PHILOSOPHICAL PSYCHOLOGY
https://doi.org/10.1080/09515089.2024.2444503

Routledge
Taylor & Francis Group

Greene's dual-process moral psychology and the modularity of mind

Moritz Bammel

Institute of Sustainability Psychology, Leuphana University Lueneburg, Lueneburg, Germany

ABSTRACT

Joshua Greene has famously argued for two distinct processes of how humans make moral judgments. Despite a lively controversy around potential normative implications of this view, less attention has been paid to those philosophical assumptions that are fundamental to Greene's dual-process theory itself. In this paper, I argue that Greene's dual-process theory hinges on a modular account of cognition and the brain, and I critically discuss the plausibility of Greene's view in light of increasing popularity of dynamical systems accounts in cognitive science. If we reject modularity and adopt a dynamical systems perspective instead, we can still hope to find relative differences in the functional specialization of dynamic brain networks within one interconnected system, but Greene's original theory in terms of two asymmetrically independent processes will no longer be tenable. This imposes constraints on the kind of explanations that we can expect from an empirically informed ethics in that only non-exclusive dual-process theories would be compatible with a dynamical systems account. Ultimately, however, the controversy around the modularity of mind should not be misconceived as a purely empirical question, but rather as a matter of conflicting epistemic standards as to what qualifies as a good explanation in cognitive science.

ARTICLE HISTORY
Received 19 May 2024
Accepted 12 December 2024

KEYWORDS
Dual-process; moral judgment; empirical ethics; modularity; dynamical systems theory; brain networks

1. Introduction

Empirically informed ethics aims at providing a new impetus to long-standing and sometimes deadlocked debates in moral philosophy by drawing on empirical findings from moral psychology and neuroscience (Bublitz & Paulo, 2020). A prime example of such an empirically informed approach to normative ethics is Greene's *dual-process theory of moral judgment* (Greene, 2013, 2014; Greene et al., 2001, 2004). According to Greene, characteristically deontological and consequentialist moral intuitions, respectively, are products of two different

CONTACT Moritz Bammel moritz.bammel@leuphana.de Institute of Sustainability Psychology, Leuphana University Lueneburg, Universitaetsallee 1, Lueneburg 21335, Germany

cognitive modules, one referring to the "automatic mode" associated with fast and emotional responses and the other referring to the "manual mode" associated with slow and deliberate thinking. Based on empirical data in support of the dual-process theory of moral judgment, and normative assumptions about when to use the "manual mode" and when to use the "automatic mode", Greene (2014) developed an argument in favor of relying on consequentialist instead of deontological moral intuitions when confronted with unfamiliar ethical problems.

Despite a lively controversy around Greene's dual-process theory over the last two decades, which mainly revolved around the question whether the existence of two separable neuropsychological processes of making moral judgments would have any genuine normative implications (Berker, 2009; Dale, 2020; Dale & Gawronski, 2022; Dean, 2009; Heir, 2023; Kamm, 2009; Knigs, 2018; Kumar & Campbell, 2012; Lott, 2016; Meyers, 2013; Mihailov, 2015; Paulo, 2019; Sauer, 2012; Singer, 2005), less attention has been paid to those philosophical assumptions that are fundamental to Greene's dual-process theory itself. In contrast to a critical reception of Greene's normative claims concerning the reliability of deontological intuitions in the context of unfamiliar moral dilemmas, many scholars who participated in the debate around Greene's dual-process moral psychology have rather uncritically accepted the empirical data presented by Greene as evidence for the existence of two separate cognitive modules. However, as I will argue in this paper, Greene's dual-process theory of moral judgment hinges on an implicit endorsement of the *modularity of mind* (Fodor, 1983), since Greene is making a claim not only about two different types of moral judgments on a behavioral level but also about two distinct cognitive modules in the brain that are in turn supposed to give rise to these different behaviors.[1]

On the classical modular view, the human brain consists of independent, anatomically and functionally clearly dissociable modules, each of which is presumed to enable different aspects of human cognition (Barrett & Kurzban, 2006; Fodor, 1983; Sternberg, 2011). Dynamical systems approaches in cognitive science and neuroscience, in contrast, reject a strictly modular view and advocate in favor of conceiving the brain as a nonlinear dynamical system instead, which is subject to constant plasticity and in which functionally specialized brain networks emerge and change over time (Chemero, 2009, 2023; Favela, 2020, 2021, 2024; Izhikevich, 2007; Kelso, 1995; Nguyen et al., 2024; Noble et al., 2024; Pessoa, 2014, 2022; Smith & Thelen, 1994, Thompson & Varela, 2001; van Gelder, 1998; Van Orden & Kloos, 2003; Van Orden et al., 2001, 2003). Thus, a dynamicist account of cognition seems to be incompatible with the basic assumption of Greene's dual-process theory – irrespective of the controversies around potential normative implications – according to which there are two distinct

brain modules that arguably give rise to either deontological or consequentialist moral intuitions.

By raising awareness to an important debate in philosophy of cognitive science that has been largely neglected in the reception of Greene's dual-process theory, I aim to stimulate a new direction of discussion in empirical ethics that takes into account those theoretical developments in cognitive science that have gained momentum since the publication of the initial study by Greene et al. (2001), first and foremost applications of dynamical systems theory in cognitive science and neuroscience. The argument developed in this paper is conditional in nature in that I focus on the implications for Greene's dual-process theory and future research in empirical ethics if we were to reject the modularity of mind and adopt a dynamical systems perspective on cognition and the brain instead, but I do not aim to resolve the debate between proponents of modularity and dynamical systems thinking at this point.

Starting in section 2, I will provide a brief summary of Greene's dual-process theory and the subsequent discussion around potential normative implications. In section 3, I will then turn to the modularity of mind and a critical examination of this implicit premise in Greene's argument. In light of growing interest in adopting a dynamical systems perspective in neuroscience and cognitive science, I will portray how a dynamicist conception of cognition and the brain differs from the modularity of mind in section 4. Finally, in section 5, I will discuss the implications of a potential rejection of a modular view of cognition for dual-process theories in empirical ethics. Although it follows that a static demarcation in terms of two separate cognitive modules will no longer be tenable, I do not intend to reject attempts to distinguish different classes of moral judgments altogether. Instead, in line with adopting a dynamical systems conception, we need to reconsider the types of explanations that we can expect from an empirically informed ethics, and we should rather look for relative, context-sensitive differences in brain and behavior dynamics that may inform similarly more nuanced differences in moral judgments.

2. Greene's dual-process moral psychology

In a series of neuroimaging studies, Greene et al. (2001, 2004) investigated whether there are differences in brain activity and reaction times associated with different moral judgments in two versions of the well-known trolley cases. In the first case, the trolley dilemma, participants were asked to decide whether they consider it morally permissible to press a button to change

a switch so that a trolley would not kill five people on the current track but only one person on the other track instead. Most participants gave an affirmative answer. In the second case, the footbridge dilemma, participants were asked to decide whether they consider it morally permissible to push a person off a bridge onto the track and sacrifice that one person in order to save the lives of five people who would have been killed by the train otherwise. In contrast to the trolley dilemma, most participants gave a negative answer in the footbridge dilemma. How come?

Greene and colleagues hypothesized that differences in moral judgment in the trolley dilemma compared to the footbridge dilemma can be explained by selective involvement of primarily fast and emotional neuropsychological processes leading to a negative judgment in the footbridge dilemma, in contrast to slower and more deliberate neuropsychological processes leading to a positive judgment in the trolley dilemma. To test their hypothesis, Greene et al. (2001, 2004) presented a series of personal and impersonal moral dilemmas to participants that were constructed analogously to either the footbridge or the trolley dilemma, respectively.[2] In line with their hypothesis, they found significantly increased hemodynamic activity in brain areas associated with emotions in personal compared to impersonal dilemmas. In contrast, those brain areas associated with working memory showed significantly stronger hemodynamic activity in impersonal compared to personal moral dilemmas. Moreover, only in personal moral dilemmas, reaction times were significantly faster for typical negative moral judgments than for untypical positive moral judgments. In the case of impersonal moral dilemmas, however, there was no significant difference in reaction times between typical positive and untypical negative moral judgments.

Based on these empirical findings, Greene (2014) developed the *dual-process theory of moral judgment*, which is inspired by the long-standing tradition of dual-process theories in psychology that do not only apply to moral judgments (for review of the dual-process literature, see De Neys, 2021; Evans & Stanovich, 2013). According to Greene's theory, there is a fast, automatic and emotion-driven process that enables efficient judgment and decision making on the one hand, and a slower, more flexible process on the other hand that enables deliberative thinking. Greene illustrates these two processes with the help of a camera analogy: the fast, emotional mode corresponds to the default settings of a camera, the "automatic mode", and thus makes it possible to quickly take decent quality pictures in many situations. On the other hand, there is the "manual mode", which takes more time but leads to better pictures in certain situations. According to Greene, these two processes are asymmetrically independent of each other in that the automatic mode works without the manual mode, whereas the manual mode does not work

without the automatic mode. Since most participants made characteristically deontological moral judgments in personal moral dilemmas which were correlated with increased activity in brain regions associated with emotions, Greene concludes that characteristically deontological moral judgments are primarily a product of the automatic mode. Characteristically consequentialist moral judgments are in turn primarily a product of the manual mode, according to Greene, since the majority of subjects in impersonal moral dilemmas made characteristically consequentialist moral judgments which correlated with increased activity in brain regions associated with working memory.[3] This is what Greene calls the *central tension principle* (Greene, 2014, pp. 697–708). Importantly, Greene does not only assume a fast (automatic) and a slow (manual) process at the behavioral level – a less controversial hypothesis in psychology – but he also assumes that these two processes are implemented in the form of different cognitive modules at the neurophysiological level, and it is precisely the latter claim that is at stake in the present paper.

With respect to potential normative implications for moral philosophy, Greene (2014) describes two lines of reasoning, the *direct* and the *indirect route*. The direct route is a classical debunking argument in that it illuminates the role of a morally irrelevant factor, in this case personal force, to which either characteristically deontological or characteristically consequentialist moral judgments seem to be susceptible. However, the direct route argument alone does not suffice to justify which class of moral judgments – characteristically deontological or consequentialist – seems to be biased (Greene, 2014, pp. 711–713). The indirect route argument therefore involves the additional normative premise that we should rely on the manual and not the automatic mode when we find ourselves confronted with unfamiliar moral problems. Greene does not consider all automatic mode, emotion-driven moral judgments as problematic, but only those made in situations for which we lack relevant previous experience that could have re-calibrated the automatic mode instead of relying on misleading evolutionary preconfigurations. For Greene, almost all moral questions for which there is no intuitive consensus on how to act count as unfamiliar, including the much-discussed variations of the trolley cases. Therefore, according to Greene, we should reject characteristically deontological intuitions, since these are a product of the automatic mode, and rely on characteristically consequentialist intuitions instead, for the latter are a product of the manual mode (Greene, 2014, pp. 713–725).

Greene's direct and indirect route arguments have triggered a lively discussion in the philosophical literature with many contributions either sympathetic (Kumar & Campbell, 2012; Paulo, 2019; Singer, 2005) or critical (Berker, 2009; Dale, 2020; Dale & Gawronski, 2022; Dean, 2009; Heir, 2023;

Kamm, 2009; Knigs, 2018; Lott, 2016; Meyers, 2013; Mihailov, 2015; Sauer, 2012) of the significance of Greene's empirical findings for normative ethics. Surprisingly, however, there has been a largely uncritical acceptance of Greene's dual-process theory itself among both supporters and critics of potential normative implications derived from such a dual-process view, i.e., the idea that there are two separate cognitive modules implemented in distinct brain areas that give rise to either characteristically deontological or consequentialist moral judgments (for exceptions, see Craigie, 2011; Crockett, 2013; Cushman, 2013; De Neys, 2023; Kahane, 2012; Kahane & Shackel, 2010; Kahane et al., 2012). In the remainder of this paper, I will show that the acceptance of Greene's dual-process theory of moral judgment, regardless of its normative (in)significance, presupposes a tacit commitment to a modular view of cognition and the brain, which is increasingly challenged by dynamical systems accounts of the brain and cognition.

3. The modularity of mind

3.1. Fodor's conception of modularity

The modularity of mind refers to a view of cognition and the brain that was, and still is, widespread in both science and the general public. The basic idea is that particular cognitive and behavioral abilities correspond to specific anatomically and physiologically distinguishable modules in the brain that function largely independently of each other (Barrett & Kurzban, 2006; Fodor, 1983; Sternberg, 2011). A common simplified description of the relationship between cognition and the brain thus looks like this: Vision takes place in the visual cortex, the prefrontal cortex controls decisions, the hippocampus is the memory center, emotional experiences result from activity in the amygdala, and so on. Jerry Fodor, one of the most prominent figures of the cognitive revolution who shaped the paradigm of cognitivism within cognitive science, laid the conceptual foundation for such a modular view of cognition and the brain in his book *The Modularity of Mind* (Fodor, 1983). Before reconstructing Fodor's conception of a cognitive module, however, I will briefly discuss the rationale of *double dissociation* studies which serve as an even earlier basis for modular thinking in cognitive science.

Single and double dissociation studies are usually based on clearly localizable brain lesions resulting from a stroke or a brain tumor, e.g., in the hippocampus, and the observation of a loss or impairment of a specific cognitive ability following that brain lesion, e.g., memory. A single dissociation refers to an inference regarding the relationship between brain area B_1 and two different cognitive abilities C_1 and C_2. If a lesion in B_1 correlates with an impairment of C_1 but C_2 is not affected, it follows that B_1 is necessary for C_1 but not for C_2. A double

dissociation, on the other hand, refers to the relationship between two brain regions, B_1 and B_2, and two different cognitive abilities C_1 and C_2. If a lesion in B_1 correlates with an impairment of C_1 but not C_2, and if a lesion in B_2 correlates with an impairment of C_2 but not C_1, it follows that B_1 is necessary for C_1 and that B_2 is necessary for C_2. Furthermore, it is often claimed on the basis of a double dissociation that B_1 is not relevant for C_2 and that B_2 is not relevant for C_1. Thus, B_1 and B_2 can be dissociated from each other in terms of their respective functional specialization – this is the classical rationale of double dissociations, which is closely related to the idea of localizability of individual cognitive abilities at the level of different neuronal subsystems (Pessoa, 2022, pp. 50–53; Van Orden & Kloos, 2003; Van Orden et al., 2001).

In picking up on the idea of isolable and clearly localizable cognitive capacities in the brain, Fodor (1983) introduces the concept of a *cognitive module*. Fodor's theory is premised on adopting cognitivism, according to which cognition refers to brain-bound information processing based on neural representations, analogous to how computers process information (Fodor, 1983, p. 39). On Fodor's view, a cognitive module has the following properties: "Roughly, modular cognitive systems are domain specific, innately specified, hardwired, autonomous, and not assembled' (Fodor, 1983, p. 37). *Domain specific* means that a specific cognitive module performs a specific cognitive function (Fodor, 1983, pp. 47–52). Furthermore, according to Fodor, cognitive modules are not abstract organizational units but hardwired at the neuronal level (*innately specified, hardwired*). Since cognitive modules are not distributed across the brain but correspond to circumscribed brain areas (*not assembled*), this results in a modularization at both the functional and neuronal level such that each cognitive module corresponds to a circumscribed brain area to which a unique functional profile can be assigned (Fodor, 1983, pp. 98–100). The most important property of a cognitive module is the attribute *informational encapsulation* (Fodor, 1983, p. 103). This means that all relevant information necessary to execute any cognitive capacity is available only to the cognitive module that implements this specific cognitive capacity (*autonomous*). It is precisely this encapsulation in terms of information processing that results in the independence of cognitive modules from each other (Fodor, 1983, pp. 64–68).

To summarize, two important features of cognitive modules are particularly relevant for the discussion of Greene's dual-process moral psychology: First, a modular view of cognition implies an unambiguous mapping between cognitive abilities on the one hand and anatomically circumscribed brain areas on the other. Second, different cognitive modules are conceived as operating independently of each other due to the encapsulation of information processing in each module. This leads to the assumption of independent, localizable, spatially segregated, and

functionally specialized modules in the brain, each associated with a specific cognitive ability.[4]

3.2. Modularity in Greene

In light of the question whether Greene's dual-process theory of moral judgment presupposes a commitment to a modular account of cognition, it is crucial to be clear about what Greene actually means by "dual-processes". Is Greene merely talking about differences on the behavioral level, i.e., two different types of how people typically make moral judgments, or is Greene talking about two fundamentally different processes on the subpersonal level (e.g., the brain)? This is an important distinction because one might endorse a dual-process theory (of moral judgment) on the behavioral level without being committed to a modular view as to how these different behaviors may be related to differences in neurophysiology (see, e.g., Craigie, 2011; Kahneman & Frederick, 2002). However, depending on the kind of empirical data that is interpreted as evidence in favor of any dual-process theory, dual-process theories and a modular account of cognition and the brain often become a package deal. In a footnote, Greene (2014) elaborates on what exactly he means by "dual-processes":

> One might ask, what, exactly, is `dual' in dual-process theories? Is it types of processing? Types of cognitive systems? Different brain regions? Different kinds of outputs? The answer is "all of the above", but the core difference, in my view, concerns the type of processing. As Turing taught us, dual-process psychology can be implemented or simulated on a computer using a single processing system occupying a single physical location and using one set of computational principles (at low levels, at least). But, as it happens, that's not how it generally works in the brain. Instead, *distinct neural systems typically engage in distinct types of processing in distinct locations*. Likewise, cognitive outputs typically reflect the kinds of processing that produced them. As a result, *a brain injury can alter behavior because it causes damage to a particular location, which houses a particular cognitive system, which processes information in a particular way, and which therefore tends to push behavior in a particular direction*. Of course, not all dual-process dissociations are so clean, but sometimes they are. (Greene, 2014, p. 697; emphasis added)

As it becomes clear from the quote above, Greene is talking not only about two different processes on the behavioral level but about two different neurophysiological subsystems, and he also makes clear that these subsystems correspond to distinct circumscribed brain areas which engage in "distinct types of processing", just as in Fodor's definition of a cognitive module. In particular, Greene specifies that he associates the automatic mode with the ventral striatum and the ventromedial prefrontal cortex (VMPFC), whereas he localizes the manual mode in the dorsolateral prefrontal cortex (DLPFC) (Greene, 2014, pp. 697–698).

Moreover, in a recent response to De Neys (2023) who criticized the presumed exclusivity of two processes in dual-process theories and the idea that one processes is necessarily faster than the other, Greene (2023) once again clarified what he considers as the most important empirical evidence in support of his dual-process theory of moral judgment, namely brain lesion studies from patients with lesions in the VMPFC or the hippocampus that were associated with increases in characteristically consequentialist or deontological moral judgments, respectively. Here again we can see the classical rationale of double dissociation studies as alleged evidence for the existence of distinct cognitive modules (Pessoa, 2022, pp. 50–53; Van Orden & Kloos, 2003; Van Orden et al., 2001).

Since modular accounts of cognition are still widespread (Barrett & Kurzban, 2006; Drayson, 2017; Palecek, 2016; Sternberg, 2011), it may not appear too surprising that Greene's dual-process theory of moral judgment rests upon a modular view of cognition and the brain. However, as I will show in the following section, the modularity of mind is increasingly challenged in those quarters of cognitive science that adopt dynamical systems theory as their preferred modeling framework to study cognition and the brain. I will now address some objections to the modularity of mind in order to highlight that modular accounts of cognition and the brain should not be misconceived as being based solely on empirical facts, such as lesion studies, but rather entail important philosophical assumptions that can be called into question.

4. Beyond modularity: A dynamical systems perspective

The starting point for skepticism about the modularity of mind are empirical observations based on a large number of neuroimaging studies which have shown that increased activity in one brain area is usually associated with a variety of cognitive abilities, and that a particular cognitive ability is in turn usually associated with activity in several brain areas (Anderson & Pessoa, 2011; Anderson et al., 2013; Pessoa, 2008). These findings give reason to deviate from the idea of a *one-to-one mapping* between cognitive abilities and brain areas and to think of a *many-to-many mapping* instead (Anderson, 2010; Pessoa, 2008, 2014, 2022). The by now less controversial assumption of a many-to-many mapping between cognitive abilities and brain areas gives rise to the so-called *reverse inference* problem in cognitive neuroscience (Amodio, 2010; Pessoa, 2014): If experiments show that cognitive ability C_1 correlates with increased activity in brain area B_1, it cannot be concluded – under the assumption of a many-to-many mapping – that if increased activity was found in brain area B_1 in another situation, C_1 would also be performed. Reverse inference is a problem for those research approaches such as Greene's that rely on earlier research that has established

an association between primarily fast, emotional processes (C_1) and brain area B_1, and another association between primarily slow, deliberative processes (C_2) and brain area B_2. If in another situation, like making moral judgments of trolley dilemmas, an association is found between B_2 and a certain behavior C_3, for example characteristically consequentialist moral judgments, it is invalid to conclude – against the background of a many-to-many mapping – that C_3 is ultimately nothing but an instance of C_2.

Furthermore, the interpretation of empirical data from double dissociation studies that arguably provide the strongest support of a modular conception of the brain is heavily theory-laden. Without additional theoretical assumptions, double dissociation studies can, at best, provide data about necessary but not sufficient conditions regarding the involvement of particular neurophysiological processes for certain cognitive abilities (Fuchs, 2018, pp. 46–52). Only if we have already accepted the assumptions of single causes, reductionism, decomposability, and localizationism, we will interpret empirical data from double dissociations as sufficient evidence for the existence of independent cognitive modules. However, such circular reasoning begs the question whether to accept the modularity of mind as a premise or not (Van Orden & Kloos, 2003; Van Orden & Paap, 1997; Van Orden et al., 2001).

Let us examine Greene's arguably strongest evidence in favor of his dual-process theory of moral judgment again to illustrate the question begging nature of the double dissociation rationale. Greene (2023) refers to studies involving patients with brain lesions that reported a relative increase in characteristically consequentialist moral judgments among patients with VMPFC lesions and a relative increase in characteristically deontological moral judgments among patients with damage to the hippocampus. An uncontroversial conclusion we can draw from these findings is that there seem to be two causal factors (VMPFC and hippocampus impairment, respectively) that influence the relative probability of characteristically deontological and consequentialist moral judgments in opposite direction. In order to follow Greene's (2023) reasoning, according to whom the former lesion studies demonstrate that characteristically deontological moral judgments are `housed' in the VMPFC, whereas characteristically consequentialist moral judgments are independently generated by the hippocampus, we have to make an additional assumption. Only if we presume that cognitive functioning can and needs to be explained by appeal to *single causes*, i.e., if we already endorse the general conception that each cognitive capacity can be attributed to a single part of the brain, we will conclude that characteristically deontological and consequentialist moral judgments are enabled by different cognitive modules, because we have identified two non-identical causal factors that seem to affect each class of moral judgments differently (Van Orden & Paap, 1997; Van Orden et al., 2001).

On the contrary, if we adopt a dynamicist conception and assume that cognitive functioning emerges from multiple, nonlinearly interdependent processes within the brain and beyond, i.e., if we reject the assumption of single causes, we will conclude from the previous double dissociation studies that we have identified one out of many causally necessary brain processes for each class of moral judgments (VMPFC functioning for characteristically deontological moral judgments and hippocampus functioning for characteristically consequentialist moral judgments). Importantly, on a dynamicist view, it is not justified to conclude that characteristically consequentialist and deontological moral judgments, respectively, emerge from *independent* or fundamentally different processes. All that has been demonstrated so far is that at least one out of those many interdependent processes differs between the two classes of moral judgments, and this is exactly what we should expect if we observed two different behaviors. What we do not know, based on the empirical data presented by Greene, is to what extent the remaining processes are similar or different, and it may turn out that the two sets of interdependent processes that eventually qualify as sufficient causes for each class of moral judgment are largely overlapping and thus far from independent.

Hence, we should not uncritically accept Greene's reference to brain lesion studies and his appeal to the traditional rationale of double dissociations as evidence for two distinct cognitive modules, and thus for two distinct processes of making moral judgments, but rather take into account the plausibility of a non-modular architecture of the brain as well. But how exactly shall we conceptualize the organization of the brain if we reject the assumption of modularity?

It would be a false dichotomy to believe that the only alternative to independent modules in the brain with clearly definable functional profiles was to think of the brain as an equipotential gray mass that lacks any functional specialization and thus makes more fine-grained investigations impossible (Noble et al., 2024; Prinz, 2006). The alternative to the modularity of mind is rather a complex systems way of thinking with a focus on dynamically forming and changing functionally specialized networks in the brain – a way of thinking that originates from dynamical systems theory and that offers an alternative to reductionist and engineering-inspired approaches in cognitive science. As Pessoa (2022) puts is:

> In a highly networked system like the brain, we need to shift from thinking in terms of isolated brain regions and adopt the language of networks: Networks of brain regions collectively support behaviors. *The network itself is the unit*, not the brain area [. . .]. Consequently, processes that support behavior are not implemented by an individual area but depend on the interaction of multiple areas, which are dynamically recruited into multiregion assemblies [. . .]. (Pessoa, 2022, pp. 176–177; emphasis original)

Two aspects are particularly important: First, the focus on networks instead of modules, and, second, the *dynamics* of these networks. In contrast to a module in the Fodorian sense, which refers to an anatomically clearly delimited brain area with a unique functional profile, a network refers to a collection of different brain areas distributed across the cortex and subcortex that do not have to be in spatial proximity to each other but are still functionally connected (Pessoa, 2014). Anatomical studies have shown that there are numerous connections between most brain regions, even if these are further removed from each other (Modha & Singh, 2010; Pessoa, 2022, pp. 168–173; Young et al., 1994). Instead of conceiving the brain as being composed of discrete units, we are dealing with a system in which, in principle, everything is connected to everything else, although there are differences in how densely different regions are connected to each other. The whole brain is thus already one large interconnected network within which there appear to be subnetworks with different functional specializations. In consequence, the multitude of anatomical connections in the brain are a prerequisite for a situationally changing *functional connectivity* between those brain areas that are engaged in the same subnetwork (Pessoa, 2022, pp. 189–190).

A possible misunderstanding would be to think of these subnetworks simply as larger modules that encompass several brain regions and have a broader functional profile. This is where the second aspect, namely *dynamics*, comes into play (Pessoa, 2014, 2022, pp. 186–187): Both structurally and functionally, the brain is constantly changing, a phenomenon known as plasticity. On a relatively slower time scale (ranging from minutes to years), structural changes take place (neurons die, synaptic connections are strengthened or weakened) and on a relatively faster time scale (ranging from milliseconds to minutes), the functional composition of networks change. A defining characteristic of such a dynamic network is that there is strong functional connectivity between those brain areas within the network, whereas there is less functional connectivity to other brain areas outside of the dynamic network. This means that during a certain cognitive activity, such as reading this text, different brain areas engage in synchronized activity and thus form a network for the duration of the reading activity. As soon as the reading activity ends, the functional unit of the network dissolves again and a functional reorganization takes place in the brain. Networks are thus not static but dynamic. For different cognitive activities, different functionally specialized networks emerge and dissolve dynamically and can be distinguished from each other, even though different networks can involve the same brain areas for different cognitive activities, depending on the exact type of dynamic functional connectivity. At the same time, the dynamic composition of brain networks is not arbitrary – vision, for example, always involves the primary visual cortex –

and yet particular cognitive capacities cannot be statically reduced to activity in individual brain regions, for cognitive functioning depends on the state of the whole system and its subnetworks which evolves dynamically (Pessoa, 2014, 2022, pp. 167–192). Metaphorically speaking, we can use the image of a dance to gain a better understanding of the dynamic network perspective:

> As the mind fluctuates from state to state, we can view networks cohering and dissolving correspondingly – not unlike a group of dancers merging and separating as an act progresses. The temporal evolution of their joint states is what is important. (Pessoa, 2022, p. 187)

The context-dependent dynamics of the formation and functional specialization of particular brain networks gives us a better understanding of the empirical observation of a many-to-many mapping between cognitive abilities and brain areas, which continues to exist at the network level (Pessoa, 2014). Since networks are based on dynamically emerging and dissolving functional connectivity, networks consisting of the same brain areas can be involved in different cognitive activities and networks consisting of different brain areas can be involved in similar cognitive activities. The exact type of functional connectivity varies dynamically and is ultimately decisive for the functional profile of a network. Against the background of such a dynamic and context-dependent functional specialization at the network level, context-invariant, static assignments between cognitive abilities and individual brain areas in the sense of distinct cognitive modules are no longer tenable. To be clear, there is functional specialization at the network level – in contrast to equipotentialism – but such a functional specialization must be understood dynamically and context-dependent (Pessoa, 2014).

A further objection that goes beyond the criticism of the modularity of mind, but includes its rejection, comes from embodied cognition approaches within cognitive science, especially from ecological psychology and enactivism (Chemero, 2009; Fuchs, 2018; Gibson, 1979/2015; Segundo-Ortin & Raja, 2024, Thompson & Varela, 2001; Varela et al., 1991/2017). Embodied cognition rejects neuroreductionism, according to which cognitive abilities and psychological phenomena can be reduced to neuronal activity. Instead, ecological psychologists and enactivists focus on the brain-body-environment system as the central unit of analysis for understanding cognition, and they also endorse dynamical systems theory as their preferred modeling framework, but at the level of the more encompassing brain-body-environment system and not just the brain. Ecological validity plays a key role from an embodied perspective, and thus a comprehensive investigation of how humans make different types of moral judgments must consider not only potential differences in neurophysiology but also the fact that particular brain processes associated with different types of moral judgments cannot be understood in isolation from a history of embodied

human to human interactions, embedded in particular socio-cultural environments which are constitutive for the development of these cognitive abilities. On this view, one could formulate a more radical critique of Greene's research program that goes beyond a critique of a modular account of cognition and instead questions the ecological validity of his experimental approach: Lying in a MRI scanner while making moral judgments about trolley dilemmas seems to be far removed from an ecologically valid investigation of moral judgments.

Overall, however, it is important to note that despite the objections raised above against the modularity of mind, a dynamical systems perspective has not yet become the new dominant paradigm in cognitive science and neuroscience. Although there is an increasing tendency in neuroscience under the umbrella term *network neuroscience* (Bassett & Sporns, 2017) to replace a modular perspective by a network approach, and although there is a growing community of cognitive scientists who have endorsed dynamical systems theory as their preferred modeling framework for almost three decades (Chemero, 2009, 2023; Favela, 2020, 2021, 2024; Izhikevich, 2007; Kelso, 1995; Nguyen et al., 2024; Noble et al., 2024; Pessoa, 2014, 2022; Smith & Thelen, 1994, Thompson & Varela, 2001; Van Orden et al., 2003, 2001; van Gelder, 1998), there are still other quarters in cognitive science who adhere to a modular account of cognition (Barrett & Kurzban, 2006; Drayson, 2017; Palecek, 2016; Sternberg, 2011). The current state of disagreement in cognitive science as to whether or not to abandon a modular account of cognition points toward the "real disagreement" between proponents of modularity and dynamical systems theory. As Van Orden et al. (2001) pointed out, we should not mistakenly believe that the controversy around the modularity of mind is a purely empirical question. What is actually at stake are two opposing views on the kind of epistemic standards that a good explanation in cognitive science and neuroscience should adhere to. The modularity of mind adheres to the epistemic ideals of reductionism, single causes, decomposability, and localizationism, whereas dynamical systems approaches in cognitive science adhere to the epistemic ideals of holism, anti-reductionism, non-decomposability, and anti-localizationism (Noble et al., 2024; Sanches de Oliveira & Chemero, 2015; Silberstein & Chemero, 2013; Van Orden & Paap, 1997; Van Orden et al., 2001). Both schools of thought – modularity and dynamicism – extensively refer to empirical evidence, and yet we should not expect that this debate will be settled on a purely empirical basis, because it was never a purely empirical debate to begin with. Hence, we have to adjudicate between modular and dynamical accounts of cognition and the brain on a conceptual or pragmatic basis.

Interestingly, Fodor (1983) already seemed to be aware that the plausibility of the modularity of mind is closely linked to the kind of epistemic standards that one implicitly or explicitly applies to good explanations in

cognitive science. Toward the end of his book, Fodor makes the following revealing remark:

> I promised a parting word or two about what the prospects for research in cognitive science might be like assuming that the modified modularity thesis is true. My point will be this: *the limits of modularity are also likely to be the limits of what we are going to be able to understand about the mind*, given anything like the theoretical apparatus currently available. (Fodor, 1983, p. 126; emphasis added)

Since our currently available theoretical apparatus has changed dramatically since the publication of *The Modularity of Mind* in 1983, especially in dynamical systems theory in cognitive science, we do not have to worry anymore that we will not be able to understand anything about cognition and the brain if we let go of the assumption of modularity. Yet, Fodor's statement is revealing in that it points toward the "real motivation" of adhering toa modular account of cognition and the brain, which seems to be primarily rooted in epistemic standards that are at odds with those epistemic standards favored by dynamicists. Moreover, Fodor is right in that particular types of explanations crucially depend on endorsing a modular view of cognition. If we reject the modularity of mind, we will have to let go of certain types of explanations as well. This brings us back to Greene's dual-process theory and a key question that has been waiting in the background so far: If we reject the modularity of mind as a premise, will this be the end to Greene's dual-process theory of moral judgment?

5. The end or the future of dual-process moral psychology?

Against the background of the previous discussion of the modularity of mind, Greene's empirical findings seem ambiguous without the addition of further theoretical assumptions. If we reject a modular account of cognition and the brain on a theoretical basis, it does not follow from Greene's neuroimaging studies and other lesion studies that there are two asymmetrically independent cognitive modules that each enable either characteristically deontological or characteristically consequentialist moral judgments. Instead, we can only conclude that the respective brain areas seem to be involved somehow in those neurophysiological processes that contribute to the respective moral judgments. On the latter interpretation, however, it does not follow that we are dealing with two fundamentally different systems of how humans make moral judgments.

As I have pointed out earlier, adopting a dynamical systems perspective does not mean questioning that there is functional specialization in the brain. What it does question, however, is a static and context-invariant mapping between cognitive abilities and isolated brain regions. It could be, for example, that characteristically deontological

and characteristically consequentialist moral judgments are related to two different dynamically forming brain networks, and such a relative difference in the functional specialization of two dynamic brain networks could be demonstrated by using other imaging techniques and analysis methods than those used by Greene (Sporns & Betzel, 2016). However, it is crucial to note that, from a dynamical systems perspective, there can never be complete independence between the components of a system, because it is precisely the *interdependence* of the components of a system that characterizes the system as a nonlinear dynamical system (Nguyen et al., 2024; Pessoa, 2022, pp. 129–143; Van Orden et al., 2003). A relative, dynamic, and context-sensitive difference in the functional specialization of two brain networks would thus be too weak for Greene's dual-process theory which postulates the existence of two asymmetrically *independent* processes, i.e., two distinct cognitive modules.

This sounds much like the end of Greene's dual-process moral psychology, but this need not be the case. The debate around Greene's dual-process theory of moral judgment has shifted in a direction that generally accepts the hypothesis of two different psychological and neurophysiological processes contributing to different types of moral judgments, but the presumed exclusivity of these processes has been questioned increasingly (Craigie, 2011; Crockett, 2013; Cushman, 2013; De Neys, 2023; see also Wiegmann & Sauer, 2021). Such deflated versions of dual-process theories may be compatible with a dynamical systems perspective, for they refer to differences in the relative contribution of different subsystems in the brain and how these are dynamically integrated, rather than positing categorical differences between isolable cognitive modules and their respective functional specialization. Although updated and less radical versions of dual-process theories in moral psychology were not motivated by a rejection of the modularity of mind, but by empirical findings that gave reasons to reconsider the presumed exclusivity and asymmetrical independence of two processes of moral judgment (for review, see De Neys, 2023), both approaches – either a rejection of the modularity of mind on a theoretical basis, or the development of less radical versions of dual-process theories to accommodate new empirical evidence conflicting Greene's original dual-process theory – ultimately seem to converge on an appreciation of relative instead of categorial differences as to how differences in moral judgments are related to neurophysiology. However, Greene's original dual-process theory of moral judgment seems to be inconceivable without a commitment to a modular account of cognition and the brain, and thus I believe it is important to pay attention to this implicit premise when it comes to the future of dual-process moral psychology.

Zooming out from the present discussion of Greene's dual-process theory of moral judgment to the ongoing debate between proponents of either single or dual-process theories in psychology more generally (for review, see De Neys, 2021; Evans & Stanovich, 2013), there seems to be a relevant but largely undiscussed parallel between the single versus dual-process literature, on the one hand, and the modularity versus dynamical systems literature, on the other hand (see also Zerilli, 2023). Despite the absence of a single, "received" dual-process theory that would apply generically to all aspects of cognition (Evans, 2012; Evans & Stanovich, 2013), the disagreement between supporters and critics of various dual-process theories in different psychological domains comes down to the question whether there are qualitative or merely quantitative differences between intuitive (analogous to Greene's (2014) "automatic mode") and deliberate (analogous to Greene's (2014) "manual mode") processes (De Neys, 2021). Critics of dual-process theories thus do not deny that there seem to be differences between more intuitive and more deliberate cognitive processes, but they hold that these differences are merely quantitative (or relative) rather than qualitative (or categorical).

Notably, most of the arguments in favor or against dual-process theories mainly rely on behavioral data or conceptual issues (De Neys, 2021, 2023; Evans, 2011; Evans & Stanovich, 2013). Although supporters of dual-process theories often refer to data from neuroimaging or lesion studies as additional support of their view in a similar vein as Greene (2014, 2023), which may be indicative of an implicit commitment to a modular account of cognition and the brain (see Evans, 2011; Evans & Stanovich, 2013), the neural perspective seems to play a secondary role in the wider single versus dual-process literature. That is, the debate on single versus dual-process theories does not solely depend on taking a stance in the modularity versus dynamicism debate, even though these two debates are closely connected and sometimes come as a package deal, as I have shown with respect to Greene's dual-process theory of moral judgment and its implicit commitment to the modularity of mind.

Moreover, against the backdrop of Greene's (2014) normative claims, the general idea that normative correctness might be attributed to deliberate rather than intuitive cognitive processes has been critically discussed in the wider dual-process literature too (De Neys, 2020; Evans, 2011, 2012). According to Evans (2012), it constitutes a normative fallacy to generally attribute normative correctness to deliberate processes and to view intuitive processes as necessarily biased. Such a normative fallacy seems to result from an unjustified generalization of those dual-process theories that were originally developed in the context of logical and probabilistic reasoning and that provided an explanation of participants' reasoning mistakes that were primarily associated with insufficient deliberation. Within the context

of logical and probabilistic reasoning, it may be justified to attribute normative correctness to deliberate rather than intuitive processes, but the normative fallacy lies in viewing normative correctness as a general feature of deliberation across all the different contexts to which dual-process theories have been applied (De Neys, 2020; Evans, 2012). On a charitable reading, however, Greene (2014) cannot simply be accused of falling prey to the normative fallacy, since Greene does not consider all intuitive moral judgments to be biased. Only if we are confronted with unfamiliar moral problems, Greene holds that deliberate moral judgments will be normatively superior compared to their intuitive counterpart (see section 2). This normative premise of Greene's (2014) indirect route argument may be evaluated in its own right, but it certainly differs in relevant respects from the canonical normative fallacy as discussed in the wider dual-process literature.

Zooming back in on Greene's dual-process theory of moral judgment, a rejection of the modularity of mind ultimately imposes constraints on the kind of insights that we can expect from an empirically informed ethics. Greene's promise to solve a long-standing debate in moral philosophy between consequentialist and deontological approaches hinges on the premise of a modular account of cognition and the brain. If we let go of the modularity of mind, we must also abandon the hope of a categorical solution of the debate between deontological and consequentialist approaches in normative ethics based on a couple of brain scans. For some, this may sound like biting the bullet, but an appreciation of modesty could also be the starting point for a new research program in empirically informed ethics that aims at illuminating relative differences in dynamic brain functioning that can still be highly informative, albeit less categorical and less bold then what Greene's dual-processes theory of moral judgment was initially aiming for.

6. Conclusion

In this paper, I have argued that Greene's dual-process theory of moral judgment presupposes a commitment to a modular account of cognition and the brain. Only under the assumption of a modular architecture of the brain, the evidence from neuroimaging and lesion studies presented by Greene supports the view that characteristically consequentialist and characteristically deontological moral judgments are enabled by asymmetrically independent processes, i.e., by two distinct cognitive modules. Against the background of growing influence of dynamical systems thinking in cognitive science and neuroscience, I have questioned the plausibility of the modularity of mind, and I have presented an alternative view of the brain as a complex dynamical system that is not composed of independent cognitive modules but of dynamically forming and dissolving, functionally specialized brain networks. Ultimately, accepting or rejecting a modular

account of cognition and the brain is not just a matter of empirical evidence, but a matter of conflicting epistemic standards that proponents of modularity or dynamicism apply to good explanations in cognitive science. If we reject the modularity of mind, it follows that Greene's dual-process theory of moral judgment is no longer tenable in its original version. However, dynamical systems accounts may be compatible with more nuanced versions of dual-process theories that focus on how different types of moral judgments might be related to relative differences in the dynamic recruitment and integration of different brain networks. A rejection of the modularity of mind thus does not necessarily put an end to dual-process theories in moral psychology, but it urges us to reconsider crucial theoretical commitments that constrain what kinds of explanations we can expect from an empirically informed ethics.

Note

1. Note that not all dual-process theories in psychology are committed to a modular account of cognition and the brain. For example, Kahneman, one of the best known dual-process psychologists, points out that "System 1" and "System 2" processes should not be misconceived as independent cognitive modules: "We adopt the generic labels `System 1' and `System 2' [...]. *These terms may suggest the image of autonomous homunculi, but such a meaning is not intended*. We use systems as a label for collections of processes that are distinguished by their speed, controllability, and the contents on which they operate." (Kahneman and Frederick, 2002, p. 51; emphasis added).
2. According to Greene et al. (2001), the key difference between the trolley dilemma and the footbridge dilemma is that the trolley dilemma exemplifies an *impersonal* moral dilemma, whereas the footbridge dilemma exemplifies a *personal* moral dilemma; the latter referring to direct physical interaction between the intervening agent and the potential victim. In subsequent works, Greene and colleagues refined the distinction between personal and impersonal moral dilemmas as being a matter of *personal force* more generally (Greene, 2014; Greene et al., 2009).
3. According to Greene (2014, pp. 699-700), the qualification "characteristic" underlines that characteristically deontological or characteristically consequentialist moral judgments must not be equated with any particular ethical theory that can be labeled as either consequentialist or deontological. Instead, according to Greene, moral judgments that can be justified by weighing up costs and benefits count as characteristically consequentialist (e.g. sacrificing one person to save five). Characteristically deontological moral judgments, on the other hand, are those that can be better justified by reference to rights and duties (e.g. "You must not kill!").
4. Fodor (1983) limits the scope of his modular account of cognition and the brain by highlighting that not all cognitive capacities are modularly organized. Fodor distinguishes between ``central'' and ``peripheral'' cognitive systems, and only the latter are modularly organized on Fodor's view. Deliberate thinking and other so-called higher cognitive abilities count as ``central systems'' and are thus non-modularly organized, according to Fodor. In contrast, Greene's dual-process theory rather seems to entail a commitment to *massive modularity*, i.e., a more radical version of Fodor's original theory, according to which *all* aspects of cognition and the entire brain are

modularly organized (Barrett & Kurzban, 2006; Carruthers, 2006; Prinz, 2006; Zerilli, 2023). What matters for the present discussion is the concept of a cognitive module as originally introduced by Fodor, even if Greene's interpretation of the modularity of mind is more in line with what is known as *massive modularity* and thus diverges in scope from Fodor's original view.

Acknowledgements

I would like to thank Norbert Paulo, Sebastian Wallot, Alex Wiegmann, and two anonymous reviewers for helpful feedback on earlier versions of this paper. Moreover, I would like to thank Maurice Pflug for many stimulating discussions that have helped to sharpen some of the points in this paper.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

The author(s) reported there is no funding associated with the work featured in this article.

ORCID

Moritz Bammel http://orcid.org/0000-0002-0321-2864

References

Amodio, D. M. (2010). Can neuroscience advance social psychological theory? Social Neuroscience for the behavioral social psychologist. *Social Cognition*, *28*(6), 695–716. https://doi.org/10.1521/soco.2010.28.6.695

Anderson, M. L. (2010). Neural reuse: A fundamental organizational principle of the brain. *Behavioral and Brain Sciences*, *33*(4), 245–266. https://doi.org/10.1017/S0140525X10000853

Anderson, M. L., Kinnison, J., & Pessoa, L. (2013). Describing functional diversity of brain regions and brain networks. *Neuroimage*, *73*, 50–58. https://doi.org/10.1016/j.neuroimage.2013.01.071

Anderson, M. L., Pessoa, L. (2011). Quantifying the diversity of neural activations in individual brain regions. *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 33. pp. 2421–2426). https://escholarship.org/uc/item/8fg7h0cr

Barrett, H. C., & Kurzban, R. (2006). Modularity in cognition: Framing the debate. *Psychological Review*, *113*(3), 628–647. https://doi.org/10.1037/0033-295X.113.3.628

Bassett, D. S., & Sporns, O. (2017). Network neuroscience. *Nature Neuroscience*, *20*(3), 353–364. https://doi.org/10.1038/nn.4502

Berker, S. (2009). The normative insignificance of neuroscience. *Philosophy & Public Affairs*, *37*(4), 293–329. https://doi.org/10.1111/j.1088-4963.2009.01164.x

Bublitz, J. C., & Paulo, N. (2020). Empirische Ethik: Hintergruende, Einwaende, Potentiale. In N. Paulo & J. C. Bublitz (Eds.), *Empirische Ethik. Grundlagentexte aus Psychologie und Philosophie* (pp. 9–72). Suhrkamp.

Carruthers, P. (2006). The case for massively modular models of mind. In R. J. Stainton (Ed.), *Contemporary debates in cognitive science* (pp. 205–225). Blackwell Publishing.

Chemero, A. (2009). *Radical embodied cognitive science*. The MIT Press.

Chemero, A. (2023). Abduction and deduction in dynamical cognitive science. *Topics in Cognitive Science*. https://doi.org/10.1111/tops.12692

Craigie, J. (2011). Thinking and feeling: Moral deliberation in a dual-process framework. *Philosophical Psychology*, *24*(1), 53–71. https://doi.org/10.1080/09515089.2010.533262

Crockett, M. J. (2013). Models of morality. *Trends in Cognitive Sciences*, *17*(8), 363–366. https://doi.org/10.1016/j.tics.2013.06.005

Cushman, F. (2013). Action, outcome, and Value: A dual-system framework for morality. *Personality and Social Psychology Review*, *17*(3), 273–292. https://doi.org/10.1177/1088868313495594

Dale, M. T. (2020). Neurons and normativity: A critique of Greene's notion of unfamiliarity. *Philosophical Psychology*, *33*(8), 1072–1095. https://doi.org/10.1080/09515089.2020.1787972

Dale, M. T., & Gawronski, B. (2022). Brains, trains, and ethical claims: Reassessing the normative implications of moral dilemma research. *Philosophical Psychology*, *36*(1), 109–133. https://doi.org/10.1080/09515089.2022.2038783

Dean, R. (2009). Does neuroscience undermine deontological theory? *Neuroethics*, *3*(1), 43–60. https://doi.org/10.1007/s12152-009-9052-x

De Neys, W. (2020). Morality, normativity, and the good system 2 fallacy. *Diametros*, 1–6. https://doi.org/10.33392/diam.1447

De Neys, W. (2021). On dual- and single-process models of thinking. *Perspectives on Psychological Science*, *16*(6), 1412–1427. https://doi.org/10.1177/1745691620964172

De Neys, W. (2023). Advancing theorizing about fast-and-slow thinking. *Behavioral and Brain Sciences*, *46*. https://doi.org/10.1017/S0140525X23000559

Drayson, Z. (2017). Modularity and the predictive mind. In T. K. Metzinger & W. Wiese (Eds.), *Philosophy and predictive processing* (pp. 1–12). MIND Group. https://doi.org/10.15502/9783958573130

Evans, J. S. B. T. (2011). Dual-process theories of reasoning: Contemporary issues and developmental applications. *Developmental Review*, *31*(23), 86–102. https://doi.org/10.1016/j.dr.2011.07.007

Evans, J. S. B. T. (2012). Dual-process theories of deductive reasoning: Facts and fallacies. In K. J. Holyoak & R. G. Morrison (Eds.), *The oxford handbook of thinking and reasoning* (pp. 115–133). Oxford University Press.

Evans, J. S. B. T., & Stanovich, K. E. (2013). Dual-process theories of higher cognition: Advancing the debate. *Perspectives on Psychological Science*, *8*(3), 223–241. https://doi.org/10.1177/1745691612460685

Favela, L. H. (2020). Cognitive science as complexity science. *WIREs Cognitive Science*, *11*(4). https://doi.org/10.1002/wcs.1525

Favela, L. H. (2021). The dynamical renaissance in neuroscience. *Synthese*, *199*(12), 2103–2127. https://doi.org/10.1007/s11229-020-02874-y

Favela, L. H. (2024). *The ecological brain: Unifying the sciences of brain, body, and environment*. Routledge.

Fodor, J. A. (1983). *The modularity of mind*. MIT Press.

Fuchs, T. (2018). *Ecology of the brain: The phenomenology and biology of the embodied mind*. Oxford University Press.

Gibson, J. J. (1979/2015). *The ecological approach to visual perception*. Psychology Press.
Greene, J. D. (2013). *Moral tribes: Emotion, reason and the gap between us and them*. Atlantic Books.
Greene, J. D. (2014). Beyond point-and-shoot morality: Why cognitive (neuro)science matters for ethics. *Ethics*, *124*(4), 695–726. https://doi.org/10.1086/675875
Greene, J. D. (2023). Dual-process moral judgment beyond fast and slow. *Behavioral and Brain Sciences*, *46*. https://doi.org/10.1017/S0140525X22003193
Greene, J. D., Cushman, F. A., Stewart, L. E., Lowenberg, K., Nystrom, L. E., & Cohen, J. D. (2009). Pushing moral buttons: The interaction between personal force and intention in moral judgment. *Cognition*, *111*(3), 364–371. https://doi.org/10.1016/j.cognition.2009.02.001
Greene, J. D., Nystrom, L. E., Engell, A. D., Darley, J. M., & Cohen, J. D. (2004). The neural bases of cognitive conflict and control in moral judgment. *Neuron*, *44*(2), 389–400. https://doi.org/10.1016/j.neuron.2004.09.027
Greene, J. D., Sommerville, R. B., Nystrom, L. E., Darley, J. M., & Cohen, J. D. (2001). An fMRI investigation of emotional engagement in moral judgment. *Science*, *293*(5537), 2105–2108. https://doi.org/10.1126/science.1062872
Heir, A. S. (2023). Rational intuitions: How reason underlies deontological moral judgments. *Philosophical Psychology*, 1–25. https://doi.org/10.1080/09515089.2023.2290172
Izhikevich, E. M. (2007). *Dynamical systems in neuroscience: The geometry of excitability and bursting*. The MIT Press.
Kahane, G. (2012). On the wrong track: Process and content in moral psychology. *Mind & Language*, *27*(5), 519–545. https://doi.org/10.1111/mila.12001
Kahane, G., & Shackel, N. (2010). Methodological issues in the neuroscience of moral judgement. *Mind & Language*, *25*(5), 561–582. https://doi.org/10.1111/j.1468-0017.2010.01401.x
Kahane, G., Wiech, K., Shackel, N., Farias, M., Savulescu, J., & Tracey, I. (2012). The neural basis of intuitive and counterintuitive moral judgment. *Social Cognitive and Affective Neuroscience*, *7*(4), 393–402. https://doi.org/10.1093/scan/nsr005
Kahneman, D., & Frederick, S. (2002). Representativeness revisited: Attribute substitution in intuitive judgement. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics of intuitive judgement: Extensions and applications* (pp. 49–81). Cambridge University Press.
Kamm, F. M. (2009). Neuroscience and moral reasoning: A note on recent research. *Philosophy & Public Affairs*, *37*(4), 330–345. https://doi.org/10.1111/j.1088-4963.2009.01165.x
Kelso, J. A. S. (1995). *Dynamic patterns: The self-organization of brain and behavior*. The MIT Press.
Knigs, P. (2018). On the normative insignificance of neuroscience and dual-process theory. *Neuroethics*, *11*(2), 195–209. https://doi.org/10.1007/s12152-018-9362-y
Kumar, V., & Campbell, R. (2012). On the normative significance of experimental moral psychology. *Philosophical Psychology*, *25*(3), 311–330. https://doi.org/10.1080/09515089.2012.660140
Lott, M. (2016). Moral implications from cognitive (neuro)science? No clear route. *Ethics*, *127*(1), 241–256. https://doi.org/10.1086/687337
Meyers, C. D. (2013). Brains, trolleys, and intuitions: Defending deontology from the Greene/Singer argument. *Philosophical Psychology*, *28*(4), 466–486. https://doi.org/10.1080/09515089.2013.849381
Mihailov, E. (2015). Is deontology a moral confabulation? *Neuroethics*, *9*(1), 1–13. https://doi.org/10.1007/s12152-015-9244-5

Modha, D. S., & Singh, R. (2010). Network architecture of the long-distance pathways in the macaque brain. *Proceedings of the National Academy of Sciences*, *107*(30), 13485–13490. https://doi.org/10.1073/pnas.1008054107

Nguyen, T. D., Magaldino, C. M., Landfair, J. T., Amazeen, P. G., & Amazeen, E. L. (2024). From cognitive agents to cognitive systems: Theoretical, methodological, and empirical developments of van Gelder's (1998) "dynamical hypothesis". *Topics in Cognitive Science*. https://doi.org/10.1111/tops.12725

Noble, S., Curtiss, J., Pessoa, L., & Scheinost, D. (2024). The tip of the iceberg: A call to embrace anti-localizationism in human neuroscience research. *Imaging Neuroscience*, *2*, 1–10. https://doi.org/10.1162/imag_a_00138

Palecek, M. (2016). Modularity of mind: Is it time to abandon this ship? *Philosophy of the Social Sciences*, *47*(2), 132–144. https://doi.org/10.1177/0048393116672833

Paulo, N. (2019). In search of Greene's argument. *Utilitas*, *31*(1), 38–58. https://doi.org/10.1017/S0953820818000171

Pessoa, L. (2008). On the relationship between emotion and cognition. *Nature Reviews Neuroscience*, *9*(2), 148–158. https://doi.org/10.1038/nrn2317

Pessoa, L. (2014). Understanding brain networks and brain organization. *Physics of Life Reviews*, *11*(3), 400–435. https://doi.org/10.1016/j.plrev.2014.03.005

Pessoa, L. (2022). *The entangled brain: How perception, cognition, and emotion are woven together*. MIT Press.

Prinz, J. J. (2006). Is the mind really modular? In R. J. Stainton (Ed.), *Contemporary debates in cognitive science* (pp. 22–36). Blackwell Publishing.

Sanches de Oliveira, G., & Chemero, A. (2015). Against smallism and localism. *Studies in Logic, Grammar and Rhetoric*, *41*(1), 9–23. https://doi.org/10.1515/slgr-2015-0017

Sauer, H. (2012). Morally irrelevant factors: What's left of the dual process-model of moral cognition? *Philosophical Psychology*, *25*(6), 783–811. https://doi.org/10.1080/09515089.2011.631997

Segundo-Ortin, M., & Raja, V. (2024). *Ecological psychology*. Cambridge University Press.

Silberstein, M., & Chemero, A. (2013). Constraints on localization and decomposition as explanatory strategies in the biological sciences. *Philosophy of Science*, *80*(5), 958–970.

Singer, P. (2005). Ethics and intuitions. *The Journal of Ethics*, *9*(34), 331–352. https://doi.org/10.1007/s10892-005-3508-y

Smith, L. B., & Thelen, E. (1994). *A dynamic systems approach to the development of cognition and action*. The MIT Press.

Sporns, O., & Betzel, R. F. (2016). Modular brain networks. *Annual Review of Psychology*, *67*(1), 613–640. https://doi.org/10.1146/annurev-psych-122414-033634

Sternberg, S. (2011). Modular processes in mind and brain. *Cognitive Neuropsychology*, *28*(34), 156–208. https://doi.org/10.1080/02643294.2011.557231

Thompson, E, & Varela, F. J. (2001). Radical embodiment: neural dynamics and consciousness. Trends in Cognitive Sciences, *5*(10), 418–425.

van Gelder, T. (1998). The dynamical hypothesis in cognitive science. *Behavioral and Brain Sciences*, *21*(5), 615–628. https://doi.org/10.1017/S0140525X98001733

Van Orden, G. C., Holden, J. G., & Turvey, M. T. (2003). Self-organization of cognitive performance. *Journal of Experimental Psychology General*, *132*(3), 331–350. https://doi.org/10.1037/0096-3445.132.3.331

Van Orden, G. C., & Kloos, H. (2003). The Module Mistake. *Cortex, A Journal Devoted to the Study of the Nervous System and Behavior*, *39*(1), 164–166. https://doi.org/10.1016/S0010-9452(08)70092-3

Van Orden, G. C., & Paap, K. R. (1997). Functional neuroimages fail to discover pieces of mind in the parts of the brain. *Philosophy of Science*, *64*(S4), S85–S94. https://doi.org/10.1086/392589

Van Orden, G. C., Pennington, B. F., & Stone, G. O. (2001). What do double dissociations prove? *Cognitive Science*, *25*(1), 111–172. https://doi.org/10.1207/s15516709cog2501_5

Varela, F. J., Thompson, E., & Rosch, E. (1991/2017). *The embodied mind* (Revised ed.). Cognitive Science and Human Experience. MIT Press.

Wiegmann, A., & Sauer, H. (2021). The psychology and rationality of moral judgment. In M. Knauff & W. Spohn (Eds.), *The handbook of rationality* (pp. 701–710). The MIT Press.

Young, M. P., Scanneil, J. W., Burns, G. A. P. C., & Blakemore, C. (1994). Analysis of connectivity: Neural systems in the cerebral cortex. *Reviews in the Neurosciences*, *5*(3). https://doi.org/10.1515/REVNEURO.1994.5.3.227

Zerilli, J. (2023). A tale of two histories: Dual-system architectures in modular perspective. *Behavioral and Brain Sciences*, *46*. https://doi.org/10.1017/S0140525X22002977