, and can, in many cases, collect data with this model in mind further tightens the perceived link between this theoretical approach and scientific research as carried out in the natural sciences. This technical sophistication adds scientific credibility to the mainstream perspective, thus crowding out potentially legitimate alternatives.

We have argued that all three epistemic features are key to understanding how analytical findings are created in economics, and how they enter and transform the policy debate in real world situations. This interpretive scheme was able to reconcile two remarkable features of the moral hazard story. These are (1) moral hazard has been seen to be a core tangible problem in a number of different real world situations and (2) it is perceived to be an abstract analytical problem found in highly stylised models of agent interaction in academic economics. Because economics is viewed as the most scientific of the social sciences, this detachment from reality is expected by the public. Economic findings are seen as emanating from a dispassionate scientific viewpoint that lies beyond interests and ideology.

This goes some way towards explaining the wide appeal of moral hazard explanations in the popular media and in policy circles in the three cases we investigated in this section. It also explains why alternative viewpoints that cannot fit into this framework are marginalised in both policy and public discussion. Dembe and Boden note this in their work on worker compensation where more complex views of the relationship between employer and employee that integrate sociological insights are marginalised. For commercial banks and the financial crisis we noted how the problem of fundamental uncertainty and trust in institutions was sidestepped as the discussion focused on misaligned incentives between the different agents. Finally, in relation to the euro sovereign debt crisis Paul De Grauwe describes how the use of moral hazard by the press has transformed public discussion in some European countries (see De Grauwe, January 2011). This has influenced the types of political solutions electorates and governments are willing to accept for resolving the crisis. The ramifications of this last intervention are even wider. The whole E.U. framework with its complex institutional structure and grand ambitions is now represented by part of the press and increasingly seen by part of the public as an arena for one-dimensional national interests, where national governments pursue well defined goals at the expense of each other.

6) Conclusion

It can be argued that economics has always had the potential to impose a view of the social world and suggest methods of intervention, due to the nature and objectives of the discipline. Any economic concept from marginal analysis to opportunity cost or scarcity could display aspects of our three defining epistemic features: *specificity*, *portability*, and *formal precision*. What is different, then, about modern modelling? Why do we need a novel

framework of analysis to understand the role of moral hazard as a device for intervening in the social world?

Part of the answer lies in the transformation that has occurred within economics, and was described in sections 2 and 3. We argue that the demise of general theories as the core analytical devices of economics and their replacement by fields populated by distinct models, creates a fundamental problem with the idea of 'intervention' itself. Models lead to specific policy recommendations derived from their structure, but there is no expectation that different analytical findings can be meaningfully synthesised, or even that they must be non-contradictory. It is in this setting that our three epistemic features tell us something different about current economic interventions as demonstrated by the use of moral hazard models in the three policy contexts described above. Moral hazard, defined and circumscribed by its formal setting, has been transported across policy contexts without reference to an overarching theoretical superstructure. There is also little or no attempt to explain why analytical findings are relevant within the policy context other than the abstract claim that the situation can be composed as a principal-agent problem.

Our illustration shows that the three epistemic features of specificity, portability and formal precision function synthetically to provide a framework for analysing how modern economic models intervene in the social sphere. Thus, the analytical findings derived from a model must display specificity, in order for them to be detachable from any specific situational scenario. Should this be the case, they can be applied to different scenarios, displaying portability. They also maintain the formal precision necessary for retaining their purportedly scientific status and delivering outcomes that can be used in policy analysis. We have shown how the idea of moral hazard displays all these characteristics and has generated key analytical findings used by economists in a variety of real world settings. In

this manner, moral hazard has been transformed from an explanatory device developed in the insurance industry, to a tool for economic policy intervention and the resolution of perceived problems in the social sphere.

References :

Akerlof, G.A. 1970. The Market for "Lemons": Quality Uncertainty and the Market Mechanism. *The Quarterly Journal of Economics*, vol. 84, no. 3, 488-500.

Akerlof, G.A. 2003. Writing the 'The Market for Lemons': A Personal and Interpretative Essay. http://www.nobelprize.org/nobel_prizes/economics/laureates/2001/akerlof-article.html
(last accessed 16/07/2013).

Archer, M.S. & R. Bhaskar. 1998. Critical realism: Essential readings. London, Routledge.

Arrow, K.J. 1963. Uncertainty and the Welfare Economics of Medical Care. *American Economic Review*, vol. 53, no. 5, 941-69.

Bachmann, R. 2001. Trust, Power and Control in Trans-Organizational Relations. *Organization Studies*, vol. 22, no. 2, 337-265.

Bachmann, R. 2011. At the crossroads: Future directions in trust research. *Journal of Trust Research*, vol. 1, no. 2, 203-13.

Baker, T. 1996. On the Geneology of Moral Hazard. *Texas Law Review*, vol. 75, no. 2, 237-92.

Barber, W. 2011. The Great Recession as a Great Enlightener. *History of Political Economy*, vol. 43, no. 2, 369-73.

Bennett, P. 2011. When a Crisis in Journalism Meets an Economy in Crisis. *History of Political Economy*, vol. 43, no. 2, 375-8.

Bhaskar, R. 1975 [2008]. *A Realist Theory of Science*. Abingdon, Routledge.

Bhaskar, R. 1979 [1998]. *The Possibility of Naturalism*. London, Routledge.

Callon, M. 1998. *The Laws of the Markets*. Oxford, Blackwell.

- Colander, D. 2000. The death of neoclassical economics. *Journal of the History of Economic Thought* vol. 22, no. 2, 127-43.
- Davis, J.B. 2006. The turn in economics: neoclassical dominance to mainstream pluralism? *Journal of Institutional Economics*, vol. 2, no. 1, 1-20.
- De Grauwe, P. January 2011. A less punishing, more forgiving approach to the debt crisis in the eurozone. *CEPS* no. no. 230, 1-4.
- Dembe, A.E. & L.I. Boden. 2000. Moral Hazard: A Question of Morality? *New Solutions*, vol. 10, no. 3, 257-79.
- Dow, S.C. 2010. Moral Hazard and the Banking Crisis. *working paper presented in FMM Conference, Berlin, 29-30 October 2010.*
- Dow, S.C. 2012a. Different Approaches to the Financial Crisis. *Economic Thought*, 1, 80-93.
<http://et.worldeconomicassociation.org/article/view/28> (last accessed 16/07/2013).
- Dow, S.C. 2012b. Economics and Moral Sentiments: The Case of Moral Hazard. In *Facts, Values and Objectivity in Economics*, eds. J.C. Caldas & V. Neves. London, Routledge.
- Downward, P. & A. Mearman. 2007. Retrodution as mixed-methods triangulation in economic research: reorienting economics into social science. *Cambridge Journal of Economics* vol. 31, no. 1, 77-99.
- European Commission. 2012. A Roadmap to Stability and Growth, Commission Staff Working Document, Impact Assessment, accompanying the document: Proposal for a directive of the European Parliament and of the Council establishing a framework for the recovery and resolution of credit institutions and investment firms and amending Council Directives 77/91/EEC and 82/891/EC, Directives 2001/24/EC, 2002/47/EC, 2004/25/EC 2005/56/EC, 2007/36/EC and 2011/35/EC and Regulation (EU) No 1093/2010. Brussels, 6.6.2012, SWD(2012) 166 final. http://ec.europa.eu/internal_market/bank/docs/crisis-management/2012_eu_framework/impact_assessment_final_en.pdf (last accessed 16/07/2013).

- Faulhaber, G.R. & W.J. Baumol. 1988. Economists as Innovators: Practical Products of Theoretical Research. *Journal of Economic Literature*, vol. 26, 577-600.
- Fine, B. & D. Milonakis. 2009. *From Economics Imperialism to Freakonomics*. London, Routledge.
- Fleetwood, S. 2002. *Critical realism in economics: Development and debate*. London, Routledge.
- Friedman, M. 1953. *Essays in Positive Economics*. Chicago, University of Chicago Press.
- Gibbard, A. & H.R. Varian. 1978. Economic Models. *Journal of Philosophy*, vol. 75, no. 11, 664-77.
- Gladwell, M. 2005. The Moral-Hazard Myth. The bad idea behind our failed health-care system. *The New Yorker*, August 29.
- http://www.newyorker.com/archive/2005/08/29/050829fa_fact?currentPage=1 (last accessed 16/07/2013).
- Granovetter, M. 1985. Economic action and social structure: The problem of embeddedness. *American Journal of Sociology*, vol. 91, 481-510.
- Guala, F. 2001. Building economic machines: the FCC auctions. *Studies in History and Philosophy of Science*, vol. 32, 453-77.
- Hausman, D.M. 1984. *The Philosophy of Economics: An Anthology*. Cambridge, CUP.
- Hausman, D.M. 1992. *The Inexact and Separate Science of Economics*. Cambridge, CUP.
- Kotowitz, Y. 1987. Moral hazard. In *The New Palgrave A Dictionary of Economics Volume 3*, eds. J. Eatwell, M. Milgate & P. Newman, 549-551. London, Macmillan.
- Latsis, S. 1976. A research program in economics. In *Method and Appraisal in Economics*, ed. S. Latsis, 1-41. Cambridge, CUP.
- Lawson, T. 1989. Abstraction, tendencies and stylised facts: a realist approach to economic analysis. *Cambridge Journal of Economics*, vol. 13, no. 1, 59-78.
- Lawson, T. 1997. *Economics and Reality*. London, Routledge.
- Lawson, T. 2003. *Reorienting Economics* London, Routledge.

- Lawson, T. 2012. Mathematical Modelling and Ideology in the Economics Academy: Competing explanations of the failings of the modern discipline? *Economic Thought*, 1, 3-22.
- <http://et.worldeconomicsassociation.org/article/view/25> (last accessed 16/07/2013).
- Leijonhufvud, A. 1997. Models and theories. *Journal of Economic Methodology*, vol. 4, no. 2, 193-8.
- Lewis, P. 2004. *Transforming Economics: Perspectives on the Critical Realist Project*. London, Routledge.
- Lewis, P. 2006. Applied economics and the critical realist critique. *Economica*, vol. 73, no. 290, 361-2.
- Lucas, R.E. 1980. Methods and Problems in Business Cycle Theory. *Journal of Money, Credit and Banking*, vol. 12, no. 4, 696-715.
- Luhmann, N. 1979. *Trust and Power: Two Works*. Chichester, Wiley.
- MacKenzie, D. 2006. *An engine not a camera: how financial models shape markets*. Cambridge MA, MIT Press.
- MacKenzie, D. 2008. *Material Markets: how economic agents are constructed*. Oxford, Oxford University Press.
- MacKenzie, D.A., F. Muniesa & L. Siu. 2007. *Do Economists Make Markets? On the Performativity of Economics*. Princeton, Princeton University Press.
- Mäki, U. 1992. On the Method of Isolation in Economics. In *Idealization IV; Intelligibility in Science*, ed. C. Dilworth, 317-51. Amsterdam, Rodopi.
- Mäki, U. 2002. *Fact and fiction in economics: models, realism and social construction*. Cambridge CUP.
- Mas-Colell, A., M.D. Whinston & J.R. Green. 1995. *Microeconomic Theory*. Oxford, Oxford University Press.
- Mata, T. 2011. Fractals in Economic Journalism. *History of Political Economy*, vol. 43, no. 2, 379-85.
- McCloskey, D., N. 1983. The rhetoric of economics. *Journal of Economic Literature*, vol. 21, no. 2, 481-517.
- Milonakis, D. & B. Fine. 2009. *From Political Economy to Economics*. London, Routledge.

- Morgan, J. 2013. Forward-looking contrast explanation, illustrated using the Great Moderation. *Cambridge Journal of Economics*, vol. in press. doi: 10.1093/cje/bes069, 1-22.
- Morgan, M. & M. Morrison. 1999. *Models as Mediators. Perspectives on natural and social science* CUP.
- Morgan, M.S. 2001. Models, stories and the economic world. *Journal of Economic Methodology*, vol. 8, no. 3, 361-84.
- Morgan, M.S. 2008. Models. *The New Palgrave Dictionary of Economics, Second Edition*. http://www.dictionaryofeconomics.com/article?id=pde2008_M000391 (last accessed 16/07/2013).
- Morgan, M.S. 2012. *The World in the Model*. Cambridge, Cambridge University Press.
- Morgan, M.S. & T. Knuuttila. 2012. Models and Modelling in Economics. In *Handbook of the Philosophy of Economics* ed. U. Maki, 49-88. London, North Holland.
- Muellbauer, J. October 2011. Resolving the Eurozone crisis: Time for *conditional* eurobonds. *Centre for Economic Policy Research*, vol. Policy Insight No. 59.
- Pauly, M.V. 1968. The Economics of Moral Hazard: Comment. *American Economic Review*, vol. 53, no. 5, 531-7.
- Persson, T. & G. Tabellini. 1996. Federal Fiscal Constitutions: Risk Sharing and Moral Hazard. *Econometrica*, vol. 64, no. 3, 623-46.
- Pratten, S. 2004. Mathematical formalism in economics: consequences and alternatives. *Economic Affairs*, vol. 24, no. 2, 37-42.
- Rowell, D. & L.B. Connelly. 2012. A History of the Term "Moral Hazard". *The Journal of Risk and Insurance*, vol. in print, 1-25.
- Santos, A.C. & J. Rodrigues. 2009. Economics as social engineering? Questioning the performativity thesis. *Cambridge Journal of Economics*, vol. 33, 985-1000.
- Sayer, A. 1981. Abstraction: a realist interpretation. *Radical Philosophy*, vol. 28, no. 2, 6-15.

Sugden, R. 2002. Credible worlds: the status of theoretical models in economics. In *Fact and Fiction in Economics*, ed. U. Mäki, 107-36. Cambridge, CUP.

Uchitelle, L. 2011. The Uses and Misuses of Economics in Daily Journalism *History of Political Economy*, vol. 43, no. 2, 363-8.

Walsh, V. 1987. Models and Theory. In *The New Palgrave A Dictionary of Economics Volume 3*, eds. J. Eatwell, M. Milgate & P. Newman, 482-3. London, Macmillan.

¹ Critical realists have recently started working in this direction, noting that while mainstream modelling cannot explain reality, it nevertheless affects “policy and broader models of thought, which then affect organisation and institutions within the economy” (Morgan, 2013, 4). Jamie Morgan (2013) explores how this interdependence influenced financial policy in the period prior to the current financial crisis.

² Tiago Mata notes an interesting distinction on how economics is reported in the press. He separates ‘economics as politics’ and ‘economics as science’ in order to mark the “ambiguous status of economic stories in newsrooms”(Mata, 2011, 382). Therefore, “at the *New York Times* one observes a distinction between the stories written for the national desk, which tend to be stories about politics or with a political cast, and those written for the specialist, finance, or business sections, which resemble reporting on science”(Mata, 2011, 383). Therefore, both the scientific claim of economics and its ideological bias are reconciled in this approach of analysing how the popular press reports academic economics.

³ The relationship between models, stories and other related concepts is complex and contested. For an up to date overview and discussion see (Morgan, 2012, 217-255).

⁴ This point has also been made by methodologists working on pluralism in mainstream economics such as Colander(2000) and Davis(2006). Though they do not focus on modeling per se, the authors have pointed out that referring to modern mainstream contributions as part of ‘neoclassical theory’ is highly problematic since almost any plausible definition of the basic assumptions of the latter would be violated by most ‘cutting edge’ mainstream contributions. Significantly, the centrality of mathematical modeling in economics is taken more or less for granted.

⁵ The debate goes back to Milton Friedman’s (1953) infamous defence of economic theory in *The Methodology of Positive Economics*, which launched the so-called ‘assumptions debate’. Echoing instrumentalist claims coming from the philosophy of science, Friedman argued that the depictions of agents and firms employed by economic theorists in their models should not be assessed as descriptive claims about the world. On this line of thought, economic models are not designed to pick out features of the world and explain their functioning. Economists should not strive to make them more complex or realistic, indeed, by doing so they may actually limit the utility of their models. Rather, Friedman’s defence of unrealistic assumptions saw the sole objective of economics as the production of accurate predictions of large scale patterns in human behaviour.

⁶ Dembe and Boden (2000) as well as Dow (2010) note this in relation to Arrow's (1963) paper and the academic discussion that ensued. (Arrow, 1963) is one of the first papers to formally introduce moral hazard in connection to health insurance. In the paper Arrow discusses the role of social conventions and professional ethics in shaping the relevant market outcome and reducing the problem of moral hazard. However, Pauly in an almost contemporary note on Arrow's paper writes "my analysis shows, however, that the response of seeking more medical care with insurance than in its absence is a result not of moral perfidy, but of rational economic behaviour" (Pauly, 1968, 535 also noted in Dembe and Boden, 2000, 263) and adds "the problem of 'moral hazard' in insurance has, in fact, little to do with morality, but can be analyzed with orthodox economic tools" (Pauly, 1968, 535).

⁷ This is noted in key economics textbooks. For example in *Microeconomic Theory*, by Mas-Colell, Whinston and Green, it is pointed out that "the literature's use of the term *moral hazard* is not entirely uniform" (Mas-Colell, Whinston and Green, 1995, 447 fn. 1).

⁸ This is not meant to be an exhaustive list of all uses of moral hazard in policy discussions. E.g. health insurance has been a field in which economists have identified moral hazard being a major problem since the work of Arrow (1963) and Pauly (1968). This has had a profound influence on how healthcare policies have developed in the US in the last 50 years. A critical view on how moral hazard has affected the discussion on universal health insurance in the US can be found in (Gladwell, 2005).

⁹ Morgan in his analysis of the period prior to the financial crises makes a similar argument noting that the framework of analysis driven by mainstream modelling created 'blind spots' on policymakers perceiving the underlying problems and understanding what is really happening. He notes how during this period of relative financial stability the focus of economists on price stability, "made an orientation on possible sources of financial system instability less urgent "(Morgan, 2013, 11). The two objectives seemed over time to have merged so that "insofar as there was price stability, there was also financial stability"(Morgan, 2013, 11).