

ARTICLE

From dual systems to dual function: rethinking methodological foundations of behavioural economics

Carsten Herrmann-Pillath 

Max Weber Centre for Advanced Cultural and Social Studies, Erfurt University, Steinplatz 2, 90085 Erfurt, Germany

Email: carsten.herrmann-pillath@uni-erfurt.de

(Received 14 November 2017; revised 19 July 2018; accepted 12 August 2018; first published online 24 January 2019)

Abstract

Building on an overview of dual systems theories in behavioural economics, the paper presents a methodological assessment in terms of the mechanistic explanations framework that has gained prominence in philosophy of the neurosciences. I conclude that they fail to meet the standards of causal explanations and I suggest an alternative ‘dual functions’ view based on Marr’s methodology of computational neuroscience. Recent psychological and neuroscience research undermines the case for a categorization of brain processes in terms of properties such as relative speed. I defend an interpretation of dualities as functional, without assigning them to specific neurophysiological structures.

Keywords: dual systems theories; mechanistic explanations; dual functions; Marr

1. Introduction

In behavioural economics and psychology, one of the most common features of current theories about the human brain and mind is duality, such as of ‘impulsive’ versus ‘reflective’ processes or ‘critic’ versus ‘doer’. Researchers divide the agent into two domains assuming different roles in generating behaviour. In psychology, dualities often play a crucial role in explaining pathologies of behaviour, such as addiction (e.g. Wiers *et al.* 2007). In the specific context of behavioural economics, they are essential in explaining divergences between ‘rational’ and ‘irrational’ behaviour (e.g. Thaler 2016). Dualism is also reflected in some approaches to neuroeconomics which distinguish between a controlled and an automatic domain, however, these tend to be more complex (Camerer *et al.* 2005 offer a four-quadrant scheme with the two other dimensions of ‘affective’ and ‘cognitive’). However, one authoritative view on neuroeconomics rejects dualistic approaches since it builds on the reverse transfer of economic models to neuroscience (Kable and Glimcher 2007; Glimcher 2009, 2011), aiming at the reduction of the economic model of decision-making

© Cambridge University Press 2019. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution, and reproduction in any medium, provided the original work is properly cited.

to neuroscientific facts via the identification of neuronal correspondences to utility. It has been aptly labelled ‘neuroclassicism’ by Camerer (2013). This struggle between two fundamentally different methodological positions in neuroeconomics is absent in behavioural economics that builds on psychology, with Kahneman (2011) as the leading protagonist who presented the most influential and popular view on dualism.

In this paper, I aim at deconstructing dualist views to create a new methodological baseline that also helps to overcome these conceptual tensions. I will accomplish this task by proposing a shift from a dual system view to a dual function view. I activate two different philosophical resources. One is the celebrated framework of cognitive sciences developed by Marr which has also been invoked by Glimcher (2003) in his early design of a methodology for neuroeconomics. The other is the mechanistic approaches to the brain sciences, seminally systematized by Craver (2007). Combining these two strands of thought is contestable philosophically, given the ongoing discussion about whether Marr’s concept of computation can be reconciled with the mechanistic approach (overviewed in Rusanen 2014). My solution builds on applying evolutionary theory as a unifying framework, which stands in the tradition of evolutionary psychology. However, the main task of the paper is to look at the accumulating empirical evidence against the dual systems view, and to present a parsimonious and hands-on philosophical perspective on it that helps to synthesize the empirics in terms of the alternative concept of dual function.

Before entering the details of my argument, however, I need to pinpoint one motivation that seems essential in creating the strong preference for dualities in behavioural economics, but which I will put aside in my further methodological reflections, only coming back to it in the conclusion. The rise of behavioural economics has been driven by the discovery of a zoo of anomalies of rationality. However, what counts as ‘anomaly’ remains defined by the reference to the standard of rationality. This standard underlies the design of behavioural policies, too, as in the much publicised ‘nudging’ interventions (Thaler and Sunstein 2009). Yet, as has been diagnosed early, this creates a tension with normative presumptions of economics about sovereignty of the rational individual (McFadden 2006). Indeed, the discussion about the idea of ‘libertarian paternalism’ rages on, such as when considering the idea that effective nudging might require hiding the true intention of the intervention (Bovens 2008). Now, for legitimizing libertarian paternalism in the context of the normative commitments of economics, the dual systems model is a condition *sine qua non*. In this model, the behavioural policy is legitimate as it appeals to the ‘rational better self’ of the individual, thus justifying the idea that the individual will voluntarily and consciously agree with the aim of the intervention, both virtually and *ex post*. In other words, the behavioural economist as therapist collaborates with the autonomous rational self to resume control over the irrational self that takes harmful decisions for the individual.

Alternatively, in a unified model of the actor, for economists who stick to the standard normative assumptions regarding individual autonomy and consumer sovereignty, there are only two alternatives. As long as they maintain the model of the rational actor, irrationality results from deficits in the information available to the actor: Hence, the only permissible policy is improving information and designing institutional settings which overcome defects of information processing (Harrison and Ross 2017). If they accepted the empirical verdict that people are irrational and cannot

appeal to a 'better self' within the individual, this would justify any kind of authoritarian intervention, thus violating the normative standard of individual freedom.

In sum, I believe that the status of dual systems theories is foundational for modern behavioural economics, mainly for reasons that transcend the goal of testing theories empirically, and that therefore a methodological evaluation is indispensable. The paper proceeds as follows. I start with a summary of dual systems views in behavioural economics. Next, I evaluate these views in the light of the mechanistic methodology: This is the deconstructive step. I continue with grounding my concept of 'function' in Marr's three-level concept of explanation in cognitive science, which I combine with an aetiological notion of function as proposed in the context of teleosemantics and biosemantics. This is my constructive move: I restate dual systems as dual functions, taking Kahneman's distinction between 'slow' and 'fast' processes as a workhorse.

2. Dualities in psychology and behavioural economics

One of the most influential theories about psychological dualism that exerts a strong impact on economics is the idea that there are two systems in the brain, one impulsive and one reflective, or defined in related terms, such as 'automatic' versus 'controlled' (Strack and Deutsch 2004; Kahneman 2011). Many of the theories take it for granted that those different properties correlate (as Spunt 2015 points out), such as 'intuitive' with 'unconscious', 'automatic' or 'not controlled'. In psychology, alternative expressions include 'dual process' (other proposals include 'types of process', see Evans and Stanovich 2013) or 'dual route' (Hommel and Wiers 2017). The duality is often referred to basic structural features of the brain, specifically, the distinction between limbic and cortical regions (seminally, Trepel *et al.* 2005). Although the most influential author in this context, Daniel Kahneman, often emphasizes that the 'systems' expression is only metaphorical and heuristic, most researchers would expect that eventually a mapping of 'systems' on brain areas would be possible, matching with our current state of knowledge about brain modularity (Kahneman 2011: 366; compare Brocas and Carrillo 2014 in economics). For stylistic simplicity, I will use the term 'dual systems' in this paper as representing the entire range of variants.

All these dualities assume that there are two fundamentally distinct causal structures in the brain that generate human action:

- One is the 'impulsive', 'intuitive' or 'automatic' structure in which a certain stimulus or information generates a response without further thought and with high speed. Often this type of behaviour is referred to the domain of 'affect' or 'emotion'. For example, a relatively simple reaction is the bodily response to pain, but many other kinds of behaviour are included, such as intuitive evaluations of people whom we meet for the first time (perhaps indicating racial, gender or other stereotyping) or desires caused by subliminal cues.
- The other is the 'reflective' or 'intentional' type of causal structure in which we experience deliberation, planning or conscious control, consuming more time than the former causal structure, and calling upon more resources of cognition ('cognitive load').

In economic applications, the duality is often directly referred to as the notion of 'rationality'. For example, individuals form intuitive assessments of probabilities based on stereotypes (beliefs), which do not match with probabilities that would be calculated by reflective processes (logical structure). The latter is equated with the 'rational' solution.

Dual process theories play a central role in modern psychological research about addiction and other behavioural dysfunctions (e.g. Wiers *et al.* 2007). Here, the impulsive processes drive addictive behaviour, such as compulsive gambling, whereas the reflective system is assumed to lose control. As with Kahneman's approach, one essential causal factor is differences in speed: The impulsive system reacts fast, and the reflective system can only deal with the consequences of action already taken. The influential dual process model by Strack and Deutsch (2004) differentiates the two systems in terms of fundamentally different cognitive processes, with the impulsive system building on associations, and the reflective system on propositional knowledge and reasoning, hence mainly operating via the medium of language and 'syllogism'. This is increasingly also received in economics (Alós-Ferrer and Strack 2014).

Aspects of time differentials are also invoked in dual systems approaches to hyperbolic time preferences (Laibson 1997; McClure *et al.* 2007). Instead of assuming a coherent hyperbolic preference function (which is unitary), some authors posit that different selves are involved, which appears to be close to the duality of reflection and impulsive behaviour (and being extended to 'multiple selves' models, e.g. Jamison and Wegener 2010). So, when an action is still in a more distant future, the reflective processes prevail, and when it approaches the present, impulsive reactions gain control. There is a large economic literature that exploits these and similar ideas (seminal contributions include Benhabib and Bisin 2005 or Fudenberg and Levine 2006).

The duality of impulsive and reflective processes does not unequivocally distinguish between action and evaluation, as both processes result in common behavioural outputs. Therefore, we need to distinguish these theories from dualities that divide responses between instances of action and instances of evaluation and decision making. This duality can be found in many disciplines, such as in economics, with Thaler's distinction between a 'planner' and a 'doer' (Thaler and Shefrin 1981; Bénabou and Pycia 2002), and in artificial intelligence, with Minsky's (2006) distinction between a special function of 'critic' apart from the processes that generate action (although in Minsky's case, the 'critic' is a function that can be dispersed itself). The different approaches are highly diverse in detail, yet we can refer to a fundamental pattern in which valuation and often decision is conceived as being systemically different from action. Again, this can be projected onto neurophysiological structures, for example, when distinguishing between neuronal structures of valuation and sensorimotor circuits. For example, Damasio's theory of somatic markers is a case in point highlighting a separate neuronal encoding of valuations (Bechara and Damasio 2005; Reimann and Bechara 2010).

In relation to economics, an interesting variant of these theories is the application of a principal-agent framework to understanding the relationship between acting and evaluating (the seminal paper is Brocas and Carrillo 2008, though the idea was first ventilated by sociologist Coleman 1990). This would suggest that internal information asymmetries would be an important feature of human action.

Information asymmetries would assign a central role to learning in mediating the duality of evaluator and actor; deficiencies of learning have also been identified as the core process in explaining addiction (Redish *et al.* 2008). The notion of information asymmetries is implicit to a related duality which has also been received in economics. This is the distinction between ‘wanting’ and ‘liking’, hence two different kinds of valuation (Camerer 2006; Berridge 2009). For example, in explaining eating disorders we may assume that people want something that they later do not like. This is not coterminous with the distinction between impulsive and reflective systems, as both wanting and liking come along in the two variants of impulsive and reflective valuations (in the literature, the terms with and without inverted commas).

In summary, dual systems views are a most influential analytical paradigm on which the modern synthesis of psychology and economics is constructed. Yet, the protagonists are very careful in assigning a clear empirical status to them. This leads to variations in labelling as previously mentioned or assigning a mere heuristic status to them (Alós-Ferrer and Strack 2014 speak of ‘simplifications’, or of ‘mere collective names for broad classes of interconnected processes’). Yet, as we have seen, factual claims seem to go beyond that, and assignments to specific brain structures are regularly done. Therefore, it is legitimate to question how far the dual systems views could be vindicated by neuroscientific research and its methodology.

3. Dual systems in the light of the mechanistic approach to the neurosciences

3.1. Introducing mechanistic explanations

In recent efforts to build a methodological basis for the brain sciences, the mechanistic approach has become influential (Machamer *et al.* 2000; Craver 2007). We can conceive this in two different ways: One is as a positive description of what neuroscientists factually do, the other is to extend the mechanistic approach to a full-fledged normative theory about causal explanations in general (Craver and Tabery 2015). In this paper, I combine these two views: I raise the question whether dual systems can be reconciled with the research strategies that neuroscience normally employs, and I use the generalized conception of mechanistic explanations in developing my alternative approach. I hasten to add that I present one specific interpretation of the mechanistic approach, against the background of the ongoing critical debate (see e.g. Waskan 2011; Rusanen 2014).

The original motivation of the mechanistic approach was the diagnosis that neuroscientific explanations do not fit into the mould of deductive-nomological explanations, variants of which are mostly regarded as the standard of scientific explanations in general: Scientists generate testable hypotheses from theories that claim universality, and then evaluate empirical tests based on predictions to modify the theories. In the neurosciences (and the life sciences in general, see Bechtel and Abrahamsen 2005), researchers rarely employ theories of this kind. Experimental neuroscientific research is mostly about identifying specific mechanisms that generate observable phenomena. That means, researchers postulate a certain causal structure embodied in neurophysiological structures, and then try to find out whether this hypothesis can be empirically vindicated.

A mechanism is a grouping of parts that are organized in a peculiar way to produce a phenomenon. Accordingly, for a strict methodological approach it is necessary (Craver and Tabery 2015):

1. To give a clear and unequivocal identification of the phenomenon in question, which relates to the systemic level, and not to the parts alone (such as addictive behaviour of the individual as manifest in choosing among different alternative actions);
2. To identify the boundaries of the mechanism, i.e. which parts are included, and which are not (for example, does a causal explanation of addictive behaviour only include parts of the brain, or does it, as in psychosocial approaches, also include external entities such as peer-groups?);
3. To identify the parts of the mechanism, which might require dissecting composite parts into smaller constituent parts (for example, in neuroeconomic approaches to addiction one would move beyond higher-level brain structure to the identification of specific dopaminergic reward circuits);
4. To identify the organization of the parts, such as the different levels on which elementary entities and composite entities interact (for example, as in the previous case, identify the entire architecture of various circuits, their location and interrelation);
5. To identify the process dynamics that generate the phenomenon, especially, initial and termination conditions (for example, do we want to explain a single act of choice, or the development of a disposition to act that results in a series of actions of a specific type?).

An important question about the relationship between mechanistic explanations and deductive-nomological explanations is the underlying notion of causality, as many conceptions of causality require reference to universal laws (Dowe 2008). The most prominent approach is to adopt the interventionist and counterfactual notion of causality as seminaly elaborated by Woodward (2003) (there is a debate over this claim, see Woodward 2011): In this view, identifying causal factors works via constructing counterfactuals by which their determining role can be identified by varying forms of interventions into the phenomenon in question. This approach is especially appropriate in the context of the neurosciences, which mainly work via interventions in identifying mechanisms; further, the interventionist account is also congenial to clinical applications of neuroscientific knowledge. This would also match with the use of behavioural economics for designing behavioural policy interventions.

A central task in mechanistic explanations is to identify parts and levels the interaction of which produces effects that are not merely additive. This includes the possibility of emergent properties on higher levels which are not identical to properties of parts. The difference lies in the specific patterns of organizing the elements into the whole of the mechanism (Craver 2007 speaks of multi-level mechanisms). This is a controversial issue, on which I take a non-reductionist position: Higher levels manifest emergent properties, but this emergence is causally explained by the interaction of the constituent parts (this is the position seminaly developed by Bunge 1977, 1979). In this view, a mechanistic approach does not imply

that psychological constructs would be identical to properties that neuronal phenomena have, but that the emergence of the relevant properties can be causally explained on the neuronal level. In other words, a phenomenon such as ‘fastness’ involves neuronal mechanisms causally without necessarily, exclusively and unequivocally assigning this property to a specific type of lower-level neurophysiological process (this issue relates to the extremely large and rich literature on supervenience in the philosophy of mind; for an overview see McLaughlin and Bennett 2018, and in the context of mechanistic explanations see Harbecke 2014).

3.2. Assessing dualist theories from the mechanistic perspective

If we evaluate dualistic theories from the perspective of mechanistic explanations, we can start from the fundamental issue of distinguishing between psychological and neuroscientific explanations. One interpretation of dualities would simply be that these are psychological constructs, and that neuroscientific reduction is not necessary. For example, Alós-Ferrer and Strack (2014) think that the dualities relate to a ‘continuum’ or a ‘continuous variable’, which leaves open the question whether this refers to varying performances of the same mechanism or a mixture of different mechanisms, and seems to stay in tension with the idea that one system intervenes after the other system has already started to operate (the so-called ‘default interventionist’ view, Evans 2010, which is not followed by Strack and Deutsch 2004 who think that the systems operate in parallel).

A rigorous interpretation of this position would claim that, for example, the ‘reflective system’ is a psychological construct that has autonomous causal powers on this ontological level, and that it might relate to many neurophysiological aspects of the brain in a very complex way, in the sense of being an emergent property that cannot be and need not be assigned to a specific part and neurophysiological structure. In other words, for explaining observed behaviour, it is not necessary to move to lower analytical levels: This position corresponds to the views of economists who ponder what additional explanatory power can be generated from the neurosciences (e.g. Bernheim 2009), and who may even reject any relevance (Gul and Pesendorfer 2008).

However, most protagonists of the dual systems view declare that the systems somehow relate to specific parts of the brain. In psychology, one reason is the need to design clinical interventions: For example, manipulating certain psychological phenomena, identified as dysfunctional, by pharmacological means requires the identification of lower-level mechanisms with which the pharmaceutical substance would interfere. Pragmatic interventionism requires deeper-level mechanistic analysis, unless it also remains on the level of the psychological constructs (for example, applying psychotherapy for addiction). In this sense and corresponding to the previous remarks on Woodward’s notion of causality, clinical approaches need to apply a mechanistic perspective at least implicitly. Considering behavioural economics, approaches such as nudging would also work without further dissecting lower-level mechanisms, which explains the somewhat loose connection between behavioural economics and neuroeconomics.

For neuroscientists and neuroeconomists, relating psychological constructs to neurophysiological structures and processes is a central goal of research. Accordingly,

dualities must refer to certain brain mechanisms. This is often done, however, in a sloppy way, if we take the methodological standards of mechanistic explanations as an assessment criterion, as sketched in the list of five criteria in the previous section. Referring to behavioural economics applications, and to the five criteria listed above, we diagnose the following.

1. *Identification of phenomena.* The phenomena in question are not clearly identified, as they are mostly specified as deviations from a scenario of ‘rational’ behaviour, defined in different ways, such as deviations from an explicit rational choice model in games, or as unhealthy choices etc. This is a problem because this means that contextual factors have at least implicit causal powers. For example, loss aversion may not be a universal phenomenon, as it vanishes in certain market contexts: Even Kahneman (2011: 297ff) concedes that once individuals are framed as ‘traders’ in games such as the famous coffee mug exchange, the endowment effect would evaporate. Accordingly, there is an intensive discussion in experimental economics about how far experimental results in the lab transfer to the field (Levitt and List 2007; Kagel 2015). In addition, when comparing experiments with real-world situations, the serious problem emerges that apparently similar actions (such as drinking a glass of beer) can be embedded in entirely different action types and goal structures. For example, drinking beer in a pub differs from drinking beer at home, as in the former case goals such as shared emotions in the group may embed or even dominate the narrower goal of drinking beer (Hommel and Wiers 2017). In human action, the goal of a specific behavioural pattern is essential for delineating the phenomenon, so that the internal perspective of the agent needs to be made explicit. Whereas in experimental economics this mainly refers to cognitive acts such as interpreting experiments (for example, interpreting a public goods game as a ‘gamble’, Karlan 2005), in behavioural economics we need to take non-cognitive goal structures into consideration that are neurophysiologically embodied.
2. *Boundaries of mechanisms.* Delineating the boundaries of the mechanisms is mostly done with reference to the rational ideal type, again. This leads to overlooking the possible interactions between different mechanisms that supposedly ‘stand alone’, even including the possibility that the true identification of the mechanism would combine two supposedly different mechanisms into one integrated causal response pattern to perceived environmental situations. This can be even demonstrated by formal economic reasoning: For example, Findley and Caliendo (2014) show that a combination of hyperbolic time preferences with myopia can produce exponential time discounting on the behavioural level, and Steiner and Stewart (2016) show that apparently distorted weighting of probabilities as posited by prospect theory can be a ‘rational’ response to uncertainty in social interactions that cause the ‘winner’s curse’, enabling an agent to avoid it. Thus, ensembles of mechanisms might be the relevant units that generate behaviour. This raises the more fundamental question whether such ensembles might include external entities, along the lines of recent theories about the extended mind and distributed cognition (Ross 2012; building on

Clark and Chalmers 1998; Clark 2011). One important issue is how far external means of representation interact with internal neuronal processing: For example, different mechanisms of representing information trigger reversals in the loss aversion function, as in the case of frequentist versus probabilistic representations (see Glimcher 2011: 383ff). Even more radically, it is an open question how and where the border line between brain and body is drawn, which is particularly important in the context of social interactions. Social neuroscience suggests direct embodied coordination across different individuals (Oullier and Basso 2010; Gallese 2014).

3. *Parts of mechanisms.* Most dualist theories remain on the higher level of brain architecture, such as assigning the reflective and logical processes to the frontal cortex: On this coarse level of analysis, one cannot distinguish between ‘irrational’ and ‘rational’ forms of intuition: Rational intuition was famously established in the Iowa Gambling task and is discernible in many ways, such as in intuitive preferences for valid logical judgements, even accompanied by affectual phenomena (Handley and Trippas 2015). In many contexts, a disaggregate approach to brain connectivity shows that lower-level and higher-level structures interact, such as in mentalizing (e.g. Lombardo *et al.* 2009). This creates difficulties in unequivocally assigning certain constituents of dualities to brain structures on a higher level of aggregation: Decomposing the processes results in the identification of parts that belong to different structures.
4. *Architecture of mechanism.* Accordingly, dualist theories mostly fail to identify a precise picture of the brain architecture that underlies certain behavioural phenomena, which is crucial because often bottom-up and top-down processes interact, with the latter referring to the cognitive categorization of external contextual factors. The classical example is empathy, which is grounded in certain basal neurophysiological mechanisms, but at the same time depends on higher-level cognitive triggers (Singer and Lamm 2009; Decety 2015). The role of cognitive mechanisms in generating impulsive reactions is salient in the strong impact of attention on realizing certain action patterns (e.g. Puglisi *et al.* 2017). As Spunt (2015) convincingly argues, this comes to the fore when realizing that the different properties that supposedly make up the dualities often are orthogonal to each other, such that, for example, ‘automaticity’ does not unequivocally and exclusively go along with ‘unconscious’ or ‘uncontrollable’.
5. *Process boundaries.* Finally, there is no clear distinction between single acts of choice and series of similar choices through time, which is crucial as it has been shown, for instance, that bundling single choices has direct consequences for the validity of assumptions about time preferences, which would entail very different identification points for initial and terminal conditions of the respective cognitive and affective mechanisms. Bundling was proposed by Ainslie (1992, 2007) as overcoming behavioural dysfunctions resulting from hyperbolic time preferences, which have also been approached in a dual systems perspective, as mentioned previously. This relates to the previous point, as process types are defined in more complex architectures, such as involving the construction of the self and personal identity (Damasio 2010).

Thus, in summarizing this section, we can conclude that dualities in behavioural economics and psychology fail to meet the methodological standards of mechanistic explanations. They navigate between the Scylla of being by far too general to formulate meaningful psychological causal explanations that properly account for the system boundaries in which action is determined and the Charybdis of failing to properly identify causal mechanisms on the neurophysiological level. But I do not want to stop at this negative conclusion: The next task is to show how the mechanistic approach can help to identify alternative and promising heuristics of research.

4. The functionalist alternative to dual systems

4.1. *Dual function as alternative conceptualization of duality*

I think that there is a straightforward solution to the methodological troubles with dualistic theories which builds on the general concept of ‘function’. This perspective relates to Marr’s (1982) celebrated methodology of cognitive sciences, which distinguishes between three analytical levels, the computational, the algorithmic and implementation: On the computational level (1), we would ask for tasks that a certain mechanism is expected to realize, and define a computational solution for that; next, we look at alternative algorithmic methods (2) which then can be implemented and materialized in different embodiments (3) (such as a human brain or AI). Marr’s approach has been invoked by Glimcher’s (2003: 133ff) seminal methodological grounding of the emerging discipline of neuroeconomics (though not followed up in Glimcher 2011). However, Glimcher’s argument reveals a tension that I can use productively in developing my alternative functionalist approach to dualities.

Neuroeconomics clearly adopts the mechanistic methodology in focusing on empirical research on the neuronal mechanisms that underlie human (and other animal) decision making (Glimcher and Rusticchini 2004; Fehr and Rangel 2011). However, in defining neuroeconomics, Glimcher exploits Marr’s approach to introduce a major methodological shift: He argues that on the computational level, economic theory is the relevant reference. This stands in direct tension with the mechanistic methodology, as economics in terms of its mathematical models now serves as a source of generalizations in the neurosciences along the lines of the deductive-nomological model. This view has been challenged by Ross (2008, 2012) in distinguishing between a ‘molecular’ and a ‘molar’ approach to neuroeconomics, which raises the intricate question on which level economic theory applies. According to Ross, economic theory may apply for specific kinds of mechanisms within the individual but does not apply for behaviour as the outcome on the individual level, as far as this is approached in de-contextualized and generic terms, that is, without behaviour being scaffolded by institutions that constrain economic choices. In other words, in generic terms we could not conceive of individuals as being economic actors, but only parts of them. Obviously, this opens the possibility that Glimcher’s use of economics in neuroscience could be reconciled with psychological theories that deal with higher-level outcomes of complex interactions between lower-level mechanisms, even if contradicting the economic model on the level of the individual: The condition is giving up the claim that economic theory can directly explain individual behaviour, unless certain specific contextual conditions are fulfilled.

This possibility is indicated by Glimcher's own extension of Marr's framework. Glimcher introduces evolutionary theory to explain the functions on the computational level. This is not directly evident from Marr's own writings but appears reasonable when considering Marr's distinction between 'what' and 'why' questions: On the computational level, we need not only to identify what the process computes, but also why, and put this in a mathematical expression. This leads us immediately to the question how far Marr's methodology harmonizes with the mechanistic approach. The 'what' question is also central in mechanistic explanations, and mechanistic philosophers have indeed argued that Marr's computational and algorithmic level, taken together, are 'sketches of mechanisms' (Piccinini and Craver 2011). However, other commentators think that mechanistic explanations only involve bottom-up reductionist explanations whereas Marr's 'why' question implies a top-down explanation (Rusanen 2014).

In fact, there are two different ways to interpret Marr's 'why'. One has been suggested by Shagrir (2010) and is referred to as a mechanistic framework by Shagrir and Bechtel (2014). Responses to the 'why' question would relate to identifying properties in the environment that define physical constraints on the range of possible computational solutions, so that we can explain why the specific form of the computation (the 'what') is a proper solution to the task. Shagrir and Bechtel think that this is essential for delineating the phenomenon that can be approached as a mechanism.

The second form of response is Glimcher's evolutionary approach. I think that the two views are not mutually exclusive since they simply interpret the 'why' question in different, though complementary ways. This is salient once we introduce the notion of 'function'. The former view asks for explanations why a specific form of computation has the capacity to fulfil a certain function for the organism in relation to its environment (such as enabling workable vision). The second view explains why this function emerged in the real world: This is 'why?' as 'what for?'. Accordingly, the notion of function seems helpful in reconciling Marr's approach with the mechanistic methodology: In analysing mechanisms, it is always important to ask for the function that a specific mechanism fulfils. For example, when analysing the composition of a mechanism, we always need to ask what the function of lower-level mechanisms is in enabling the higher-level mechanism to operate properly. Eventually, we also ask what the function of the higher-level mechanism is. This relates to the phenomenon in question, and therefore relates mechanism with environment, as envisaged by Shagrir. In other words, Shagrir's interpretation of the 'why' question relates to the structure of selective constraints, and how these map into mechanisms, and Glimcher's relates to the process of how this mapping has been realized.

Functionalism is one of the most influential strands of thought in the philosophy of mind, with a rich variety of various approaches that were embroiled in a lively debate for decades (overview in Van Gulik 2009). I cannot deal with these debates here and remain satisfied with the simplest possible understanding that might prepare a common ground for my move from dual systems to dual functions. In this view, explaining a phenomenon would start out from asking the 'why' question first, i.e. identify its function, and then proceed to explain how this function is realized. This is different from some uses of the term in psychology and philosophy of mind which concentrate on the internal relationships between functions within

the brain/mind (this is also emphasized in Piccinini and Craver 2011). Although I include this meaning, I establish the close connection between function and evolutionary analysis as in concepts of aetiological function developed by theorists of teleosemantics and biosemantics (Millikan 2009; MacDonald and Papineau 2006; which differs from the previous internal view, with representatives such as Cummins 1975). This is relevant in the current context, because behavioural economics mainly concentrates on failures of representation and the resulting supposed malfunctioning. Teleosemantics uses a notion of 'proper functioning' in order to distinguish between these cases and regular realizations of functions on the level of components of the system that generates the behaviour (for example, what makes a frog respond properly to an environmental cue indicating a fly, and how to account for cognitive failures; see Neander 2006). In this sense, applying the notion of proper function on the aggregate level necessarily leads to a mechanistic reconstruction of causes of failure, which in turn requires the modeller to apply the notion also on lower levels. Notice that this does not imply the functionalist fallacy in simply assuming that lower-level mechanisms comply with the higher-level function: The possibility of failure is crucial for defining mechanistic research strategies. On first sight, this corresponds to the behavioural economics approach.

However, the functional view is radically de-constructionist about dualistic theories because it would guide research to identify specific mechanisms that realize certain functions, relative to certain contexts of actions. That means, we would no longer try to characterize entire domains of the brain-mind in terms of a certain system with a list of given properties (for example, relating the prefrontal cortex to reflective processes such as planning that have properties such as being 'slow') but aim at identifying mechanisms that are specific to certain functions. This is close to the approach of evolutionary psychology which confronts a modularized view of the brain with all kinds of theories that approach the brain as an integrated information processing system (Tooby and Cosmides 2005). This is where I depart from Glimcher's (2003) use of Marr in a principled way: Glimcher argues that the economic model of choice can be justified by referring to evolution as a process that resolves trade-offs in maximizing fitness. Evolutionary psychology rejects this reference to the notion of 'general purpose rationality' and suggests a modular alternative. This corresponds to a mechanistic methodology, since modules in evolutionary psychology are approached as multi-level mechanisms, such as when approaching emotions as higher-level coordinators of various lower-level mechanisms generating a certain behavioural pattern.

In sum, a simplest analytical move would be to re-interpret dual systems as dual functions: All properties that are used for characterizing the systems would be conceived as functional properties.

4.2. A workhorse: 'fast' versus 'slow' functions

Let us move to the constructive stage of my argument. I will approach the pivotal duality between 'fast' and 'slow' systems in terms of functions. 'Fastness' is a function that is necessary for certain aspects of survival, such as being able to respond to the threat of a predator. 'Slowness' is a function that may be necessary if larger quantities of information need to be processed, with no immediate threat in

place. If we consider speed as a function, we do no longer assign 'speed' as a property to specific types of brain processes, but we look at how the brain evolved specific mechanisms that can realize this function in relation to a task: This is Marr's computational level. Impulsive processes may realize this function, but so do certain kinds of reflective processes. There is no absolute notion of speed, but only speed relative to the fulfilment of specific functions.

I run through the list of requirements for mechanistic explanations again and ask how we can make use of the distinction between fast and slow functions to arrange the empirical material of neuroscientific research and draw the methodological conclusion.

1. *Identification of phenomenon.* Speed is always relative to the action that is necessary to fulfil a certain function properly. In the standard neuropsychological model of action, reflection is not necessarily slower than the action taken, even if triggered by an impulsive response in the first place. Action requires the formation of an action potential, and then realization. Reflective processes can set on and even reach conclusion while the action potential is still building, thus enabling continuous real-time reflective control of behaviour, which might appear to be as 'impulsive' to the external observer as the 'original' impulsive urge. This even applies for higher-level brain structures, such as the dualism between the cortical and limbic regions: It is perfectly possible that reflective feedback loops across brain areas operate faster than the action potential translates into action (for strong empirical support of this important point, see Cunningham and Zelazo 2007). This is also important when considering interactions between impulsive and automatic processes, such as when learning to suppress pain, even to the extent that feelings of pain are completely blocked.

Methodological conclusion: In delineating phenomena, speed is a functional requirement that allows for multiple realizability of functions. Absolute speed as a property of neuronal mechanisms should not matter for identifying adequate mechanistic explanations of observed behaviour.

2. *Boundaries of mechanisms.* One of the most challenging issues in dual systems views is the role of memory in guiding behaviour. In Kahneman's approach, the distinction between 'experienced utility' and 'decision utility' (Kahneman *et al.* 1997) does not harmonize well with his own dual systems concept, because decision utility is seen as a source of dysfunctions, yet it relates to the reflective and even conscious type of process. Experienced utility relates with the impulsive system in the sense that immediate sensory inputs are involved, such as in the famous experiments when test persons put their hands into cold water, later incorrectly remembering this experience according to the 'peak-end rule' which would underlie decision utility. One could argue, as Kahneman apparently does, that the retrieval of information from memory is a 'fast' process of its own kind, but then we would need to assume a large variety of different kinds of interacting systems (compare Berridge's 2009 addition of a 'learning system'). This is certainly true for memory (overview in Michaelian and Sutton 2017). Two aspects loom large in functional analysis. The first is that memory can be distributed over external media. This is salient in the emergent discipline of cultural neuroscience that

focuses on the interaction between external cultural media and human neuroplasticity, implying that similar situations, as seen from the perspective of the external observer, can go along with very different neurophysiological structures across cultures (for a survey, see Han *et al.* 2013). Hence, functional analysis would suggest that the boundaries of mechanisms need to be extended to include those external media (language, symbolic artefacts etc.). The second aspect is that in most experimental research informed by the dual systems approach, working memory is in focus, as cognitive load matters. But memory is a very complex multi-systems phenomenon, and, for example, retrieval from information stored in episodic memory can interplay with external media in creating fast responses to environmental settings which are even conscious, though not planned along the lines of the economic model of decision making. For example, Shohamy and Daw (2015) suggest a 'retrospective integration model' as opposed to the standard neuroeconomic 'prospective integration model'. This works via the continuous real-time 'replay' of decision scenarios triggered by similarities across past and present contexts, thus creating a preparedness for action which allows for fast responses.

Methodological conclusion: In realizing functions, there is no unequivocal mapping between properties of a behaviour identified by the external observer and the boundaries of mechanisms generating it.

3. *Parts of mechanisms.* Many dualistic theories posit a clear distinction between the types of mechanism that define the two systems, which is not vindicated empirically. For example, recent work on relative speeds of belief-based and logical structure-based behaviour in solving cognitive tasks has shown that logical processing often operates on an intuitive level, even unconscious, and finds expression in affective valuations (Handley and Trippas 2015). The duality is most salient in Strack and Deutsch's (2004) canonical model that distinguishes between 'associative' information processing in the impulsive system and a propositional mode of information processing in the reflective system. Yet, they also recognize a bidirectional linkage between conceptual content and behavioural schemata, without pursuing possible implications for the overall validity of the duality. If we consider the semantics that is operative in propositional structures, the boundaries between different types of information processing become blurred or even obsolete. Meanings are practices and actions which go along with valuations that are affective in turn (Pinker 2007). Hence, 'reflectivity' may be related with associative processes as well, as in those theories of thought and meaning that build on metaphor as the fundamental notion (Fauconnier and Turner 2002; Lakoff 2008). In this perspective, a supposedly linguistically mediated reflective system would build on the same elements of information processing as the impulsive system, so that the conceptual difference between the systems would collapse. One example for this is Lades' (2014) attempt at broadening the scope of the concept of 'impulsivity' to include reflective reference to self-images and identity-related needs, which are always symbolically mediated, and can signal states of deprivation of similar strength as states of physiological deprivation.

Methodological conclusion: Functional analysis should avoid any essentialist pre-assignment of types of parts to types of mechanisms.

4. *Architecture of mechanism.* Concepts such as selfhood point towards the question of architectures of function fulfilment. One empirical problem with dual systems theories is that they either assume a time lag between the two responses of the system (impulsiveness first, control via reflection later) or do not detail the mechanisms that govern the parallelism between the two. The latter implies we would need to introduce a higher-level mechanism that chooses among the two reaction patterns. This idea has been emerging in dual systems theories recently (Stanovich 2011); one systematic model is Foxall's (2016) who suggests a tripartite model where the notion of reflection refers to a meta-cognitive level which decides which of the two types of processes, impulsive or algorithmic, will operate. Evidently, this implies that the higher-level function operates with at least the same speed as the impulsive system. In the psychological literature, two phenomena are especially difficult to align with the dualism between fast and slow: attention and motivation (Spunt 2015; Wiers and Gladwin 2017). Attention is a complex mechanism of its own standing that plays a central role in triggering other mechanisms, including even reflexes, so that one needs to speak of 'prepared reflexes' (Hommel and Wiers 2017). Motivation is similarly, if not more complex than attention, and is crucial to understand differential performances across individuals in manifesting dysfunctional behaviour. These observations lead us to consider the more general question of how to account for parallelism systematically. A functional approach allows for the possibility that behaviour is generated by competing mechanisms of which one is selected according to a 'winner takes all' principle. This can explain behavioural variety within and across individuals.

Methodological conclusion: Functional analysis implies that speed and complexity are not unequivocally related; complex architectures can operate with high speed, especially if parallelism is considered.

5. *Process boundaries.* Equating intuition with fast responses exclusively cannot be vindicated empirically. For example, in interpersonal relations, many intuitions build up slowly, even undergirded by reflection. We might feel spontaneous sympathy with a person, yet this emotion also unfolds through time, though still being driven by intuitions. In comparison, reflective or propositional classifications of persons might operate with much higher speed. For example, if I know that a person is a policewoman, I might automatically reach certain conclusions about predicting her behaviour, whereas building expectations on personal trust as an intuition would need much longer time. This may be generalized in the sense that often a misplaced assumption is made that 'emotions' are states, whereas reflection is a process, which immediately seems to suggest that the former are 'fast', once they are triggered. But in fact, emotions are also processes, and depending on the nature of the emotions these can work at different speeds (Goldie 2014).

Methodological conclusion: Identification of functions implies complex time structures of function fulfilment that map into a large variety of possible mechanisms enabling proper fulfilment.

5. Conclusion

The move to a dual functions perspective is highly significant for rethinking the relationship between neuroeconomics and behavioural economics. A large part of neuroeconomic research is focused on identifying single mechanisms of choice, based on certain reward circuits in the brain (Glimcher 2009; Fehr and Rangel 2011). This is extended to adopting a unitary approach to decision and behaviour that aims at reducing the economic model to neuroscience (Glimcher 2011). However, leading researchers in the field have also recognized that this approach fails to consider higher levels of brain organization, such as complex cognitive performances or emotions (Camerer 2013). On the other hand, we have broad high-level generalizations such as the dualist theories, which explicitly are treated as alternatives to the unitary model of economics, thus resulting in a tension with the leading neuroeconomic paradigm. What is missing, is the intermediate level. This comes into focus, when we introduce the mechanistic perspective. That corresponds to the ‘modular view’ championed by Don Ross (2008, 2012). In this approach, we concentrate on the interaction between levels of organization in the brain and even include external entities in complete causal explanations.

Coming back to the normative aspects that I have touched in the introduction; the dual function approach alleviates the tension between the normative foundations of economics and the methods of behavioural policy. This is vindicated by empirical research: For example, Loewenstein *et al.* (2015) have shown that informing experimental subjects about being nudged does not make the intervention ineffective, even in the sense of the response pattern. In my dual functions approach, this is straightforward to explain, such as when considering the parallelism of various response patterns to a certain situational context, and the immediate activation of higher-level cognitive structures.

Author ORCID.  Carsten Herrmann-Pillath 0000-0001-6078-0078

Acknowledgements. I am grateful to insightful discussions with the INSOSCI team members at the project meeting at Käsna, Finland, in July 2017. Special thanks go to the invited guest speaker, Reinout Wiers, for showing the way through the forest of dual system theories, and to Jens Harbecke for clarifying the mechanistic approach in philosophy of science. I am indebted to the *Economics and Philosophy* reviewers whose comments have helped to improve the manuscript substantially.

Financial Support. This work was funded by ERA-NET NEURON in the context of the INSOSCI project ‘The integration of cross-disciplinary research in neuroscience and social science – a methodological case study on economic policies and the neuroscience of agency’, Grant number 01GP1625.

References

- Ainslie G.** 1992. *Picoeconomics: The Strategic Interaction of Successive Motivational States within the Person*. Cambridge: Cambridge University Press.
- Ainslie G.** 2007. Thought experiments that explore where controlled experiments can’t: The example of will. In D. Ross, D. Spurrett, H. Kincaid and G. L. Stephens (eds), *Distributed Cognition and the Will. Individual Volition and Social Context*. Cambridge, MA: MIT Press, pp. 169–198.
- Alós-Ferrer C. and F. Strack** 2014. From dual processes to multiple selves: implications for economic behaviour. *Journal of Economic Psychology* **41**, 1–11.
- Bechara A. and A. R. Damasio** 2005. The somatic marker hypothesis: a neural theory of economic decision. *Games and Economic Behavior* **52**, 336–372.

- Bechtel W. and A. Abrahamsen** 2005. Explanation: a mechanist alternative. *Studies in History and Philosophy of Biological and Biomedical Sciences* **36**, 421–441.
- Bénabou R. and M. Pycia** 2002. Dynamic inconsistency and self-control: a planner–doer interpretation. *Economics Letters* **77**, 419–424.
- Benhabib J. and A. Bisin** 2005. Modeling internal commitment mechanisms and self-control: a neuroeconomics approach to consumption-saving decisions. *Games and Economic Behavior* **52**, 460–492.
- Bernheim B. D.** 2009. The psychology and neurobiology of judgment and decision making: What’s in it for economists? In P. W. Glimcher, C. F. Camerer, E. Fehr and R. A. Poldrack (eds), *Neuroeconomics: Decision Making and the Brain*. Amsterdam: Elsevier, pp. 115–126.
- Berridge K. C.** 2009. ‘Liking’ and ‘wanting’ food rewards: brain substrates and roles in eating disorders. *Physiology and Behavior* **97**, 537–550.
- Bovens L.** 2008. The ethics of nudge. In T. Grüne-Yanoff and S. O. Hansson (eds), *Preference Change: Approaches from Philosophy, Economics and Psychology*. Berlin: Springer, pp. 207–220.
- Brocas I. and J. D. Carrillo** 2008. The brain as a hierarchical organization. *American Economic Review* **98**, 1312–1347.
- Brocas I. and J. D. Carrillo** 2014. Dual-process theories of decision-making: a selective survey. *Journal of Economic Psychology* **41**, 45–54.
- Bunge M.** 1977. *Treatise on Basic Philosophy, Volume 3. Ontology I: The Furniture of the World*. Dordrecht: Reidel.
- Bunge M.** 1979. *Treatise on Basic Philosophy, Volume 4. Ontology II: A World of Systems*. Dordrecht: Reidel.
- Camerer C.** 2006. Wanting, liking, and learning: neuroscience and paternalism. *University of Chicago Law Review* **73**, 87–110.
- Camerer C.** 2013. A review essay about Foundations of Neuroeconomic Analysis by Paul Glimcher. *Journal of Economic Literature* **LI**, 1155–1184.
- Camerer C., G. Loewenstein and D. Prelec** 2005. Neuroeconomics: How neuroscience can inform economics. *Journal of Economic Literature* **XLIII**, 9–64.
- Clark A.** 2011. *Supersizing the Mind: Embodiment, Action, and Cognitive Extension*. Oxford: Oxford University Press.
- Clark A. and D. J. Chalmers** 1998. The extended mind. *Analysis* **58**, 10–23.
- Coleman J.** 1990. *Foundations of Social Theory*. Cambridge, MA: Belknap.
- Craver C.** 2007. *Explaining the Brain. Mechanisms and the Mosaic Unity of the Neurosciences*. New York, NY: Oxford University Press.
- Craver C. and J. Tabery** 2015. Mechanisms in science. In E. N. Zalta (ed.), *Stanford Encyclopedia of Philosophy*. Available at <http://plato.stanford.edu/archives/win2015/entries/science-mechanisms/>.
- Cummins R.** 1975. Functional analysis. *Journal of Philosophy* **72**, 741–764.
- Cunningham W. A. and P. D. Zelazo** 2007. Attitudes and evaluations: a social cognitive neuroscience perspective. *Trends in Cognitive Sciences* **11**, 97–104.
- Damasio A. R.** 2010. *Self Comes to Mind. Constructing the Conscious Brain*. New York, NY: Pantheon.
- Decety J.** 2015. The neural pathways, development and functions of empathy. *Current Opinion in Behavioral Sciences* **3**, 1–6.
- Dowe P.** 2008. Causal Processes. In E. N. Zalta (ed.), *Stanford Encyclopedia of Philosophy*. Available at <http://plato.stanford.edu/archives/fall2008/entries/causation-process/>.
- Evans J. S. B. T.** 2010. *Thinking Twice: Two Minds in One Brain*. New York, NY: Oxford University Press.
- Evans J. S. B. T. and K. E. Stanovich** 2013. Dual-process theories of higher cognition: advancing the debate. *Perspectives on Psychological Science* **8**, 223–241.
- Fauconnier G. and M. Turner** 2002. *The Way We Think: Conceptual Blending and the Mind’s Hidden Complexities*. New York, NY: Basic Books.
- Fehr E. and A. Rangel** 2011. Neuroeconomic foundations of economic choice– recent advances. *Journal of Economic Perspectives* **25**, 3–30.
- Findley T. S. and F. N. Caliendo** 2014. Interacting mechanisms of time inconsistency. *Journal of Economic Psychology* **41**, 68–76.
- Foxall G. R.** 2016. Metacognitive control of categorial neurobehavioral decision systems. *Frontiers in Psychology* **7**. doi: [10.3389/fpsyg.2016.00170](https://doi.org/10.3389/fpsyg.2016.00170).

- Fudenberg D. and D. K. Levine** 2006. A dual-self model of impulse control. *American Economic Review* **96**, 1449–1477.
- Gallese V.** 2014. Bodily selves in relation: embodied simulation as second-person perspective on intersubjectivity. *Philosophical Transactions of the Royal Society B: Biological Sciences* **369**, 20130177
- Glimcher P. W.** 2003. *Decisions, Uncertainty, and the Brain: The Science of Neuroeconomics*. Cambridge, MA: MIT Press.
- Glimcher P. W.** 2009. Choice: towards a standard back-pocket model. In P. W. Glimcher, C. F. Camerer, E. Fehr and R. A. Poldrack (eds), *Neuroeconomics: Decision Making and the Brain*. Amsterdam: Elsevier, pp. 503–522.
- Glimcher P. W.** 2011. *Foundations of Neuroeconomic Analysis*. Oxford: Oxford University Press.
- Glimcher P. W. and A. Rusticchini** 2004. Neuroeconomics: the consilience of brain and decision. *Science* **306**, 447–452.
- Goldie P.** 2014. *The Mess Inside: Narrative, Emotion, & the Mind*. Oxford: Oxford University Press.
- Gul F. and W. Pesendorfer** 2008. The case for mindless economics. In A. Caplin and A. Schotter (eds), *The Foundations of Positive and Normative Economics*. Oxford: Oxford University Press, pp. 3–41.
- Han S., G. Northoff, K. Vogeley, B. E. Wexler, S. Kitayama and M. E. W. Varnum** 2013. A Cultural neuroscience approach to the biosocial nature of the human brain. *Annual Review of Psychology* **64**, 335–359.
- Handley S. J. and D. Trippas** 2015. Dual processes and the interplay between knowledge and structure: a new parallel processing model. *Psychology of Learning and Motivation*, **62**, 33–58.
- Harbecke J.** 2014. The role of supervenience and constitution in neuroscientific research. *Synthese* **191**, 725–743.
- Harrison G. W. and D. Ross** 2017. The empirical adequacy of cumulative prospect theory and its implications for normative assessment. *Journal of Economic Methodology* **24**, 150–165.
- Hommel B. and R. W. Wiers** 2017. Towards a unitary approach to human action control. *Trends in Cognitive Sciences* **21**, 940–949.
- Jamison J. and J. Wegener** 2010. Multiple selves in intertemporal choice. *Journal of Economic Psychology* **31**, 832–839.
- Kable J. W. and P. W. Glimcher** 2007. *Matching Mental and Neural Representation of Value during Intertemporal Choice*. New York, NY: Center for Neural Science. Available at <http://www.cns.nyu.edu/events/neuroecon/KableGlimcher.pdf>.
- Kagel J. H.** 2015. Laboratory experiments: the lab in relationship to field experiments, field data and economic theory. In G. R. Schotter and A. Schotter (eds), *Handbook of Experimental Economic Methodology*. Oxford: Oxford University Press, pp. 339–359.
- Kahneman D.** 2011. *Thinking, Fast and Slow*. London: Penguin Books.
- Kahneman D., P. P. Wakker and R. Sarin** 1997. Back to Bentham? Explorations of experienced utility. *Quarterly Journal of Economics* **112**, 375–405.
- Karlan D. S.** 2005. Using experimental economics to measure social capital and predict financial decisions. *American Economic Review* **95**, 1688–1699.
- Lades L. K.** 2014. Impulsive consumption and reflexive thought: nudging ethical consumer behavior. *Journal of Economic Psychology* **41**, 114–128.
- Laibson D.** 1997. Golden eggs and hyperbolic discounting. *Quarterly Journal of Economics* **112**, 443–477.
- Lakoff G.** 2008. The neural theory of metaphor. In R. W. Gibbs (ed.), *The Cambridge Handbook of Metaphor and Thought*. New York, NY: Cambridge University Press, pp. 17–38.
- Levitt S. D. and J. A. List** 2007. What do laboratory experiments measuring social preferences reveal about the real world? *Journal of Economic Perspectives* **24**, 31–64.
- Loewenstein G., C. Bryce, D. Hagmann and S. Rajpal** 2015. Warning: you are about to be nudged. *Behavioral Science and Policy* **1**, 35–42.
- Lombardo M. V., B. Chakrabarti, E. T. Bullmore, S. J. Wheelwright, S. A. Sadek, J. Suckling, MRC AIMS Consortium and S. Baron-Cohen** 2009. Shared neural circuits for mentalizing about the self and others. *Journal of Cognitive Neuroscience* **22**, 1623–1635.
- MacDonald G. and D. Papineau** (ed.). 2006. *Teleosemantics: New Philosophical Essays*. Oxford: Oxford University Press.
- Machamer P., L. Darden and C. F. Craver** 2000. Thinking about mechanisms. *Philosophy of Science* **67**, 1–25.

- Marr D.** 1982. *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. New York, NY: Henry Holt.
- McClure S. M., K. M. Ericson, D. I. Laibson, G. Loewenstein and J. D. Cohen** 2007. Time discounting for primary rewards. *Journal of Neuroscience* **27**, 5796–5804.
- McFadden D.** 2006. Free markets and fettered consumers. *American Economic Review* **96**, 5–29.
- McLaughlin B. and K. Bennett** 2018. Supervenience. In E. N. Zalta (ed.), *Stanford Encyclopedia of Philosophy*. Available at <https://plato.stanford.edu/archives/spr2018/entries/supervenience/>.
- Michaelian K. and J. Sutton** 2017. Memory. In E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*. Available at <https://plato.stanford.edu/archives/sum2017/entries/memory/>.
- Millikan R. G.** 2009. Biosemantics. In B. P. McLaughlin, A. Beckermann and S. Walter (eds), *The Oxford Handbook of Philosophy of Mind*. Oxford: Clarendon Press, pp. 394–406.
- Minsky M.** 2006. *The Emotion Machine: Commonsense Thinking, Artificial Intelligence, and the Future of the Human Mind*. New York, NY: Simon & Schuster.
- Neander K.** 2006. Content for Cognitive Science. In G. Macdonald and D. Papineau (eds), *Teleosemantics: New Philosophical Essays*. Oxford: Oxford University Press, pp. 167–194.
- Oullier O. and F. Basso** 2010. Embodied economics: how bodily information shapes the social coordination dynamics of decision-making. *Philosophical Transactions of the Royal Society B* **365**, 291–301.
- Piccinini G. and C. Craver** 2011. Integrating psychology and neuroscience: functional analyses as mechanism sketches. *Synthese* **183**, 283–311.
- Pinker S.** 2007. *The Stuff of Thought: Language as a Window into Human Nature*. New York, NY: Penguin.
- Puglisi G., A. Leonetti, A. Landau, L. Fornia, G. Cerri and P. Borroni** 2017. The role of attention in human motor resonance. *PLoS ONE* **12**, e0177457. doi: [10.1371/journal.pone.0177457](https://doi.org/10.1371/journal.pone.0177457).
- Redish A. D., S. Jensen and A. Johnson** 2008. A unified framework for addiction: vulnerabilities in the decision process. *Behavioral and Brain Sciences* **31**, 415–487.
- Reimann M. and A. Bechara** 2010. The somatic marker hypothesis as a neurological theory of decision-making: review, conceptual comparisons, and future neuroeconomics research. *Journal of Economic Psychology* **31**, 767–776.
- Ross D.** 2005. *Economic Theory and Cognitive Science: Microexplanations*. Cambridge, MA: MIT Press.
- Ross D.** 2008. Two styles of neuroeconomics. *Economics and Philosophy* **24**, 473–483.
- Ross D.** 2012. Neuroeconomics and economic methodology. In J. B. Davis and D. W. Hands (eds), *The Elgar Companion to Recent Economic Methodology*. Cheltenham: Edward Elgar, pp. 61–93.
- Rusanen A.** 2014. On Explaining Cognitive Phenomena: The Limits of Mechanistic Explanation. PhD dissertation, Helsinki. Available at <http://ethesis.helsinki.fi>.
- Shagrir O.** 2010. Marr on computational-level theories. *Philosophy of Science* **77**, 477–500.
- Shagrir O. and W. Bechtel** 2014. Marr's computational level and delineating phenomena. Available at http://philsci-archive.pitt.edu/11224/1/shagrir_and_bechteli.Marr's_Computational_Level_and_Delineating_Phenomena.pdf.
- Shohamy D. and N. Daw** 2015. Integrating memories to guide decisions. *Current Opinion in Behavioral Sciences* **5**, 85–90.
- Singer T. and C. Lamm** 2009. The social neuroscience of empathy. *Annals of the New York Academy of Sciences* **1156**: 81–96.
- Spunt R. P.** 2015. Dual-process theories in social cognitive neuroscience. In *Brain Mapping*. Amsterdam: Elsevier, pp. 211–215. doi: [10.1016/B978-0-12-397025-1.00181-0](https://doi.org/10.1016/B978-0-12-397025-1.00181-0)
- Stanovich K. E.** 2011. *Rationality and the Reflective Mind*. New York, NY: Oxford University Press.
- Steiner, J. and C. Stewart** 2016. Perceiving prospects properly. *American Economic Review* **106**, 1601–1631.
- Strack F. and R. Deutsch** 2004. Reflective and impulsive determinants of social behaviour. *Personality and Social Psychology Review* **8**, 220–247.
- Thaler R. H.** 2016. Behavioral economics: past, present, and future. *American Economic Review* **106**, 1577–1600.
- Thaler R. H. and H. M. Shefrin** 1981. An economic theory of self-control. *Journal of Political Economy* **89**, 392–406.
- Thaler R. H. and C. R. Sunstein** 2009. *Nudge. Improving Decisions about Health, Wealth, and Happiness*. London: Penguin.
- Tooby J. and L. Cosmides** 2005. Conceptual foundations of evolutionary psychology. In D. M. Buss (ed.), *The Handbook of Evolutionary Psychology*. Hoboken, NJ: Wiley, pp. 5–67.

- Trepel C., C. R. Fox and R. A. Poldrack** 2005. Prospect theory on the brain? Toward a cognitive neuroscience of decision under risk. *Cognitive Brain Research* **23**, 34–50.
- Van Gulik R.** 2009. Functionalism. In B. P. McLaughlin, A. Beckermann and S. Walter (eds), *The Oxford Handbook of Philosophy of Mind*. Oxford: Clarendon Press, pp. 128–151.
- Waskan J.** 2011. Mechanistic explanation at the limit. *Synthese* **183**, 389–408.
- Wiers R. W., B. D. Bartholow, E. van den Wildenberg, C. Thush, R. C. M. E. Engels, K. J. Sher, J. Grenard, S. L. Ames and A. W. Stacy** 2007. Automatic and controlled processes and the development of addictive behaviors in adolescents: a review and a model. *Pharmacology, Biochemistry and Behavior* **86**, 263–283.
- Wiers R. W. and T. W. Gladwin** 2017. Reflective and impulsive processes in addiction and the role of motivation. In R. Deutsch, B. Gawronski and W. Hofmann (eds), *Reflective and Impulsive Determinants of Human Behaviour*. New York, NY: Routledge, pp. 173–188.
- Woodward J.** 2003. *Making Things Happen: A Theory of Causal Explanation*. Oxford: Oxford University Press.
- Woodward J.** 2011. Mechanisms revisited. *Synthese* **183**, 409–427.

Carsten Herrmann-Pillath is Professor and Permanent Fellow at the Max Weber Centre for Advanced Cultural and Social Studies at Erfurt University, Germany. His work spans across various disciplines, with philosophy of economics at the core. His recent books include: *Foundations of Economic Evolution: A Treatise on the Natural Philosophy of Economics* (Edward Elgar, 2013), and (with Ivan Boldyrev) *Hegel, Institutions and Economics: Performing the Social* (Routledge, 2014). His current research focuses on cross-disciplinary integration between the sciences and economics, both ‘macro’ (ecological economics and Earth system sciences) and ‘micro’ (neurosciences and behavioural economics). His personal website is: cahepil.net.