CHAPTER 3

Performative Mechanisms

Carsten Herrmann-Pillath

3.1 Towards Constitutive Explanations in Economics

This paper advances the idea that the current literature on performativity can be put on a stronger methodological footing if it is combined with the literature on mechanisms in the social sciences. I think that what authors such as Donald MacKenzie or Michel Callon actually did in their seminal contributions is presenting thick descriptions of *performative mechanisms*. Yet, what is missing is a general conceptual framework that allows to extend these thick descriptions into analytical approaches to causal explanations of the observed phenomena. This framework is provided by the methodology of constitutive explanations. In this paper, I merely sketch a few bare bones of this. The core task is to relate the notion of performativity to established bodies of research in economics. I think that the pivotal notion is that of incentives working on a given set of preferences in order to generate a certain behaviour, which underlies the different kinds of mechanisms about which economists propose generalizations that aspire to assuming the form of universal laws. Against this idea, I present the view that incentives and preferences are embedded into performative

C. Herrmann-Pillath
Witten/Herdecke University, Witten, Germany

I.A. Boldyrev, E. Svetlova (eds.), *Enacting Dismal Science*,
DOI 10.1057/978-1-137-48876-3_3

mechanisms that generate behaviour, which implies that universal laws cannot be formulated, as performative mechanisms are local and contextualized in essence. In particular, performative mechanisms endogenize the causal loops between incentives and preferences, thus suspending the analytical independence between the two, which is the core condition for the possibility of generalizations about the causal link to behaviour.

Most economists maintain the implicit methodological stance of the 'covering law' benchmark for scientific explanations. This mostly means that given certain *ceteris paribus* conditions, economics can identify regularities in economic phenomena that are universal. As a consequence, economists also believe that there are certain causal mechanisms with most general scope that can be implemented in policy interventions to achieve a certain goal. For example, if economists can show that private property rights have certain universal efficiency features, they would recommend privatization as a standard policy under most circumstances. In practical applications, this might require to make the *ceteris paribus* conditions explicit, which would result in a much more detailed analysis of conditions of application, but without affecting the theoretical and methodological core.[1]

Consider as one example that I will further detail in Sect. 3.6 of this chapter, the 'legal origin' theories of corporate governance which claim that common law institutions are more efficient than civil law institutions in arranging for external finance. This claim is based on a conception of causality that underlies the design of pertinent econometric testing, suggesting recommendations for interventions that were rapidly picked up by organizations such as the World Bank. By this reasoning, a universal regularity was stated, and its application would follow the universal claim while judging certain conditions of application (such as the state of the court system). However, when this research was later scrutinized in more detail, it turned out that the causal parameters could not be defined in a de-contextualized way. For example, considering certain elements of corporate governance schemes, functional equivalences between apparently different elements in social practices were not identified, or the impact of extra-legal embedding determinants was overlooked. Once researchers try to catch these aspects in a more exact way, a principle of indeterminacy seems to hold: The causal relations in the econometrics vanish, and no universal regularity can be established anymore, resulting in an apparent trade-off between exactness and universality. The fundamental methodological issue that comes to the fore here is that of contextuality: Against

the background of the standard econometric approaches, the different parameters and variables are mutually and externally contextualized, so that in the end, one can certainly maintain the idea that there is a causal interdependence between them, but has to acknowledge that this causality cannot be covered by generalizations that hold for a larger number of cases.

There are different possible reactions to problems like this one, which easily crop up in many specific areas of economic research. One is to maintain the covering law stance (and mostly refer to protective arguments such as the *ceteris paribus* clause), another is to search for alternative methodological frameworks. In this paper, I argue that this framework is the notion of constitutive explanations, or the analysis of 'mechanisms'. The notion of mechanisms has made a rapid career in the social sciences recently, but has rarely been received in economics so far.[2] For sure, there is the notion of 'mechanism design', but this differs fundamentally from the approach of constitutive explanations. 'Mechanisms' in mechanism design are mathematical structures that identify certain rules of games that relate individual strategic choices with a social value function defined by the designer, aiming at achieving the social optimum while incentivizing all agents to reveal truthful information. These are equilibrium solutions that do not identify causal mechanisms in the real world. If these mathematical models are applied to design real-world institutions, they actually refer to what is a causal mechanism, yet without identifying this directly. In other words, the 'mechanism' is a mathematical structure that is projected on real-world mechanisms, but without firstly trying to identify those causal mechanisms by means of theory-driven empirical research.[3]

In comparison, the notion of mechanism in the social sciences has been received from the sciences, in particular, the neurosciences. Here, mechanisms are conceived as multi-level complex causal structures, with different levels being approached by different disciplines or disciplinary subfields. The covering law criterion has been questioned by many philosophers of science over the recent decades, not only for principled reasons but also for the empirical observation that beyond physics, most sciences do not meet this benchmark, at least in current practice. In our context, this is certainly true not only for the social sciences in general but also for another field that focuses on the explanation of human behaviour, the neurosciences.[4] Both areas have got into close touch recently via the emergence of a new field in economics, namely neuroeconomics. This is part of a broader movement to introduce science-based methods into economics,

behavioural and experimental economics. The question arises whether in such cross-disciplinary interactions, unification of methodological principles is necessary in order to achieve conceptual and empirical integration. Against this backdrop, the recent convergence of views about mechanisms in both the neurosciences and the social sciences is highly significant.

In this paper, beyond elaborating on the mechanism methodology in economics, I advance the additional thesis that in a general taxonomy of mechanisms, the specificity of the social sciences is that certain core mechanisms are 'performative'. I distinguish between two categories of mechanisms, causal mechanisms in general and performative mechanisms as a subset. The general notion of mechanism already includes the defining feature that mechanisms are productive: This means, the coming together of parts and levels in a composite structure generates effects that are novel in the sense of new combinations of causes and effects. Therefore, the notion of mechanism also plays an important role in evolutionary theories which aim at understanding the emergence of novelties: A neuronal structure is a mechanism that produces novel phenomena in the physical world. Mechanisms in the social sciences are a special case because the productive function has the additional property of performativity. A most important class of social mechanisms are institutions in the broadest sense, and, following Searle, we can approach these as being observer-relative facts.[5] A mechanism in the general sense is an observer-independent fact (neurons fire independently from observers watching them), whereas a social mechanism is an observer-relative fact in the sense that causes and effects are necessarily mediated by the cognitive or, more general, neuronal systems of the individuals whose interactions are part of the mechanism.

Stated in this way, it is important avoiding the conceptual short-cut that a performative mechanism is simply a more complex chain of causes and effects that includes physical phenomena of neuronal systems. This would define a reductionist position; constitutive explanations, however, are non-reductionist.[6] Neuronal systems enable symbolic behaviour, and so mediation means semiotic causation: An observer-relative fact is a fact that comes into existence because the productive mechanism incorporates signs that are produced by neuronal systems and that establish channels for information transmission between neuronal systems that are themselves physically mediated not by neuronal mechanisms but by signs (such as body movements, utterances, and artefacts). Semiotic causation enables performativity of social mechanisms. The crucial phenomenon in establishing semiotic causation is interpretation: The effect is constituted

by the interpretation of the receiver of the sign, and hence in principle independent from the original intention of the sender. This theoretical framework, I argue, allows to provide the analytical foundation for the phenomenon of contextuality that I introduce in the example of 'legal origin' theories.[7]

Subsequently, I continue with outlining the basic conceptual framework for analysing performativity in terms of semiotic causality. I show that this is reflected in the cross-disciplinary literature on the relationship between incentives and preferences, thus violating the standard assumption in economics that preferences and incentives are perfectly separable. I demonstrate that an incentive structure is performative, implying that there are no universal regularities that allow transplanting certain general models of incentive structures into different contexts while keeping the chain between causes and effects stable and uniform. Finally, I draw conclusions for the analysis of corporate governance mechanisms and the related incentive structures.

3.2 A Basic Conceptual Frame for Modelling Performative Mechanisms

In this section, I add more detail to the basic framework for analysing performative mechanisms. I start out from further clarifying the two elementary terms: 'mechanism' and 'performativity'.

- A mechanism-based explanation is a *constitutive* one in the sense that explanations are based on the analytical and empirical identification of causal processes that are specific to time and space, hence do not result in universal regularities ('covering laws'). Mechanisms operate under the constraint of universal laws, but for the explanation of the observed regularities, the universal laws are not sufficient: Under conditions of scientific and disciplinary specialization, therefore, particular disciplines such as the neurosciences focus on the identification of mechanisms as the primary epistemic goal. Mechanisms are complex as they include different levels of aggregation, and mostly are part of larger structures in relation to which the mechanisms are separated via boundaries; across these, further causal interactions occur which involve the inputs and the outputs of the mechanisms. The most important aspect of disciplinary methodological standards is how they delineate the criteria of acceptable mechanistic explanations and

for causal relevance. Generally, this defines a naturalistic ontology, structured according to the criterion which entities are seen as having causal powers.[8]

- In the social sciences, some mechanisms have the special property that human cognitive systems are involved which enable intentional actions towards other individuals that are based on cognitive states through which interpretations of those actions are mediated. As a result, mechanisms include those cognitive states, and a full explanation requires the reconstruction of the specific ways how a mechanism emerges from cognitively mediated interactions between individuals. I call this 'performativity', and the causal process involved is semiotic or is 'semiosis'. Hence, whereas in the sciences mechanisms are givens (such as the physical structure of neurons), in the social sciences mechanisms are part and parcel of a social ontology of observer-relative facts that is continuously being created and reproduced by the individuals involved in interactions. In a nutshell, chains between external causes as inputs and behaviour as outputs are always mediated via cognitive states, which are distributed across many individuals, and are thus also determinants of the external causes.[9]

As a consequence, a social mechanism always manifests what I call a 'triadic' pattern of causation, involving both physical interactions between individuals and semiotically mediated interactions which can be embodied in the same physical phenomenon, but need not be. For example, if an individual hands over a banknote to another individual, this is the cause of a behavioural effect, but the effect cannot be explained by the mere physical fact of moving the banknote in space by means of bodily movements. The banknote is a sign, and semiotic causation necessarily involves the physical movement; both causal modes are indispensable for producing the effect, namely, a particular action of the receiver of the banknote. At first glance, that would suggest that semiotic causation supervenes on the physical movement. But a brief reflection shows that this is not the case, unless one adopts a naive sender–receiver perspective on the relationship. In this case, one would assume that the sender has the intention to use the banknote as a sign, and that this meaning is transferred to the receiver together with the physical item. However, as economists well recognize for the case of money, this does not match with the way how banknotes actually adopt the role of a sign, as this is only constituted on the level of the collective of individuals who mutually recognize this sign. Hence, we

cannot apply the simple framework of efficient causality here that underlies the explanation of the physical movement of the banknote, although that might appear to be reasonable as long as we look at the interaction in isolation.[10]

Social mechanisms have a triadic structure that can be conceptualized as in Fig. 3.1, which depicts a semiotically mediated performative mechanism$_s$ in which a physical mechanism is a constituent part.[11] We consider a physical object *O* that is a cause in the mechanism that produces an effect. This is mediated via the physical structure of the mechanism$_p$ *M*. In our example, this is the physical transfer of the banknote. I emphasize that in analysing social mechanisms, the explicit treatment of physical aspects is indispensable, such as arrangements of individuals in space, temporal sequences of actions, the shape and properties of artefacts, or technologies of interaction.[12] However, the ultimate effect of the mechanism$_p$ is determined by the semiotic causation mediated via the sign *S* (mechanism$_p$ is a necessary but not a sufficient cause of the effect). This effect is relative to the interpretant *I* which refers to the sign *S*. Thus, between *O* and *I*, a relationship is established that is conventionally called 'meaning'. However, this is dependent on the embedding of the performative mechanism into a larger context, the social system, in which the effect in terms of the subsequent action has a function. Thus, semiotic causation establishes a conjunction of meaning and function. This is enabled by the role of the sign in categorizing the physical mechanism *M*. For example, the banknote has a particular value, and depending on the value,

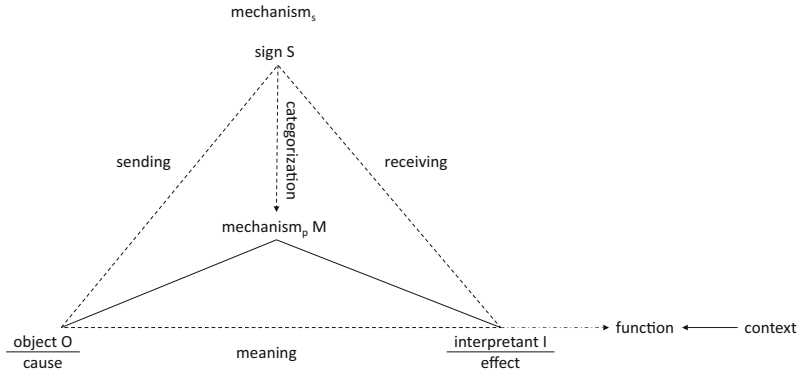

Fig. 3.1 Triadic causation in semiosis

a physically similar movement will lead to different effects in terms of subsequent actions of the interpretant *I*. The assignment of value is context-dependent, more specifically established by the usages of the sign in a population of interacting individuals.

I call the causal relationship between *O* and *I* 'bimodal', both physically and semiotically mediated. This turns *O* into an observer-relative fact, and is therefore, constitutive of the performativity of social mechanisms. It is important to emphasize that this does not depend on the individual intentionality of sign usages in social relationships: For example, the colour of the skin is a sign that operates independently from intentionally sending the sign. Further, the general category of 'sign' subdivides into different kinds, with different physical manifestations of the relationship between sign and object.[13]

I argue that performative mechanisms build on semiotic causation. Semiotic causation establishes performativity, which more specifically roots in the collective assignment of meaning to signs that trigger certain actions in a community of sign users. In this process, collectives of agents establish observer-relative facts, that is, social entities, and thereby, enrich the social ontology by creative acts. If we refer this view with the mechanism approach, we notice central elements, in particular, the identification of different levels (individual vs. collective) and objects (artefacts, embodied actions, individuals) and particular causal pathways that embed a mechanism into a larger unit (such as the embeddedness of the single action in a network of recurrent actions involving a collective). The triadic framework is just a most general conceptual structure guiding the more detailed analysis of performative mechanisms.[14]

3.3 Performing the Ultimatum Game: Which Way is the 'Right' One?

Analysing performative mechanisms as being based on semiosis has far-reaching methodological implications for understanding the relationship between external causes of actions and the actions that result from the causal impact. A covering law approach to incentives would assume that all human individuals are following the same principles of decision-making, so that a particular incentive structure would produce regular outcomes under *ceteris paribus* conditions. This is also the assumption that underlies the practical uses of 'mechanism design' theory: Then, results of game theory would be interpreted as stating universal laws across different

actualizations of human behaviour. Interestingly, in practical applications of mechanism design theory, many additional activities are necessary that fine-tune the contexts and even the behavioural stances of actors in order to make sure that they 'perform' the mechanism in an appropriate way. This is what we have to expect against the background of the triadic model.[15]

In the triadic model, semiotic causation makes the role of the context explicit, thus rendering arguments obsolete that would refer such necessary adaptations of models to the real world as taking place in the *ceteris paribus* domain or during the necessary tinkering in turning theory practical. Some well-known examples are the simple experiments such as the ultimatum game: In approaching models in terms of the covering-law methodology, human individuals would be expected to manifest similar behaviour, independent from their actual contexts in everyday life, allowing for random variations. If this cannot be proven, explanations would have to consider the *ceteris paribus* conditions. These conditions include states of knowledge in the experiment, such as the beliefs of the individuals (hence cognitive states). However, these beliefs can differ from the conceptualization of the experiment by the researcher, resulting in different interpretations. A common phenomenon is that the individuals subsume the experimental situation under familiar types of interactions outside the experimental setting. This can explain systematic variations across different groups of individuals which share certain social contexts that result in these beliefs. Hence, the experiment cannot be fully controlled by the experimenter, in the sense that she would be the conductor who fully determines the way how the experiment is performed. The individuals perform the experiment autonomously, and consequently, we cannot identify universal regularities over different applications of a standard experimental setting. This performance is mediated via the semiotic causation that is driven by the interpretive acts of the test persons: The incentive structure causes their behaviour, but incentives are simultaneously signs.[16]

The game-theoretic structure underlying the experiment describes only one part of the real-world causal mechanism that links the incentives (pay-offs) with the results. The game-theoretic description of the performative mechanism is incomplete. Contrary to the expectations of the experimenter, the cognitive states are not fully described by the information that the participants obtain from the description of the game provided by the experimenter. The question is how far we can say that their state of knowledge is simply 'false beliefs': This may become evident if

games are played recurrently, and individuals learn, so that they might finally converge to the 'rational' solution, hence perform the game according to the expectations of the experimenter. Yet, if this means de-contextualization in an artificial environment, we are not allowed to draw any conclusions about behaviour in a recontextualized real-life environment. Individuals simply perform the experiment differently. In the triadic framework, the experimenter achieves to impose another function on the interpretant *I*, which is to play the game properly; but this also changes the meaning of the sign, namely, the pay-offs.[17]

In the social sciences debate about mechanisms, this interdependence between cognitive states and outcomes of interactions is mostly referred to as 'self-fulfilling prophecy' in the sense of Merton's.[18] However, I argue that this is misleading, as the notion of self-fulfilling prophecy means that there is a belief that is initially false, but leads towards actions that changes the beliefs of others, resulting in further actions that ultimately confirm the original beliefs, thus rendering them truthful (like in the example of the bank-run triggered by wrong assessments of the financial status of the bank). If we call a mechanism 'performative', this is a much stronger proposition in the sense that the reference for the truth value of a belief is endogenous to the process even to the degree that it does not exist independently from the process under scrutiny. This reflects the creation of observer–relative facts via semiosis. The standard example for this is money, which does not exist before being performed collectively (to the opposite, the financial status of a bank in the bank-run does exist already).

Consider the case of the ultimatum game, again. Can we really say that playing the game 'correctly' turns originally false beliefs into true beliefs? This question touches upon a foundational issue in behavioural and experimental economics, namely the existence of social preferences in human individuals. Do subjects 'unlearn' false preferences in correct treatments of the experiment? Are these therefore 'errors'? Are we justified in judging the learned preferences as the 'true' ones? I argue that these questions do not grasp the real meaning of the ultimatum experiment, which in all the realizations actually involves the working of performative mechanisms.

3.4 Incentives Performing Preferences

Interestingly, the issue of human sociality is raised in many different disciplinary contexts, under different labels. What is 'social preferences' in behavioural economics is 'collectivism' or 'allocentrism' in psychology, or

'empathy' in the neurosciences. In the neurosciences, interest is directed at the question whether there are universal human (neuro)biological features that condition certain kinds of social behaviour. If that could be proven, this certainly would have implications for the other disciplines. In psychology, the issue is partly seen as an aspect of individual psychological properties, but in the context of collectivism, mainly as a cultural feature. This differs from the neuroscience question about universal human properties in hypothesizing that there are cultural properties shared by individuals belonging to the same social group which are stable in the long run; some groups might manifest higher degrees of collectivism than others. At first sight, both the neurosciences and the psychological approach would suggest that there is a set of fixed properties that would also be reflected in certain 'true beliefs' about the degree of other-mindedness among a group of human individuals. However, this expectation has not materialized.

In the neurosciences, research on empathy has resulted in a complex mechanism-based explanation that combines bottom-up and top-down processes, hence multi-level causal feedback loops. In a nutshell, there are certain species-specific neuronal mechanisms that enable other-oriented cognitive states and respective behaviours, but at the same time, these are only triggered under certain conditions which depend on higher-level states of knowledge mediated by symbolic systems, hence semiotically caused, in my parlance. In particular, this refers to cognitive categorizations of other individuals into in-group and out-group members, with no 'natural' delineation of the group in question. Although this behavioural tendency towards 'groupishness' is presumably a universal feature of humans, this fact alone cannot explain group boundaries under specific circumstances. Hence, one cannot identify universal regularities of human social behaviour rooted in shared biological properties but only complex mechanisms of empathy that involve different ontological levels and kinds. This result can be easily put into the triadic framework: The mechanism_p is the neurophysiological structure that is triggered by certain sensory inputs and generates certain bodily reactions, but these reactions can only be fully explained via the mechanism_s, that is, the semiotic categorizations. Emphatic behaviour has a function in the larger context of social systems, and there is no universal mechanism, as these functions are specific to particular social systems located in space and time. In other words, empathy is a performative mechanism that includes neuronal structures as mediating physical entities, but does not simply supervene on these structures, being

embodied in mechanisms that reach far beyond the brain, including social systems and the signs used therein.[19]

In psychology, suffice to mention one line of thought dubbed the 'ecological approach to culture'. This rejects the idea that people carry along certain inherent cultural characteristics such as 'values' that define their degree of 'collectivism', but posits that in social interactions, individuals take actions that create a certain environment unintendedly which incentivizes their behaviour in a way such that a stable pattern of interaction is achieved reflecting certain regularities which are context-specific (the 'niche'). This results in a conjunction between those incentives and cognitive states, that is only broken when fundamental parameters of the interaction are changed, especially involving the symbolic representations through which the interactions are mediated. As a result, what appears to be an internal 'value' explaining behaviour in terms of efficient causality (the value causes an action pattern), turns out to be a context-dependent regularity that is triggered by certain semiotic mechanisms. Once the experimenter achieves de-contextualization, the 'value' disappears.[20]

Both strands of research on human sociality therefore result in a general hypothesis about performative mechanisms that underlie social behaviour: Incentives, cognitive states, and symbolic media work together in generating performative actions that establish degrees of sociality specific to time and place. Therefore we cannot state a universal regularity about the degree of 'other-mindedness' of individuals. The 'mechanism' that is stated in the game-theoretic structure of the ultimatum game cannot be extended to a 'covering law' that is empirically meaningful. This also implies that we certainly can de-contextualize the behaviour of individuals by appropriate experimental settings, thus apparently producing evidence on the validity of 'rationality' as a universal human characteristic. However, this is only another instance of performativity, imposed by the experimenter.

Fortunately, a related view has been also articulated in economics recently, mostly under the heading of 'framing'. There is a large and growing literature showing that individuals do not have a fixed endowment with preferences that manifest a certain degree of 'sociality', but that 'social preferences' are endogenous to the context of a certain situation of choice, and in particular how choice is framed. In their comprehensive review of this literature, Bowles and Polanía-Reyes (2012) employ a general structure which directly matches with my model of triadic semiotic causation (Fig. 3.2). They argue that the standard economic view

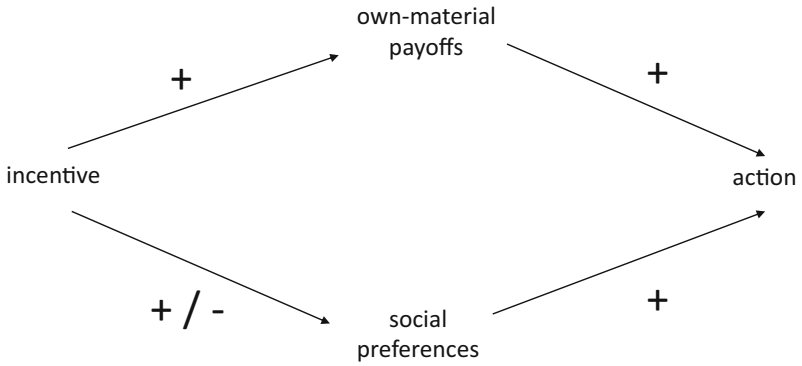

Fig. 3.2 Incentives and social preferences (following Bowles and Polanía-Reyes 2012)

posits the 'separability' thesis: Incentives work independently from the given level of social preferences in generating behavioural outcomes. To the contrary, there is ample evidence that incentives also change the level of social preferences, so that there can be both crowding in and crowding out in terms of the resulting behaviour. This results in a two-channel or, in my parlance, 'bimodal' model of causation that connects incentives and actions.

Bowles and Polanía-Reyes categorize the precise effects into four groups, always considering the introduction or the strengthening of a material incentive to perform a certain action.[21]

- The incentive changes the framing of an action such that self-interested motivation becomes salient and is seen as legitimate. This even applies for behaviour that does not involve social preferences at all: For example, introducing a fine on a certain behaviour can increase its frequency because the fine is interpreted as a price that is paid for allowing deviance.
- The incentive is interpreted as an information about the relatively low share of individuals with social preferences in a population, and increases uncertainty about the motivation of observed cooperative behaviours, that is, whether this behaviour is signalling social preferences. Therefore, the incentive decreases individual motivation to act socially.

- The incentive signals information of the designer about her assessment of the distribution of types in a population, and also gives information about how she perceives the nature of the action to be performed.
- The incentive reduces the sense of individual autonomy and therefore triggers resistance.

Resulting from these four constellations, there is no universal regularity between certain incentive structures and observed behaviour. We notice that economic research concurs with the aforementioned research in other disciplines, and suggest the interpretation that there is no fixed pattern of sociality in human individuals, but that sociality is based on performative mechanisms that result from the interaction between the environment of choice and distributed cognitive states in the group of interacting individuals, mediated by semiotic causation. Thus, incentives are always being interpreted, and interpretation is always contextualized. Scientific analysis, however, has to move beyond this general statement and needs to make the performative mechanisms explicit that result in particular causal chains between incentives and actions.

3.5 Semiotic Causation and Performativity of Incentive Systems

The economic problem of sociality raises an intricate question about the nature of incentives as signs. This is straightforward to grasp in the triadic model: The incentive as object *O* is at the same time the material embodiment of the sign *S*. Hence, there is a direct function of the incentive in creating the performative mechanism that produces the 'performance' of the incentivized individual. Bimodal causality works embodied in one single physical form. This is not a necessary feature of incentive systems but the most interesting case for our discussion.[22]

In terms of the economic conceptualization of preferences and choices, bimodal causality means that the incentive exerts causal effects via two channels. One channel is the direct effect on choice mediated by the given structure of preferences. The other channel works via the circumstance that this structure of preferences is also shaped by the effects of the incentive on cognitive states. This implies that preferences are not independent from the incentives. The following action is caused bimodally, ending

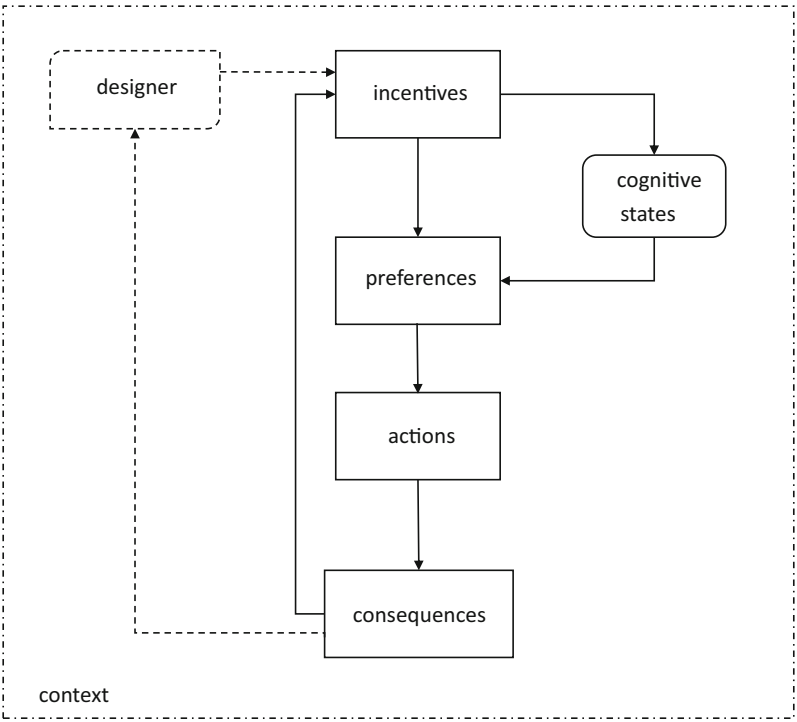

Fig. 3.3 Bimodal causation in incentive systems

up with certain consequences (pay-offs). These consequences work as incentives, in turn, thus resulting in feedback mechanisms. We end up with a recursive structure as pictured in Fig. 3.3. Here, I add the role of the designer of the system, because the interpretation of incentives certainly is changed fundamentally when interpretation also involves reference to the supposed intentions of the designer to create the system. Yet, this remains a simplified picture because I do not include the obvious mediating role of the designer's cognitive states in the perception of consequences and in her response.

How do the incentives influence the structure of preferences? Building on the previous summary of Bowles and Polanía-Reyes, I add some detail and emphasize the following.

- Under conditions of incomplete and imperfect information, all individuals are aware of the fact that their action consequences will depend on the behaviour of others; therefore, they need to construct beliefs about the preferences of others and their cognitive states. Then, the incentive system provides information about the composition of types of individuals in a population. This can directly involve reference to the designer of the system, because individuals may reasonably conclude that the designer of the system has designed it according to her information about that distribution. For example, as an employer, she may have even selected individuals according to those criteria. If the incentive system is geared towards individualistic behaviours, individuals may adapt their preferences to this information.[23]
- The qualitative nature of incentives influences the way how situations of choice are perceived, especially in terms of interactions with others. In particular, it is not warranted that monetary values are just approximations to underlying preferences. Monetary incentives change frames of valuations and even induce shifts across categories of valuations. For example, monetary expressions of value induce behaviour that is more individualistic, and they shift the reference frame towards an exchange context focusing on notions of reciprocity, in comparison with the direct benefits of action consequences.[24]
- Most generally, any kind of incentive system shifts the balance between intrinsic and extrinsic motivation. This effect is the more pronounced if the incentive system is cast in quantitative terms and can be anticipated in terms of the action consequences. If we consider the previous point, the effects of monetary incentives are actually twofold: one is the effect of the quantification and targeting of the incentive, the other is its explicit casting into monetary terms.[25]

The result of this analysis is unequivocal. Incentive systems are social mechanisms, and therefore, are performative. As performative mechanisms, they involve distinct elements, such as the roles of designer and actor, the physical entities that represent the incentive system (such as operating procedures), or the entities in which the incentives are embedded and embodied (such as numbers in bank accounts). These elements are causally connected in a bimodal way, mediated via semiosis, that is, interpretations on part of the actors. Finally, these mechanisms are embedded into the context of the social system, such as the organization that

implements the system. There is no way to predict the effects of certain incentives based on a universal regularity of causation. Causation is always semiotically mediated, and the same incentives can lead to very different behavioural responses, depending on the way how they are interpreted.

Incentive systems are important elements in the real-world design of economic institutions and organizations. The upshot of my analysis is that one cannot directly transfer results of generic economic models of incentive systems into particular social contexts. A real-world incentive system is a performative mechanism that extends beyond the part of the social system that is subject to the direct intervention of the designer of the system. In practice, this means that the designer would have to contextualize the incentive system. Naturally, this raises the question how to identify the relevant contexts. A case in point is the aforementioned issue of corporate governance institutions.

3.6 Implications for the Analysis and the Design of Corporate Governance Mechanisms

The case of corporate governance is of interest here, because it involves different forms of performativity. In the recent decades, corporate governance schemes have diffused across the globe that build on modern economic theory, such as principal-agent theory, transaction cost economics, and incomplete contract modelling. The problem driving these developments is the division between ownership and management in the modern public corporation: How can managerial behaviour be aligned with shareholders', that is, owners' goals? The theoretically grounded corporate governance schemes are themselves complex mechanisms, involving institutional regulations of the interaction between different groups and bodies of the corporation, and including a range of high-powered incentive systems for top-level management. These schemes originated in the Anglo-Saxon world, and are being promoted as best practices grounded in theory until today, with prominent cases as the ongoing reforms of corporate governance in Japan. One conspicuous feature of this diffusion is that on the one hand, the corporate governance schemes are seen as particular mechanisms that can be copied in other societal, cultural, and political contexts, and that on the other hand, these contexts are often blamed for dysfunctional performances of corporate governance. At the same time, however, there is no general agreement about the idea that the corporate governance schemes being promoted are factually the most

efficacious ones. One issue highlighted in the 'post-Piketty' world is the possible impact on rising inequality and a disconnection between managerial compensation and actual performance.[26]

Compared to the real-world use of mechanisms derived from generic reasoning in economics, according to the results of the previous sections, a mechanism in the domain of the economy would be conceptualized as a causal structure that involves a complex social ontology, namely constituent units and levels of the underlying performative mechanisms. For example, one would consider individuals, groups, or institutions as parts of the mechanism of corporate governance, as well as different specific mechanisms that establish their interactions, in particular focusing on semiotic causation. We would no longer approach a corporate governance mechanism based on de-contextualized economic models. Instead, we would reconstruct empirically the social ontologies that embed these mechanisms and show how they result in a certain pattern of 'performing performances', in the sense of performatively producing certain outcomes of actions.

In the first section, I have already referred to the 'legal origin' theories of corporate governance which are of special interest here as they start out from empirical research about determinants of external finance, and ground the interpretation of the results on economic theory.[27] This research posits a causal linkage between a set of variables that affect performance in terms of making external finance available at lowest costs and largest scope. At the same time, the argument operates on a higher level of aggregation in claiming that entire legal systems historically determined the emergence of these differences in performance, such as civil law versus common law. This is a multi-level analysis which was then tested statistically without making the underlying mechanisms explicit. When this research was put under closer scrutiny, serious empirical flaws became apparent. These flaws can be interpreted in terms of inaccurate and biased conceptualization, identification, and measurement of social mechanisms, in particular[28]:

- The list of potentially relevant elements of corporate governance mechanisms was incomplete;
- Functional equivalences of included elements were not properly identified; vice versa, different functions of similar elements were not recognized;
- Measurement criteria were implicitly referring to US conditions as benchmark;

- The boundaries of the mechanisms were not properly identified and the role of extra-legal factors in determining performance was not recognized.

If researchers improve the empirical approach, they reach the conclusion that, on the one hand, it is possible to explain improvements of corporate governance and their effects for single countries through historical times, but that at the same time, statistical regularities across countries are difficult to establish. This clearly indicates that the research factually resulted in the identification of mechanisms that are specific to time and place, but cannot achieve generalizations over time and place.

I take this result as an indication that corporate governance schemes are performative: The 'legal origins' theory failed to take account of the contextualization of elements of corporate governance. This compares with the development of mainstream theory of corporate governance that converged on a shareholder-value paradigm in the recent decades. Against the background of the aforementioned problem structure resulting from the separation of ownership and management in the public corporation, the proposed mechanism mainly focuses on directly activating the investor's valuations of companies on the capital markets in incentivizing the executive managers' behaviour via specific compensation schemes. This transition was part and parcel of a broader trend of 'financialization' of capitalist economies.[29]

This process has met a lot of criticism, both scholarly and public. As in the case of the 'legal origins' theories, which are closely connected to it, the idea is that one single mechanism is the most efficient one, and that the entire institutional set-up has to be geared towards this goal. However, if corporate governance mechanisms are performative, this assumption is not warranted. This point can be made in two different, though closely related ways. The first is to consider the interaction between corporate governance institutions and the economic system in general, the other is to scrutinize specific incentive systems which are parts of the corporate governance arrangements.

Without referring to performativity explicitly, Aoki's model of the cognitive division of labour is actually presenting corporate governance as a performative social mechanism.[30] Aoki argues that there are different ways how the knowledge is distributed in a company that determines its competitive success. In my framework, competitive success is a function, and the governance mechanism is designed to mediate between certain incentives,

behaviours, and this function. Aoki uses the term ‘cognitive assets’ and adopts the perspective of distributed cognition theorists: Cognitive assets are not strictly individualized but are distributed across individuals and technological artefacts.[31] This has implications for the distribution of decision rights. For example, if cognitive assets are mostly individualized on part of the managers, workers’ participation would introduce inefficiencies in incentivization, and vice versa, if workers’ cognitive assets are highly complementary to managers’ cognitive assets, a participatory corporate governance mechanism would be more efficient.

However, these interdependencies cannot be simply cast into a universal regularity in turn, because they are performative. Performativity emerges on different levels, thus revealing a complex structure of performative mechanisms. First, within the company, the distribution of cognitive assets is endogenous to the corporate governance scheme. For example, if workers are excluded from decision-making procedures in the company, they are lacking incentives for adopting skills and knowledge that are highly complementary to managers’ cognitive assets, and vice versa. Second, the distribution of cognitive assets is partly determined by other institutions in the economy, such as education and training and the labour market structures in general. The more portable skills are across companies, the lower are the incentives for forming company-specific or complementary skills. As a result of these and similar effects, similar levels of corporate performance can be achieved by different mechanisms of corporate governance. However, this is not simply a self-fulfilling mechanism, because performativity depends on certain determinants that remain givens, in particular technology. Even though technology can also be performative, the question is how far the temporal and spatial contiguities play together in enabling performativity. For example, the Silicon Valley Hi-Tech model typically even enables the outsourcing of the entire production process, because there are very low complementarities of cognitive assets between workers and high-skilled managers and engineers.[32]

Evidently, it is not possible to define a ‘one size fits all’ corporate governance scheme only taking theoretical analyses of the generic problems of principal–agent relations, asymmetric information, and so on into consideration. On the systemic level, corporate governance schemes become performative via the endogenous adaptation of the ‘cognitive division of labour’ in the economic system, thus changing the perception of the incentives that emanate from a certain corporate governance structure.

The other aspect to be discussed here is the incentive systems in the narrow sense, such as the stock option schemes. Here, we can directly apply the results of the previous section. Interestingly, it has been argued that the transition to those systems was also bolstered by management education, thus also establishing a possible case for the performativity of economic theory in conjunction with incentive structures. This argument runs in the following way. If education of managers is based on the mainstream theories about corporate governance and incentive schemes, as explicitly done in textbooks of managerial economics, students actually learn about the distribution of types at least in the environment of companies. So they will adopt the respective patterns of individualistic preferences. This confirms the expectations of the theorists and the designers of incentive systems, and therefore even empirically vindicates the underlying theories. This mechanism can be also supported by separating equilibria in sorting individuals with different levels of social preferences into different environments.[33]

Further, we can apply the entire range of particular hypotheses about the performativity of incentive systems here, such as regarding the trade-off between intrinsic and extrinsic motivation when extrinsic incentives are quantified and announced in advance, or on the priming effects of money on behaviour. In the light of these hypotheses, one can argue that the introduction of high-powered incentive systems that directly aim at individual behaviour will also change the regularities that are assumed to hold for this behaviour. Indeed, one would expect that the behavioural patterns become more similar to the assumptions of opportunism and individualist rationality.

If we take together these two perspectives on corporate governance schemes, the institutional and the incentive structure, we realize that such schemes are in fact complex performative mechanisms with a high degree of contextualization. One important consequence is that similar incentive systems can operate in a different way at different times and places. This applies also on different levels of analysis: A corporate governance scheme can operate under the contextualization of single companies, working for one case and failing in another, or can be contextualized on higher levels, such as referring to national-level institutions and culture. For example, the Japanese system was working well after World War II into the 1980s. Since then, strong pressures emerged reforming the system, theoretically conceived as convergence to the Anglo-Saxon model. However, the

process is slow and protracted, thus reflecting the complexity of the causal determinants. New mechanisms will also show idiosyncratic features, though of a different kind.

3.7 Conclusion

In this paper, I propose that the literature on performativity should be combined with the literature on social mechanisms in order to create a powerful approach to understanding and explaining the variance of human behaviour across different institutional contexts. This requires a fundamental shift in the methodological conceptualization of the economist's work, namely from a 'covering law' view to a 'constitutive explanations' view.

In more detailed work, constitutive explanations require the precise identification of levels of social ontology, kinds of particular mechanisms and types of social entities that are involved in a constitutive explanation. I have provided a few hints with the example of corporate governance schemes. One important consequence of this is that for the analysis of mechanisms in the economy, economics as it stands is not sufficient. There is a huge explanatory gap between economic theories and the derived models and the mechanisms that work in the real world. As in the reference case for constitutive explanations, the neurosciences, the analysis of economic mechanisms is multi-disciplinary, involving the entire range of the social sciences, and also disciplines such as psychology or biology. Therefore, the mechanism approach is also providing a framework of disciplinary integration. However, it is important to get the direction of the underlying theoretical effort right here: The theoretical achievement in terms of providing explanations of real-world phenomena is the identification and substantiation of a mechanism. It is the mechanism that defines the patterns of cross-disciplinary integration, and hence these patterns can differ across different mechanisms. We cannot achieve cross-disciplinary integration in directly linking theoretical premises and results of the different disciplines.

This is especially important when we consider the phenomenon of performativity. As has been amply demonstrated by experimental economics, human behaviour can be shaped by the proper establishment of mechanisms that trigger certain performances. As such, the experiments do not test given theoretical hypotheses, but are actually implicit instructions how to design real-world institutions in order to generate similar behavioural results. If we create incentive systems based on the assumption of rational

opportunism, we also create the agents that behave according to these predictions. By implication, the patterns of cross-disciplinary linkages are also endogenous to the application of the theories in the real world, again making any vision of a particular pattern of cross-disciplinary integration obsolete.

In other words, theorizing about human behaviour and social systems means to work on inventories of performative mechanisms, possibly resulting in taxonomies, historical and evolutionary trajectories. This is the true sense in which the economist becomes a naturalist. The first task of the economist-as-naturalist is to grasp the complexity and diversity of human behaviour in institutionalized contexts, and only then to work out regularities that apply across them. These regularities may in turn be rooted in universal laws, which apply for certain aspects and elements of the mechanisms that define the social ontology of the economy. In spite of being universal, however, they can only offer partial explanations of the causal processes that are mediated by complex performative mechanisms.

Notes

1. On the central role of *ceteris paribus* assumptions in the covering law approach to economic hypotheses, see Hausman (1992: 131ff, 2013: 14ff). Hausman argues that these clauses enable economics to maintain 'inexact laws'. Of course, the precondition is that there is a precise and reliable method to distinguish between acceptable and unacceptable c.p. clauses (on the complex questions here, see Reutlinger et al. (2015)).
2. For a comprehensive survey of this literature, see Hedström and Ylikoski (2010) or the volume edited by Demeulenaere (2011). This is mostly pursued under the heading of 'analytical sociology', but contributing strands of thought are broader in scope, including seminal works such as Elster (1989). Philosophically, an important pacesetter was Bhaskar (1989), although this example also shows how the reception in economics was ending in a heterodox cul-de-sac that left no impact on mainstream economics (so-called 'realism' à la Tony Lawson). For a rare reception of the mechanism methodology in economics, see Vromen's (2011) analysis of routines as multi-level mechanisms.
3. For example, an auction is a 'mechanism' in mechanism design theory with certain optimality features. In the real world, auctions often do not work as designers imagine. For example (see The Economist, August 29, 2015: 60), on the eBay website, the share of auctions has been declining continuously, partly because of the 'hassle costs' of auctions. One cause of these costs is

'sniping', when users wait until the last minute in order to submit a winning bid, thus disappointing other bidders. The experience of sniping drives many users away from auctions. Systematic research on empirical issues of mechanism design models include, for example, the experimental testing of the solutions of the hold-up theorem provided in the theory of incomplete contracts (Maskin and Tirole 1999; Maskin 2002), which were widely regarded to resolve this problem; Hart (2009) therefore declared it as obsolete for theories about contracting and the firm. Yet, this strong conclusion was only tested recently by experimentalists (e.g. Fehr et al. 2014; Erlei and Roß 2014). Erlei and Roß, for example, show that the sheer complexity of the theoretical mechanisms may give a role to bounded rationality in determining the experimental subjects' choices, which systematically and strongly diverge from the theoretical predictions.

4. Neuroscience research has been the most important object of studying mechanistic explanations in philosophy of science, with path-breaking contributions such as Craver (2007).
5. Searle (1995) distinguishes between observer-independent and observer-relative facts, thus assigning the ontological status of existence to both. For example, a tree is an observer-independent fact, a holy tree is an observer-relative fact. The property of 'holiness' can cause a change of state on part of the observer (such as fear), and thus exists in terms of having causal powers.
6. Craver (2007: 107ff, 233ff) distinguishes between the reductionist and the systems tradition in neuroscience, showing that in spite of the fact that many neuroscientists pursue a reductionist agenda, the field advances in developing multi-level integrative theories about complex mechanisms that produce a certain phenomenon in question.
7. Typically, reference to interpretation appears to entail hermeneutic approaches. However, even purely naturalistic theories of communication such as Aunger's (2002: 255ff) argue that communication in populations of agents communicating via signals cannot be viewed in the sender–receiver paradigm, but as a population-level phenomenon in which the effects of communication events on the receivers determine the meaning, and not the intentions of the senders. In fact, this amounts to the naturalization of Wittgenstein's approach to meaning. In my definition of a 'social' mechanism, I actually stay in line with Max Weber's definition of a 'social action' as being a type of action which intrinsically relates to actions of others.
8. This understanding of 'naturalism' follows Bhaskar (1989) and should not be misunderstood as 'physicalism', although the general assumption of physical closure of the world would hold (Papineau 2009). In Bhaskar's view, assigning causal powers to entities is constitutive for defining the ontology that underlies the design and testing of theories. So, in the social sciences constructing mechanisms is tantamount to creating a social ontology.

9. The current literature on performativity includes a range of different uses of the term, which my most general definition covers. One important strand is to investigate into the performativity of economic theories, following seminal contributions such as MacKenzie (2006). I treat a 'theory' simply as an instance of a cognitive state which is mediated via the artefacts that embody the theory (such as books and experimental devices). Callon (2007) extends this approach by including all ideas, practices, and devices that enable economic action, such as accounting practices. I have further expanded this approach to include materially mediated cognitive states in general (Herrmann-Pillath 2010, 2012a). This notion is also more general in not only referring to strong 'Barnesian' performativity in the sense of MacKenzie (2007) but also including all phenomena that relate to the emergence of collective intentionality via human interaction, in the sense of Searle (2010) or Tuomela (2007). This use is grounded in the original meaning of performativity in speech act theory.
10. In fact, this analysis is standard lore in the philosophy of language, referring to the overcoming of referential theories of meaning to rule-based theories which relate meaning to conventions and practices in communities of language users, for a survey, see e.g. Lycan (1999). Searle's theory of institutions transfers this fundamental shift of perspective to the analysis of institutions.
11. This diagram is a modification of standard graphic representations of Peirce's semiotics in biosemiotics, see, for example, El-Hani et al. (2006) or Salthe (2009). For a more detailed exposition, see Herrmann-Pillath (2012b). Peirce laid the ground for the analytical distinction between two modes of causality that are involved in social interactions, efficient and final; for a comprehensive discussion of Peirce's views on causality, see Short (2007).
12. This point corresponds to the revival of 'materiality' in sociology, see the seminal volume edited by Pinch and Swedberg (2008), which also plays an important role in performativity theory, partly reflecting the intellectual impact of actor-network theory that emphasizes the emergence of agency in networks of human individuals and artefacts (Latour 2005). This goes back to the origins in science and technology studies, where the physical location and structure of the laboratory is a central concern. In economic sociology, this has centred interest on the role of 'market devices' in enabling economic interactions (see the contributions in Callon et al. 2007).
13. Peirce's major contribution in creating the discipline of semiotics was to elaborate on a complex taxonomy of signs that starts out from studying the nature of the underlying mechanisms. For example, a sign can be embodied information, such as a facial expression signalling an emotion, or purely

conventional, such as a linguistic expression. For a survey of the Peircian taxonomy, see Short (2007).

14. It is straightforward to relate the semiotic model to sketches of mechanistic explanations such as Schmid (2011). Schmid distinguishes four steps in analysing social mechanisms: The explanation of the individual action, the analysis of the interaction patterns, the aggregation process, and the feedbacks between the aggregate level and individual action.
15. See Muniesa and Callon (2007) who give many examples of how theoretical models of game theory need to be supported by transformational measures aiming at the particular group of actors that are intended to perform the models in a particular context. This can refer to design of locations, design of forms of interactions, and also the training of participants. For example, in real-world spectrum auctions even the economists themselves who designed the mechanisms would be hired by participating companies to perform the mechanism properly.
16. This exposition summarizes the famous experiments in testing the ultimatum game predictions across a number of 'small scale' societies (Henrich et al. 2005). Deviations from the predictions of the model are a standard result which is mostly interpreted in a twofold way. One is to argue that humans are more altruistic than assumed by standard theory. This would be an alternative covering law approach in trying to substitute one universalization by another. The other is to include a learning dimension, showing that after some period of learning, experimental subjects will not commit the 'mistakes' anymore and produce the predicted result. The importance of the Henrich et al. study lies in showing up a third solution: This is that the response pattern is systematically influenced by socially embedded interpretations of the subjects (e.g. in societies with cooperative hunting offers in the ultimatum game would also be higher). I think that this also introduces a third alternative to Guala's (2007) methodological evaluation of experimental economics: He distinguishes between 'testers' and 'builders' and emphasizes that 'builders' aim at transforming the context of an experimental game in order to make subjects performing it. He thinks that this does not invalidate the predictions of the model, as one can see this procedure as an attempt to isolate behavioural determinants, which, after all, actually appear to work, given the setting of the experiment. The Henrich et al. experiments show that collectives of experimental subjects might systematically create autonomous 'performances' of the models. This is what is expected in the mechanism approach, such as argued by Little (1992) who champions the idea of medium-level theoretical conceptions in the social sciences, with limited reach in space and time.
17. Although the learning argument would be the most straightforward one in dealing with these issues from the viewpoint of standard economic theory,

in fact even this only works under special 'performative' conditions, see Camerer (2003: 59f). I pointed to Guala's (2007) assessment in the previous footnote. This problem is an aspect of the issue of external validity of experiments which is certainly taken very seriously by experimental economists. But this results in a very strong impact of basic convictions and intentions of experimenters on the actual empirical strategies and interpretation of results, which I would see as another instance of performativity, in this case with reference to the collective or community of researchers.

18. See Hedström and Ylikoski (2010: 61f). For the original contribution, see Merton (1948: 195): 'The self-fulfilling prophecy is, in the beginning, a false definition of the situation evoking a new behavior which makes the original false conception come true. This specious validity of the self-fulfilling prophecy perpetuates a reign of error. For the prophet will cite the actual course of events as proof that he was right from the very beginning.'
19. Singer and Lamm (2009) provide a concise statement of this interaction between top-down and bottom-up processes in triggering empathy. It is particularly interesting because this is a clear case where we cannot say that a mental phenomenon supervenes on a neuronal structure, because the mental phenomenon involves extra-somatic mechanisms. On the consequences of this research for economics, see Kirman and Teschl (2010). Interestingly, they point to experimental evidence that emphatic behaviour even varies for the same individuals depending on specific interactions with others. On the biological foundations of in-group/out-group distinction, see Bowles et al. (2003).
20. For a survey of the ecological approach, see Yamagishi (2012). In a large number of experiments Yamagishi has shown that Japanese subjects only appear to act more collectivistic than Americans if they receive certain contextual clues. If the context is entirely anonymous and de-contextualized, they often even act less other-oriented than Americans. This contradicts a long tradition in social psychology (e.g. Triandis 1995) in assuming that culture imbues individuals with certain sets of internalized values that explain certain behavioural patterns in comparison to people from other cultures (for a survey in the context of economics, see Beugelsdijk and Maseland 2010).
21. It is important to notice that economists normally treat all incentives as equivalent to monetary incentives; in the context of experimental economics, this is most explicitly done so, as monetary pay-offs are seen as indirect indicators of utility. Beyond economics, the notion of incentive is much broader and includes, for example, praise, awards, prizes, fame, and so on. As is well known from psychological research, these incentives can work very differently on motivation than material incentives that are directly

targeted at producing a certain level of activity. I come back on this point below.

22. There are very complex constellations that need further scrutiny in terms of Peircian semiotics, as I have previously mentioned (footnote 13). Basically, every incentive is a physical object, namely a physically embodied stimulus. There are incentives in which object and sign are physically united such as in the case of extending bodily caress to a person as a positive gratification. In other cases, the sign is separate, such as bestowing a medal on a person, where the original incentive would be the psychological and social states that are expressed by that sign. Money raises very tricky issues here, as economists, but also some neuroscientists treat money as directly reflecting the underlying utility. Although this is regarded as a technical device that is limited to certain experimental settings, the use of money is normally extended far beyond them. That would imply that money plays the intricate role of a culturally conditioned 'primary reinforcer', coming close to an oxymoron (see Camerer et al. 2005: 35). In standard economic theory, to the contrary, money is treated as a sign (a 'veil') that represents other underlying incentives, which are the things money can buy (see the discussion in Harrison 2008: 306f).
23. A concise argument on this has been presented by Sliwka (2007), compare also Falk and Kosfeld (2006). Psychologists have shown that in such a setting, there are many degrees of freedom: For example, when the incentive system signals the dominance of cooperative types, this might induce more people to free-ride (see Chen et al. 2005).
24. There is ample psychological evidence of strong framing effects of money, as in priming experiments; for a survey, see Vohs et al. (2006). Framing effects can be various and differentiated (e.g. even framing with clean or dirty banknotes can make a difference in behaviour); as exemplary studies, see Yang et al. (2012). This literature also refutes the typical assumption in experimental economics that money directly reflects underlying utilities, see Amir et al. (2008).
25. Since the seminal survey of Deci et al. (1999), the general notion of a conflict between extrinsic reward and intrinsic motivation is well accepted, but rarely received in economics, with exceptions such as Frey (1997) or Falk and Kosfeld (2006). Most importantly, for our discussion of managerial incentive system in the next section, these effects are especially strong if extrinsic rewards are very large and fall into the same category as the goal pursued by the incentivized action (James 2005). Bonus systems for managers should have an especially strong negative effect on their intrinsic motivation.
26. As one example of the discussion, see the influential work of Bebchuk and Fried (2004); on the public debates, see Joutsenvirta (2013). This has

already triggered many regulatory responses, such as imposing caps on bonuses and strengthening the role of shareholders in fixing remuneration schemes for CEOs. But all these measures so far do not question the fundamental principles of these schemes.

27. The seminal contributions were La Porta et al. (1997, 1998) and the generalization by Djankov et al. (2003). Interestingly, in the latter contribution, contextualization creeps back into the analysis because the performance of certain legal systems is seen as being dependent on the stock of civic capital.
28. For a devastating empirical critique of the legal origins theory, see Siems and Deakin (2010). Aoki (2010: 71ff) is a good survey of the discussion.
29. This refers to the phenomenon that the investors' perspective tends to dominate the entire institutional design of the corporate sector, hence also changing the strategic orientation of business towards financial goals, see Krippner (2005). In Herrmann-Pillath (2013) I offer an interpretation of this in terms of performativity theory, including aspects such as the accounting and financial reporting institutions or the patent system.
30. Aoki (2010: 26ff). Aoki's theory can be related to formal mathematical approaches to firm structure that point towards positive externalities between production factors, thus creating supermodular production functions, see e.g. Milgrom and Roberts (1990). Interestingly, in practical application of this thinking, 'high commitment human resource management systems' would not take the shape of high-powered extrinsic incentives, see Roberts (2004: 174f).
31. This is an important issue in social ontology: Economics just accepts the idea that cognitive states are confined to individuals, hence brains, whereas in recent developments of cognitive science these are seen as being partly externalized, see e.g. Clark (2011). Again, this can only properly appreciated in a mechanism view.
32. These arguments have been already deployed in Aoki's (1988) classical comparison between the A-Firm and the J-Firm. In principle, the two governance schemes in the USA and Japan achieved similar performance levels at that time, because they were embedded into different institutional structures of the labour market, different distributions and contents of skills, and so on, hence were contextualized differently. However, it is also partly a question of technology which scheme works best: At that time, Aoki pointed towards automotive industry as an example for a good match with the Japanese model, whereas chemical industry might be better governed by the US model. This observation points towards the possibility of 'counterperformativity' (MacKenzie 2007), which is an important phenomenon in rendering performativity theory empirically meaningful.

33. Ghoshal (2005) accused business schools in educating students to become opportunistic agents, based on advanced theory. It is important to notice that in the light of performativity theory, this does not need to suppose that students actually 'become' more opportunistic, which is probably not the case empirically (Guala 2007). Ghoshal's reasoning matches with the aforementioned theory by Sliwka (2007) in that the education transmits information about the composition of types in the economy, which triggers strategic responses of individuals. Such effects can be leveraged by endogenous sorting of different types across market and non-market domains of the economy, see Kranton (1996).

References

Amir, On, Daniel Ariely, and Ziv Carmon. 2008. The Dissociation between Monetary Assessment and Predicted Utility. *Marketing Science* 27(6): 1055–1064.

Aoki, Masahiko. 1988. *Information, Incentives, and Bargaining in the Japanese Economy*. New York: Cambridge University Press.

Aoki, Masahiko. 2010. *Corporations in Evolving Diversity: Cognition, Governance, and Institutions*. Oxford: Oxford University Press.

Aunger, Robert. 2002. *The Electric Meme: A New Theory of How We Think*. New York: Free Press.

Bebchuk, Lucian, and Jesse Fried. 2004. *Pay Without Performance: The Unfulfilled Promise of Executive Compensation*. Cambridge, MA: Harvard University Press.

Beugelsdijk, Sjoerd, and Robbert Maseland. 2010. *Culture in Economics: History, Methodological Reflections, and Contemporary Applications*. Cambridge: Cambridge University Press.

Bhaskar, Roy. 1989. *The Possibility of Naturalism. A Philosophical Critique of the Contemporary Human Sciences*. New York: Harvester Wheatsheaf.

Bowles, Samuel, and Sandra Polanía-Reyes. 2012. Economic Incentives and Social Preferences: Substitutes or Complements? *Journal of Economic Literature* 50(2): 368–425.

Bowles, Samuel, Jung-Kyoo Choi, and Astrid Hopfensitz. 2003. The Co-evolution of Individual Behaviors and Social Institutions. *Journal of Theoretical Biology* 223: 135–147.

Callon, Michel. 2007. An Essay on the Growing Contribution of Economic Markets to the Proliferation of the Social. *Theory, Culture & Society* 24(7–8): 139–163.

Callon, Michel, Yuval Millo, and Fabian Muniesa (eds.). 2007. *Market Devices*. Oxford: Blackwell.

Camerer, Colin. 2003. *Behavioral Game Theory: Experiments in Strategic Interaction*. Princeton, NJ: Princeton University Press.

Camerer, Colin, George Loewenstein, and Drazen Prelec. 2005. Neuroeconomics: How Neuroscience Can Inform Economics. *Journal of Economic Literature* 43(1): 9–64.

Chen, Xiao-Ping, S. Arzu Wasti, and Harry C. Triandis. 2007. When Does Group Norm or Group Identity Predict Cooperation in a Public Goods Dilemma? The Moderating Effects of Idiocentrism and Allocentrism. *International Journal of Intercultural Relations* 31(2): 259–276.

Clark, Andy. 2011. *Supersizing the Mind. Embodiment, Action, and Cognitive Extension*. Oxford: Oxford University Press.

Craver, Carl F. 2007. *Explaining the Brain. Mechanisms and the Mosaic Unity of the Neurosciences*. New York: Oxford University Press.

Deci, Edward L., Richard Koestner, and Richard M. Ryan. 1999. A Meta-analytic Review of Experiments Examining the Effects of Extrinsic Rewards on Intrinsic Motivation. *Psychological Bulletin* 125(6): 627–668.

Demeulenaere, Pierre (ed.). 2011. *Analytical Sociology and Social Mechanisms*. Cambridge; New York: Cambridge University Press.

Djankov, Simeon, Edward Glaeser, Rafael La Porta, Florencio Lopez-de-Silanes, and Andrei Shleifer. 2003. The New Comparative Economics. *Journal of Comparative Economics* 31(4): 595–619.

El-Hani, Charbel Niño, Joao Queiroz, and Claus Emmeche. 2006. A Semiotic Analysis of the Genetic Information System. *Semiotica* 160(1–4): 1–68.

Elster, Jon. 1989. *Nuts and Bolts for the Social Sciences*. Cambridge and New York: Cambridge University Press.

Erlei, Mathias, and Wiebke Roß. 2014. Bounded Rationality as an Essential Component of the Holdup Problem. *Working Paper*, Clausthal University.

Falk, Armin, and Michael Kosfeld. 2006. The Hidden Costs of Control. *American Economic Review* 96(5): 1611–1630.

Fehr, Ernst, Michael Powell, and Tom Wilkening. 2014. Handing Out Guns at a Knife Fight: Behavioral Limitations of Subgame-Perfect Implementation. *Working Paper* 171, University of Zurich.

Frey, Bruno S. 1997. *Not Just for the Money. An Economic Theory of Personal Motivation*. Brookfield: Edward Elgar.

Ghoshal, Sumantra. 2005. Bad Management Theories are Destroying Good Management Practices. *Academy of Management Learning & Education* 4(1): 75–91.

Guala, Francesco. 2007. How to Do Things with Experimental Economics. In *Do Economists Make Markets? On the Performativity of Economics*, edited by Donald MacKenzie, Fabian Muniesa, and Lucia Siu, 128–162. Princeton, NJ: Princeton University Press.

Harrison, Glenn W. 2008. Neuroeconomics: A Critical Reconsideration. *Economics & Philosophy* 24(3): 303–344.

Hart, Oliver. 2009. Hold-Up, Asset Ownership, and Reference Points. *Quarterly Journal of Economics* 124(1): 267–300.

Hausman, Daniel. 1992. *The Inexact and Separate Science of Economics*. Cambridge and New York: Cambridge University Press.

Hausman, Daniel. 2013. Philosophy of Economics. In *The Stanford Encyclopedia of Philosophy* (Spring 2013 Edition), edited by Edward N. Zalta. Accessed November 1, 2015. http://plato.stanford.edu/archives/spr2013/entries/economics/

Hedström, Peter, and Petri Ylikoski. 2010. Causal Mechanisms in the Social Sciences. *Annual Review of Sociology* 36: 49–67.

Henrich, Joseph, Robert Boyd, Samuel Bowles, Ernst Fehr, Herbert Gintis, Richard McElreath, Michael Alvard, et al. 2005. 'Economic Man' in Cross-cultural Perspective: Behavioural Experiments in 15 Small-Scale Societies. *Behavioral and Brain Sciences* 28(6): 795–855.

Herrmann-Pillath, Carsten. 2010. A Neurolinguistic Approach to Performativity in Economics. *Journal of Economic Methodology* 17(3): 241–260.

Herrmann-Pillath, Carsten. 2012a. Institutions, Distributed Cognition and Agency: Rule-Following as Performative Action. *Journal of Economic Methodology* 19(1): 21–42.

Herrmann-Pillath, Carsten. 2012b. Towards an Externalist Neuroeconomics: Dual Selves, Signs, and Choice. *Journal of Neuroscience, Psychology and Economics* 5(1): 38–61.

Herrmann-Pillath, Carsten. 2013. Performativity of Economic Systems: Approach and Implications for Taxonomy. *Journal of Economic Methodology* 20(2): 139–163.

James Jr., Harvey S. 2005. Why Did You Do That? An Economic Explanation of the Effect of Extrinsic Compensation on Intrinsic Motivation and Performance. *Journal of Economic Psychology* 26(4): 549–566.

Joutsenvirta, Maria. 2013. Executive Pay and Legitimacy: Changing Discursive Battles Over the Morality of Excessive Manager Compensation. *Journal of Business Ethics* 116(3): 459–477.

Kirman, Alan, and Miriam Teschl. 2010. Selfish or Selfless? The Role of Empathy in Economics. *Philosophical Transactions of the Royal Society, B* 365: 303–317.

Kranton, Rachel E. 1996. Reciprocal Exchange: A Self-Sustaining System. *American Economic Review* 86(4): 830–851.

Krippner, G. 2005. The Financialization of the American Economy. *Socio-Economic Review* 3(2): 173–208.

La Porta, Rafael, Florencio Lopez-de-Silvanes, Andrei Shleifer, and Robert Vishny. 1997. Legal Determinants of External Finance. *Journal of Finance* 52(3): 1131–1150.

La Porta, Rafael, Florencio Lopez-de-Silanes, Andrei Shleifer, Robert Vishny. 1998. Law and Finance. *Journal of Political Economy* 106(6): 1113–1155.

Latour, Bruno. 2005. *Reassembling the Social. An Introduction to Actor-Network Theory*. Oxford et al.: Oxford University Press.

Little, Daniel. 1992. *Understanding Peasant China. Case Studies in the Philosophy of Social Science*. New Haven: Yale University Press.

Lycan, William G. 1999. *Philosophy of Language. An Introductory Text*. London: Routledge.

MacKenzie, Donald. 2006. *An Engine, Not a Camera: How Financial Models Shape Markets*. Cambridge, MA: MIT Press.

MacKenzie, Donald. 2007. Is Economics Performative? Option Theory and the Construction of Derivatives Markets. In *Do Economists Make Markets? On the Performativity of Economics*, edited by Donald MacKenzie, Fabian Muniesa, and Lucia Siu, 54–86. Princeton, NJ: Princeton University Press.

Maskin, Eric. 2002. On Indescribable Contingencies and Incomplete Contracts. *European Economic Review* 46(4–5): 725–733.

Maskin, Eric, and Jean Tirole. 1999. Unforeseen Contingencies and Incomplete Contracts. *Review of Economic Studies* 66(1): 83–114.

Merton, Robert K. 1948. The Self Fulfilling Prophecy. *Antioch Review* 8(2): 193–210.

Milgrom, Paul, and John Roberts. 1990. The Economics of Modern Manufacturing: Technology, Strategy, and Organization. *American Economic Review* 80(3): 511–528.

Muniesa, Fabian, and Michel Callon. 2007. Economic Experiments and the Construction of Markets. In *Do Economists Make Markets?: On the Performativity of Economics*, edited by Donald MacKenzie, Fabian Muniesa, and Lucia Siu, 163–189. Princeton, NJ: Princeton University Press.

Papineau, David. 2009. The Causal Closure of the Physical and Naturalism. In *The Oxford Handbook of Philosophy of Mind*, edited by Ansgar Beckermann, Brian P. McLaughlin, and Sven Walter, 53–65. Oxford: Oxford University Press.

Pinch, Trevor, and Richard Swedberg (eds.). 2008. *Living in a Material World, Economic Sociology Meets Science and Technology Studies*. Cambridge, MA: MIT Press.

Reutlinger, Alexander, Gerhard Schurz, and Andreas Hüttemann. 2015. "Ceteris Paribus Laws", The Stanford Encyclopedia of Philosophy (Fall 2015 Edition), Edward N. Zalta (ed.), http://plato.stanford.edu/archives/fall2015/entries/ceteris-paribus/

Roberts, John. 2004. *The Modern Firm. Organizational Design for Performance and Growth*. New York: Oxford University Press.

Salthe, Stanley N. 2009. The System of Interpretance: Naturalizing Meaning as Finality. *Biosemiotics* 1(3): 285–294.

Schmid, Michael. 2011. The Logic of Mechanistic Explanations in the Social Sciences. In *Analytical Sociology and Social Mechanisms*, edited by Pierre Demeulenaere, 136–153. Cambridge: Cambridge University Press.

Searle, John R. 1995. *The Construction of Social Reality*. London: Penguin.

Searle, J.R. 2010. *Making the Social World. The Structure of Human Civilization*. Oxford: Oxford University Press.

Short, Thomas L. 2007. *Peirce's Theory of Signs*. New York: Cambridge University Press.

Siems, Mathias M., and Simon Deakin. 2010. Comparative Law and Finance: Past, Present, and Future Research. *Journal of Institutional and Theoretical Economics* 166(1): 120–140.

Singer, Tania and Claus Lamm. 2009. The Social Neuroscience of Empathy. *Annals of the New York Academy of Sciences* 1156: 81–96.

Sliwka, Dirk. 2007. Trust as a Signal of a Social Norm and the Hidden Costs of Incentive Schemes. *American Economic Review* 97(3): 999–1012.

Triandis, Harry C. 1995. *Individualism & Collectivism*. Boulder, CO: Westview Press.

Tuomela, Raimo. 2007. *The Philosophy of Sociality*. Oxford: Oxford University Press.

Vohs, Kathleen D., Nicole L. Mead, and Miranda R. Goode. 2006. The Psychological Consequences of Money. *Science* 314(5802): 1154–1156.

Vromen, Jack. 2011. Routines as Multilevel Mechanisms. *Journal of Institutional Economics* 7(2): 175–196.

Yamagishi, Toshio. 2012. Micro–Macro Dynamics of the Cultural Construction of Reality. A Niche Construction Approach to Culture. In *Advances in Culture and Psychology: Volume 1*, edited by Michele J. Gelfand, Chi-yue Chiu and Ying-yi Hong, 251–308. Oxford and New York: Oxford University Press.

Yang, Qing, Xiaochang Wu, Xinyue Zhou, Nicole L. Mead, Kathleen D. Vohs, and Roy F. Baumeister. 2012. Diverging Effects of Clean Versus Dirty Money on Attitudes, Values, and Interpersonal Behavior. *Journal of Personality and Social Psychology* 104(3): 473–489.