Phenomenology and the Cognitive Sciences
https://doi.org/10.1007/s11097-024-10000-3

Dual process theory and the challenges of functional individuation

James D. Grayot[1] · **Lukas Beck**[2] · **Thijs Heijmeskamp**[3]

Accepted: 31 May 2024

Abstract

Despite on-going debates in philosophy and cognitive science, dual process theory (DPT) remains a popular framework for theorizing about human cognition. Its central hypothesis is that cognitive processing can be subsumed under two generic types. In this paper, we argue that the putative success and popularity of this framework remains overstated and gives rise to certain misunderstandings. If DPT has predictive and/or explanatory power, it is through offering descriptions of cognitive phenomena via functional analysis. But functional descriptions require an individuation strategy. To date, there has been no systematic exploration of how Type 1 and Type 2 are functionally individuated. Following recent debates in philosophy of cognitive science, we consider three individuation strategies (i.e., abstraction, reification, fictionalization) and assess the legitimacy of each in relation to DPT. This leads us to the verdict that the most viable route for justifying DPT is to construe Type 1 and Type 2 processes as reifications. We conclude that, construed as reifications, the common rationales offered by proponents of DPT for demarcating Type 1 and Type 2 processes do not escape criticism and require further theoretical justification.

Keywords Dual Process Theory · Type 1 Processing · Type 2 Processing · Functionalism · Functional Individuation · Reification · Cognitive Processing · Higher Reasoning · Philosophy of Cognitive Science

✉ James D. Grayot
james.grayot@gmail.com

Lukas Beck
L.Beck1@lse.ac.uk

Thijs Heijmeskamp
thijsheijmeskamp@gmail.com

[1] Institute of Philosophy, University of Porto, Porto, Portugal

[2] Department of Philosophy, Logic and Scientific Method, London School of Economics, London, England

[3] Erasmus School of Philosophy, Erasmus University Rotterdam, Rotterdam, Netherlands

Published online: 25 June 2024

1 Introduction

Dual process theory (henceforth: DPT) is one of the most important developments in modern psychology and cognitive science. Its central hypothesis is that human reasoning, judgment, and decision-making can be subsumed under two cognitive processing *types*. Type 1 processing is understood to be *fast*, *automatic*, and *reactive* whereas Type 2 processing is *slow*, *controlled*, and *deliberative* (Evans, 2006, 2008; Evans & Stanovich, 2013a, 2013b). Supporters of DPT argue that the framework not only provides a sound schema for categorizing diverse styles of thinking, but also that it reliably predicts and explains how the differential processing of information plays various unique roles in shaping everyday behaviors.

But what is DPT, exactly? Whereas early interpretations of the DPT framework cast cognitive processing in terms of separate 'systems', i.e., *System 1* and *System 2* (Kahneman & Frederick, 2002; Stanovich, 1999), many supporters, especially Evans and Stanovich, have withdrawn from this convention and instead construe processing in terms of 'types'. This shift follows from criticisms that the language of 'systems' wrongly implies that such processes are (i) well-defined with respect to their associated tasks, and (ii) packaged in discrete, domain-specific modules, which have clear neurophysiological correlates in the human brain (see, e.g., Osman, 2004; Keren & Schul, 2009; Keren, 2013). The preferred language of types is meant to bypass these issues and be more neutral with respect to the discreteness of cognitive processes *vis-a-vis* their associated tasks and possible neural correlates. Thus reconstrued, *Type 1* processing consists of autonomous processes that are not under an individual's conscious control, whereas *Type 2* processing consists of reflective processes that are demanding of working memory, such as language-use, mental simulation, and complex problem-solving (Evans & Stanovich, 2013a, 2013b).

Nevertheless, there are important issues within the DPT framework that still need resolving. For instance, it is ambiguous and thus debated what the precise nature of the organization and interaction of the two processing types is (De Neys, 2021; Grayot, 2020). Proponents of the "default-interventionist" model maintain that Type 1 and Type 2 are organized sequentially, with Type 1 inputs serving as catalysts for Type 2 outputs—this places Type 2 processes in a position to 'monitor' and 'intervene' upon Type 1 processes (Evans, 2008; Kahneman & Frederick, 2005). By contrast, proponents of the "parallel-competitive" model suggest that Type 1 and Type 2 processes operate concurrently and have to 'compete' for control over different aspects of an agent's behavior (Sloman, 1996; Smith & DeCoster, 2000). This model supposes that Type 1 and Type 2 processes receive separate informational inputs and do not (on most readings) functionally overlap in their characteristic outputs. Although these debates are far from being settled, proponents of DPT are relentlessly confident that the architecture of human cognition *is* fundamentally

dualistic—it's just a matter of time until the right model vindicates the general framework (Evans, 2018, 153).[1]

This unwavering optimism by proponents of DPT stands in contrast to the growing corpus of critiques from scholars working outside the paradigm. Among other concerns, it has been argued that (i) the evidential support for DPT is primarily limited to controlled laboratory settings (Gigerenzer, 2015), (ii) that DPT takes an *as-if* approach to cognitive modeling (Berg & Gigerenzer, 2010, 2010), and, most damningly, (iii) that the framework offers little more than a list of general dichotomies in lieu of any "real theory" (Gigerenzer, 2010). Although there have been laudable efforts by proponents to deal with these critiques—or "hostile criticisms" as Evans (2018, 137) calls them—those in opposition to DPT remain mostly unconvinced (for overviews, see Mugg, 2016; Grayot, 2020).

Far from providing another addition to this skirmishing, in this paper we aim to provide a new perspective on DPT by drawing on contemporary philosophy of cognitive science debates concerning the legitimacy of *functional individuation*. We take it that Type 1 and Type 2 processes need to be understood as referring to functionally individuated processes—that is, cognitive processes that are individuated based on the role(s) they play rather than on what they are constituted of or realized by. Portraying Type 1 and Type 2 processes as functionally individuated processes is, we think, uncontroversial.[2] As Samuels claims: "[within DPT] psychological mechanisms are, by very widespread consensus, individuated by their functional and computational properties…" (2009, 138). What is surprising, though, is that there has been no systematic assessment of how these properties *are* individuated.[3] As such, we aim to assess whether current defenses of DPT provide convincing rationales for demarcating specifically *two* processing types. We thereby hope to provide a novel and systematic perspective on the evidence needed to establish qualitative differences between Type 1 and Type 2 processes.[4]

Towards this aim, we first provide a brief overview of functionalism in philosophy (Section 2). Inspired by recent work in philosophy of cognitive science (e.g., Weiskopf, 2011a, 2011b), we consider three strategies for functionally individuating Type 1 and Type 2 processes (abstraction, reification, and fictionalization) and assess the prospects for each in relation to DPT's broader cognitive scientific goals. This leads us to the verdict that the most viable route for justifying DPT's core dichotomy is to construe Type 1 and Type 2 processes as reifications (Section 3). We then show that, even under a reificationist interpretation, renewed efforts to demarcate Type

[1] Note that, in this paper, we are not concerned with debates over parallel vs. competitive models of DPT as this depends on whether DPT is cast in terms of task-specificity or task-neutrality. This becomes an overarching theme in the concluding sections of our paper (Sections 4 and 5).

[2] There are two concepts of functional individuation: individuation by causal-functional traits and individuation by normative-functional properties. (see Neander 2017). We here rely on the first.

[3] Samuels (2009) is an exception here. But we plan to go beyond the type/token distinction he draws in his analysis.

[4] Note that in making this argument we do not want to argue for a single-process view but instead simply point out that DPT proponents lack evidence for the dual nature of human cognition. We thank Wim De Neys for pointing out the need to clarify this.

1 and Type 2 processes do not escape existing criticisms of DPT (Section 4). We conclude by further considering how task-specific interpretations of DPT may come into conflict with some of the broader explanatory goals of cognitive and psychological science (Section 5).

2 Making sense of functionalism in DPT

2.1 Dual processing and functional individuation

DPT individuates Type 1 and Type 2 processes via their function. Whether some process counts as 'Type 1' or 'Type 2' depends on its cognitive characteristics, e.g., whether it is fast, automatic, and reactive, or whether it is slow, controlled, and deliberative (among other properties). Moreover, not only are Type 1 and Type 2 processes individuated functionally, but the processing architecture that realizes the interaction of the different modes of processing (serial vs. parallel) is also specified in predominately functional terms. It's perhaps not surprising that proponents of DPT frequently acknowledge their intellectual debt to early defenders of functionalism in philosophy of mind and cognitive science. For instance, Evans remarks that Jerry Fodor anticipated "a dual-process theory that distinguished between input modules (such as those involved in vision and language) and general purpose, central cognition" (2008, 260).

2.2 Beyond Marr's three levels

Reading through the recent literature on DPT, it is apparent that functionalist language is not merely helpful, but *essential* for depicting the qualitative differences (cognitive characteristics) between Type 1 and Type 2 processes. Across the literature, one finds references to 'functional properties', 'functional characteristics', 'functional autonomy', and hordes of other functional assortments (see, e.g., Evans & Frankish, 2009; Stanovich, 2009; De Neys, 2018).

It's easy to see how Marr's (1982) three-fold analysis of the cognitive processing of vision played a prominent role in psychologists' later focus on the functional characteristics of mind and various mental actions. According to Marr, any cognitive system can be analyzed (i.e., explanatorily decomposed) into three distinct levels. The *computational level* identifies worldly relations that act as inputs (i.e., perceptual stimuli and their broader contexts) and outputs (i.e., judgments and behavior). The *algorithmic level* identifies the mechanisms that translate (or transduce) inputs according to rules into a form that readies them for the appropriate output. The *implementational level* identifies the components or broader systems that physically realize the rules according to which said mechanisms transduce worldly input into new output.

Proponents of DPT frequently invoke Marrian language, even using the terminology of 'levels' of analysis, for parsing the functional and computational properties of cognitive types (or systems). For instance, Stanovich states that:

> I argue that it is useful to distinguish the algorithmic level of processing from the reflective level when discussing Type 2 processing. Individual differences in the former reflect the efficiency of the functional cognitive machinery that carries out mental tasks. (2010, 26)

However, although Marr's framework has been useful for situating the ascription of functional properties and characteristics to Type 1 and Type 2, it doesn't *justify* individuating 'mental mechanisms' located at the algorithmic level into two distinct types of processes. This is because Marr's framework is neutral with respect to the internal organization of cognitive processing at the algorithmic level—in other words, it does not support (or deny) that mental mechanisms at the algorithmic level are necessarily dualistic in nature. Proponents of DPT will respond that Marr's framework, while providing a general scaffolding for the functional analysis of cognitive processing, is not what DPT relies on to justify the dual nature of cognition.

Yet, our concern here is precisely that the type of justification needed to establish the dual nature of cognition will ultimately depend on how one approaches functional individuation in detail. This concern is not new. Even former defenders of DPT have begun to express doubt over the possibility of establishing rigorous conceptual foundations for individuating dual processing types. For instance, De Neys claims that:

> [The] debate has not been resolved; it can be questioned whether the debate can be resolved, and even if it were to be resolved, *it will not inform theory development about the critical processing mechanism underlying human thinking*. This implies that the debate is irrelevant for the empirical study of thinking. In a sense, the choice between a single-process model and dual-process model boils—quite literally—down to a choice between two different religions. Scholars can and may have different personal beliefs and preferences as to which model serves their conceptualizing and communicative goals best. However, what they cannot do is claim there are good empirical or theoretical scientific arguments to favor one over the other (2021, X, emphasis added).

While we empathize with De Neys, we are less pessimistic about resolving debates over DPT. In part, this is because we think there are unexplored avenues open to us. What's missing is a systematic analysis of *how* functional processes are individuated. To fill this lacuna, we now consider a taxonomy of different strategies for functional individuation that has its roots in the philosophy of cognitive science (e.g., Weiskopf, 2011a, 2011b).[5] The taxonomy is as follows:

> **Abstraction** occurs "when we decompose a modeled system into subsystems and other components on the basis of what they do, rather than their correspondence with organizations and groupings in the target system" (Weiskopf, 2011a, 329).

[5] Weiskopf develops this taxonomy in an attempt to avoid the questionable 'boxology' seen throughout philosophy of mind and cognitive science with regard to depicting functional individuation.

Reification is the "act of positing something with the characteristics of a more or less stable and enduring object, where in fact no such thing exists" (Weiskopf, 2011a, 328).
Fictionalization is the process of "putting components into a model that are known not to correspond to any element of the modeled system, but which serve an essential role in getting the model to operate correctly" (Weiskopf, 2011a, 331).

This taxonomy allows us to further concretize what kind of functional processes Type 1 and Type 2 are meant to be, i.e., abstractions, reifications, or fictions. To see why this is relevant consider that a distinction between Type 1 and Type 2 processes *could easily* be maintained under the abstraction interpretation. If one could establish that Type 1 and Type 2 are abstractions, then the dichotomy could be supported based on the fact that it reflects neural-physiological mechanisms and networks at the implementational level. Yet, we will argue that the two types cannot be plausibly construed as abstractions. On top of this we argue that fictionalization, while being very close to reification, will be the less favored option by proponents of DPT. Hence, we will make the case that reification is the most tenable individuation strategy for supporters of DPT. Yet, we shall argue that, if we construe Type 1 and Type 2 as reifications, the common rationales for the functional dichotomy at the core of DPT remain unsatisfactory.

3 Three strategies for individuating Type 1 and Type 2 processes

In this section, we will ask what it would mean for DPT to individuate Type 1 and Type 2 processes via the strategies of *abstraction*, *reification*, and *fictionalization* and critically analyze each strategy.

3.1 Three functionalist interpretations of DPT

3.1.1 Individuation via abstraction

To say that DPT individuates functions via abstraction would be to say that it decomposes a modeled system on the basis of what it does, rather than on how its component parts are actually structured. While the process of abstraction might distort to some degree the underlying mechanics of the systems under investigation, they are, nevertheless, intended to capture the "essential operations" of those systems given the descriptions of their components (Weiskopf, 2011a, 329). It is for precisely this reason that many philosophers of cognitive science have come to interpret abstractions as 'mechanism sketches' (Buckner, 2015). In picking out the essential operations of a system, a functional theory or model provides a mechanism sketch if it identifies the primary mechanisms responsible for the system's operation, and thereby, provides an explanation for how those mechanisms cause or give rise to phenomena of interest.

While few in the philosophical literature cast DPT in terms of mechanism sketches, it's clear that at least some supporters take DPT to indicate the existence of distinct neural mechanisms governing separate cognitive processing types. There are three separate avenues one might take to establish this existence: (i) via evidence of *automatic and/or autonomous processing mechanisms*, (ii) via evidence of *executive functioning and conscious control mechanisms*, and/or (iii) via evidence of *discrete competing cognitive architectures*. Here are brief examples of each:

Concerning (i), new research on what is known as the *default mode network* (DMN) supports the view that autonomous and/or automatic processing systems exist in the brain (Buckner et al., 2008). The DMN–which includes the *medial prefrontal cortex*, the *posterior cingulate cortex*, the *inferior parietal lobule*, the *lateral temporal cortex*, the *dorsal medial prefrontal cortex*, and the *hippocampal formation*–is theorized to play a central role in the organization and expression of preplanned, reflexive behaviors that are characteristic of fast and reactive thinking (Raichle, 2015). Further, research by Vatansever et al. (2017) suggests that the DMN might be thought of as an 'autopilot mode' of brain functioning, which is specifically active during automatic behaviors (i.e., during rapid selection of a response to a predictable context) but not during controlled decision- making, which involves other neural mechanisms.

Concerning (ii), growing research on the computational (O'Reilly et al., 2010) and implementational foundations of selective attention, memory, language-use, and action-orientation related faculties (Botvinick & Braver, 2015) separately supports the existence of *executive functioning* and *conscious control* (EF/CC) networks in the brain. More specifically, while the regions thought to be involved in executive function and cognitive control are proving to be wide-ranging and diverse, there is growing consensus that both cortical and subcortical structures appear to work together to implement cognitive control in support of functions that are generally thought to be demonstrative of higher thinking and reasoning. These regions include the *lateral prefrontal cortex*, the *dorsal anterior cingulate* and *presupplementary cortices*, *dorsal premotor cortex*, the *anterior insula*, and the *intraparietal cortex* (Power & Petersen, 2013) among others.

Concerning (iii), some researchers have interpreted what appears to be an inverse relationship in activity patterns between the DMN and EF/CC networks as evidence that the brain has a fundamentally dual cognitive structure (Lieberman, 2003). Accordingly, it is speculated that functionally separate cognitive networks literally interact, i.e., compete, at the implementational level in order for activation to occur in the respective mechanisms (Goel, 2008). Many supporters of DPT take this inverse relationship in neural activity as (partial) evidence that the brain does implement something akin to a default-interventionist model of cognitive architecture (Evans, 2008; Evans & Stanovich, 2013a; Stanovich, 2009).

3.1.2 Individuation via reification

To say that DPT explains via appealing to reification is to say that it offers descriptions of phenomena whose actual component parts or structures may be quite different from the modeled system's. That is, unlike abstraction, reification does not

sketch discrete mechanisms. Instead, it seeks to describe a system's essential operations in terms of functions that transcend or blur the distinction between mechanisms, thereby positing something with the characteristics of a more or less stable and enduring object where, in mechanistic terms, there is none (Weiskopf, 2011a, 338).

Those supporters of DPT who eschew a (neural) mechanistic interpretation may have in mind something like reification when they defend the functional roles of Type 1 and Type 2 processing. Consider Evans' claim:

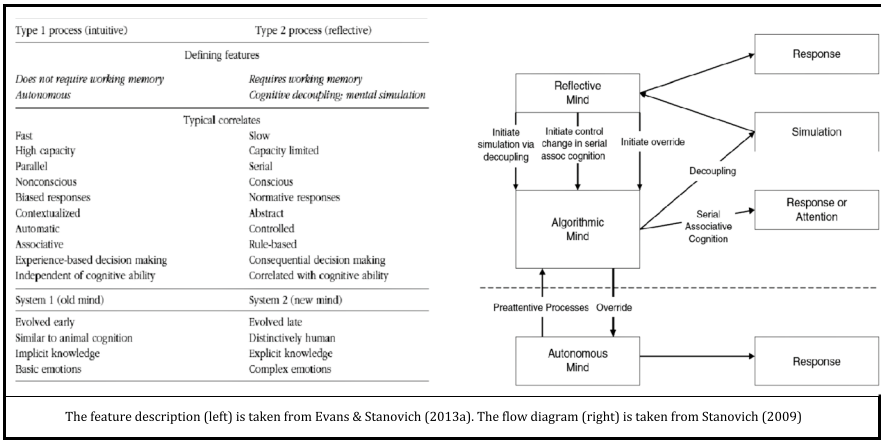
> Close inspection of the evidence suggests that generic dual-system theory is currently oversimplified and misleading. In particular, (a) it is not possible coherently to link together all the attributes associated with Systems 1 and 2, respectively…, and (b) there are at least two quite distinct forms of dual-process theory to be found in these various literatures that cannot readily be mapped on to each other. We might be better off talking about type 1 and type 2 processes since all theories seem to contrast fast, automatic, or unconscious processes with those that are slow, effortful, and conscious… (Evans, 2008, 270)

This passage suggests that the fundamental distinction between Type 1 and Type 2 processing can be maintained by positing something with the characteristics of a more or less stable and enduring object when, in fact, no such thing exists at the level of token mechanisms. More specifically, the view that DPT is a theory of cognitive *types*–as opposed to *systems*–embodies a form of reification known as *fusional* reification (Buckner, 2015, 3931). This happens when a theory or model introduces an object or operation whose defining properties and characteristics are realized by a diverse set of discrete components or resources distributed throughout a system. The reified object or operation is a fusion of component properties and characteristics.

A question for the reificationist interpretation of DPT is this: what reifies Type 1 and Type 2 – i.e., what allows us to treat them as real and distinct? One move proponents of DPT can make is to view Type 1 and Type 2 processing as *descriptions* of the algorithms carried out by mechanisms and networks at the implementational level (Evans, 2008; Stanovich, 2011). On this view, DPT individuates Type 1 and Type 2 processes by articulating (i) the problems, i.e., cognitive tasks, that they are developed (naturally or socially) to solve and (ii) the rules according to which those processes are carried out.

Of course, since such algorithms are articulated at the level of theory, reification principally involves appeals to two things: *feature descriptions*, which inventory and correlate the attributes of contrasting processing types, and *flow diagrams*, which portray the input–output relations among different processing levels as key operations. See Fig. 1 for examples of each both below:

Feature descriptions and flow diagrams like those pictured above are critical for theorists to reify Type 1 and Type 2 processes as real and distinct phenomena as such tools render the core concepts deployable for scientific theorizing. In other words, feature descriptions and flow diagrams are what bring Type 1 and Type 2 into being. For instance, one may consult a feature description chart when devising psychological experiments and, more importantly, when interpreting and analyzing behavioral



Type 1 process (intuitive)	Type 2 process (reflective)
Defining features	
Does not require working memory	*Requires working memory*
Autonomous	*Cognitive decoupling; mental simulation*
Typical correlates	
Fast	Slow
High capacity	Capacity limited
Parallel	Serial
Nonconscious	Conscious
Biased responses	Normative responses
Contextualized	Abstract
Automatic	Controlled
Associative	Rule-based
Experience-based decision making	Consequential decision making
Independent of cognitive ability	Correlated with cognitive ability
System 1 (old mind)	System 2 (new mind)
Evolved early	Evolved late
Similar to animal cognition	Distinctively human
Implicit knowledge	Explicit knowledge
Basic emotions	Complex emotions

The feature description (left) is taken from Evans & Stanovich (2013a). The flow diagram (right) is taken from Stanovich (2009)

Fig. 1 The feature description (left) is taken from Evans and Stanovich (2013a). The flow diagram (right) is taken from Stanovich (2009)

results as it provides a ready-made schema for interpreting evidence. Similarly, one may also consult a flow diagram to make sense of why a particular stimulus invoked a 'fast' rather than 'slow' response to some reasoning or decision task. Insofar as these scientific tools aid in grounding theory in behavioral evidence, reification offers a plausible justification for individuating Type 1 and Type 2 processes.

3.1.3 Individuation via fictionalization

To say that DPT individuates via appealing to fictions is to say that it offers characterizations that are known not to correspond to any element of the modeled system. Kahneman, famously, takes this view to the extreme, stating that:

> System 1 and System 2 are… fictitious characters. [They] are not systems in the standard sense of entities with interacting aspects or parts. And there is no one part of the brain that either of the systems would call home. (2011, 31, see also Kahneman & Frederick, 2005)

In fact, Kahneman goes so far as to claim that the properties and functions of Systems 1 and System 2 are ultimately "intended as description, not an explanation" (2011, 31). This illustrates the difference between individuation via *reification* and via *fictionalization*: whereas reification can at least make some claims to a theory-independent reality (we might call it genuine ontological status), individuation via fictionalization is contingent on its use-value in a given theory or model.[6]

[6] While we are relying on Weiskopf's notion of reification, the distinction between fiction and reification remains controversial in philosophy of cognitive science and philosophical psychology (see, e.g., Buckner, 2015; Zahnoun, 2020; Beck & Grayot, 2021).

3.2 Applying the individuation strategies

Having summarized three possible strategies for individuating Type 1 and Type 2 processes, we now consider whether, and if so, how each justifies the functional foundations of DPT.

3.2.1 Why abstraction doesn't work as an individuation strategy

Although proponents of DPT, like Evans, Stanovich, and Kahneman, maintain that DPT does not pick out discrete mechanisms in the brain, they don't shy away from appealing to neural evidence when it suits their purposes (Evans, 2008, 2010; Kahneman, 2011; Stanovich, 2009). This builds support for the idea that Type 1 and Type 2 processes are actually grounded in, even isomorphic with, neural- physiological processes at the implementational level. This is clearly an attractive rhetorical move for justifying the individuation of Type 1 and Type 2 processes as it would provide theorists with strong empirical support for the functional dichotomy at the core of DPT. So, the question is: is it legitimate to think of Type 1 and Type 2 processes as mechanism sketches and, thereby, as individuated via abstraction?

Recall (3.1.1), that there are several candidate networks associated with both Type 1 and Type 2 processes: Type 1 processes are correlated with activity in the default mode network (DMN) whereas Type 2 processes are correlated with activity in the executive functioning and cognitive control networks (EF/CC). As such, appealing to either network could be taken as support that DPT individuates the functions of Type 1 and Type 2 processes via the strategy of abstraction.

The issue we want to raise here is not that these different brain networks aren't differentially correlated with what we might be loosely labeled as 'intuitive' and 'reflective' types of information processing; yet, we maintain that such correlations aren't strong enough to underscore *bona fide* mechanism sketches, and so, these correlations don't independently justify individuating Type 1 and Type 2 processes on the basis of abstraction. Here's why:

For a theory or model to count as a mechanism sketch, it needs not only to capture the essential operations of the hypothesized mechanisms, but it needs to provide a plausible account of how the system is actually structured at the component level (even construed as a 'how-possibly' model of cognitive functioning, it should reliably parse the relevant mechanisms for some cognitive task). Even granting that mechanism sketches are elliptical and incomplete with respect to the target system they model (Piccinini & Craver, 2011), mechanism sketches are built upon the expectation that the modeled system in question can be, at least in principle, decomposed into discrete mechanisms.

So, why don't Type 1 and Type 2 processes count as mechanism sketches? The reason is that the framework of DPT, according to its most ardent proponents, isn't designed to identify the mechanisms responsible for the essential operations of Type 1 or Type 2 processing. In saying, for instance, that Type 1 processes are strongly correlated with activity in the DMN *(i.e., medial frontoparietal network*) what theorists are really saying is that there appears to be a correspondence between this network and a set of attributes that are associated with a particular

cognitive profile, e.g., Type 1 attributes. But notice that the identification between activity in the DMN and Type 1 depends on assimilating two levels of function-based inference: Not only must it be (first) established that Type 1 attributes are solely governed by the DMN, but it must also be (second) established that the DMN is *not* involved in any other cognitive operations that aren't associated with Type 2 processing, such as those which consume working-memory resources. Hence, to establish that the dichotomy upon which DPT is based is an abstraction, it would need to be demonstrated that the subregions of the brain that make up the DMN have no (statistically relevant) involvement in Type 2 processes. The problem here is that not only are the same subregions that comprise the DMN also revealed in studies to be involved in Type 2 processing, like goal-based reasoning and conceptual reflection (Botvinick & Braver, 2015), but it is, at present, empirically underdetermined to what degree Type 1 processes are dependent on other brain regions not typically associated with autonomous or automatic cognition (Murphy et al., 2018).

To make matters more complicated, it is equally difficult to say whether the various regions that make up the EF/CC networks (supporting Type 2 processing) can be neatly ascribed to individually tractable mechanisms or sets of mechanisms. This is because the EF/CC networks span multiple subregions of the *anterior cingulate cortex*, *lateral prefrontal cortex*, and *presupplementary cortices* (Botvinick & Braver, 2015), implicating very diverse mechanisms (many of which are still being investigated). The fact that different subregions are recruited for quite different sorts of cognitive tasks typically associated with Type 2 processing indicates that much is still unknown about the neural-physiological details of the EF/CC networks.

In saying that DPT is not plausibly viewed as offering mechanism sketches, we're not claiming that the underlying networks, i.e., DMN and EF/CC, can't be understood mechanistically. What we're saying is that Type 1 and Type 2 processes, construed as cognitive types, shouldn't be identified with token mechanisms or neural networks (cf. Samuels, 2009).

Further, because it is an open question whether or not Type 1 and Type 2 processes can be readily identified with discrete neural mechanisms or networks, the story of their interaction is also mired in conceptual and ontological ambiguity. This raises additional problems for individuation via abstraction. Recall that the majority of proponents of DPT adopt either a default-interventionist model of DPT or a parallel-competitive model of DPT. Yet, on-going analyses and replication studies indicate that neither model is singularly equipped to predict and explain how individuals engage in higher reasoning tasks (Thompson & Newman, 2018). We suspect this stems from the difficulty of correlating Type 1 and Type 2 processes with discrete neural mechanisms or networks.[7]

[7] To emphasize this point further, De Neys (2023) has argued that even if we were able to establish reasonable anatomical correlations between Types 1 and Type 2 processes and specific neural mechanisms, this would not resolve the explanatory problem of whether a behavioral solution to some reasoning/logic problem necessarily depends on switching between processing types. This, to parrot classic arguments by Keren & Schull (2013) violates the exclusivity of functions with their anatomic parts.

3.2.2 On the possibility of fictionalization as an individuation strategy

Because abstraction seems not to be a viable strategy for individuating the functional properties and operations of Type 1 and Type 2 processes, supporters of DPT are left with either reification or fictionalization as individuation strategies. Let's consider the latter first.

Prima facie, there are no problems with saying that DPT individuates via fictionalization. We suspect that most supporters of DPT will not take issue with the former claim insofar as the term 'type' is seen merely as a label for categorizing properties or operations. As mentioned above, some of the most vocal proponents of DPT have admitted that such terminology may indeed be misleading, giving the impression that the properties of these categories are naturally fixed (which they clearly aren't—Evans, 2008; Evans & Stanovich, 2013a). But this then suggests that at least part of DPT's theoretical appeal comes from the use-value that such categories afford for researchers looking to systematize diverse behavioral phenomena.

Yet, while proponents may be willing to bite the bullet when it comes to the claim that types are useful fictions, it's far less likely that they are willing to admit that the *token* cognitive processes underlying Type 1 and Type 2 processing are fictions (cf. Samuels, 2009). After all, *something* is taking place, at the algorithmic level when humans engage in higher reasoning tasks, even if that something can't be neatly explained in terms of the operations of discrete neural mechanisms. Hence, seeing token cognitive processes as fictions would raise serious questions for the theoretical validity of DPT let alone its predictive and explanatory power. This indicates that proponents have only one choice when it comes to individuating Type 1 and Type 2 processes and that's to see them as reifications. However, as we argue in the next section, under the reification interpretation DPT still lacks a clear rationale for individuating Type 1 and Type 2 processes.

4 Type 1 and Type processes as reifications

The final individuation strategy to be discussed is reification. In contrast to abstraction, under which neural-physiological mechanisms and networks *could* provide strong empirical support for there being two distinctive processing types, establishing support for the reification interpretation is more challenging. This is because, under reification, the relevant functional properties and algorithms are not intended to describe the properties and operations of mechanisms at the implementational level. Therefore, the same underlying system could, in principle, be reified into one, two, or any number of processing types depending on one's explanatory goals. Consequently, in this section we examine the most plausible justifications for individuating Type 1 and Type 2 processes via reification.

First, we look at the *conceptual criteria* that DPT theorists have so far used to establish the distinction between distinct processing types. We then argue that these standard conceptual criteria do not provide a clear dichotomy, i.e., qualitative discreteness and conceptual opposition, that would license the reification into two processing types. We then turn to a more recent proposal by Dewey (2023) who argues

that methodological analyses of single- versus dual-process reasoning tasks point toward *metacognitive control* as the most plausible candidate for establishing qualitative discreteness and opposition. In relation to this very innovative proposal, we argue that it relies on highly contentious assumptions that make it doubtful that it allows us to go beyond the limitations of the more traditional proposals.

4.1 Automaticity, autonomy, and working memory

One of the main discussions within the DPT framework is how to distinguish Type 1 from Type 2 processes based on *internal* features. This is also referred to as the "mapping problem" (Moors, 2014) or the problem of "feature alignment" (De Neys, 2021; Evans & Stanovich, 2013a). DPT has generally relied on the feature descriptions for making qualitative distinctions between processing types. However, the contemporary dual-process literature no longer takes all of the features described in Fig. 1 as *necessarily* correlating; instead, Type 1 and Type 2 possess at least one defining feature that stands in a dichotomous relationship with the defining feature of its counterpart.

The original qualitative distinction was based on the belief that the notion of 'automaticity' would capture the features of cognitive processing that were autonomous, fast, required minimal or no attention, efficient, and otherwise uncontrolled. However, this distinction proved untenable on the basis of the vagueness of the concept of automaticity: either the underlying features failed to correlate systematically or did not conceptually support a clear qualitative distinction between Type 1 and Type 2 processes (Moors & de Houwer, 2006). For instance, experiments involving the Stroop task, where participants must name the print color of a written word, or vice versa, showed that participants could implicitly adapt their performance based on the number of congruent or incongruent trials (Blaise et al., 2012). Conscious awareness of the number of congruent or incongruent trials did not seem to have a significant effect, leaving open the possibility that people *could* exercise implicit control, violating automaticity as a conceptual criterion.

As such, Evans & Stanovich and others switched to categorizing Type 1 processes as essentially *autonomous*, which often—but not always—correlate with the other features of automaticity. Bargh (1992) conceptualized autonomy as a minimal criterion for automaticity: A process is autonomous if, once started, the process runs to completion without any conscious guidance or monitoring. Whenever a respective stimulus is perceived, the corresponding process will be triggered regardless of a person's intention. Thus, the distinction of the two qualitatively different sets of processes is typically made accordingly: Type 1 processes are autonomous, and therefore usually faster; Type 2 processes, which involve cognitive decoupling and mental simulation are much more demanding of working memory resources, and therefore tend to be slower (Thompson & Newman, 2018). Hence, the relevant features do not need to be perfectly aligned (indicating necessary co-occurrence) but can merely correlate to the assigned type of mental process. In other words, proponents of DPT need only appeal to a single feature, namely autonomy, to differentiate Type 1 from Type 2 processing.

However, De Neys (2021) points out that autonomy is not a universal feature of Type 1 processing but is rather highly context dependent (see also Keren, 2013). Whether a stimulus leads to a specific autonomous response is dependent on the current task a person is engaged in or the 'goal' context. Encountering a stimulus is not enough, as the stimulus must be attended to and considered relevant. Stimuli that are not relevant to the task a person is currently engaged in are ignored.

Cognitive decoupling and mental simulation serve as the counterpart to autonomy. Cognitive decoupling is the ability to differentiate stimulus-bound representations from those which are imagined or recalled, as when thinking or reasoning hypothetically or counterfactually (Evans & Stanovich, 2013a). However, the features of cognitive decoupling and mental simulation suffer, like autonomy, from vague definitions. According to De Neys, "this theorizing has not clarified how one can operationally measure whether a mental simulation process has occurred. The available empirical evidence focuses on measuring correlated features such as the effortful, cognitively demanding nature of mental simulation" (2021, 1417–1418). Therefore, mental simulation processes cannot function as a necessary criterion for Type 2 mental processing.

There are two options for a DPT theorist to deal with these problems. The first option is to find a set of operational criteria that determines whether a (functionally individuated) process exhibits the qualitative feature to a *sufficient degree*. However, empirically establishing these threshold values between the supposed qualitative differences is extremely difficult (De Neys, 2021; Keren, 2013). The other option is to link one of the processes to the operation of another mental entity or process. Because of the vagueness of the proposed essential features of Type 1 and Type 2 processes, the first option is inevitably dependent on the second. For instance, the establishment of mental simulation is done via correlated features of the process being *effortful*. In this regard, Evans and Stanovich (2013a) propose that high demands on working memory is a defining feature of Type 2 processing. The general idea here is that Type 1 processes, being autonomous, do not require controlled attention and therefore make minimal demands on working memory resources (Evans & Stanovich, 2013a; Stanovich, 2011). In turn, Type 2 processes are dependent on the engagement of working memory.

However, appealing to working memory as a criterion is controversial as there are different views on the nature and characteristics of working memory (De Neys, 2021; Gomez-Lavin, 2021). Furthermore, controlled attention, the process underlying working memory, also seems to be involved in many autonomous processes (Barrett et al., 2004). Goal-directed attention seems often to be a prerequisite for more automatic forms of attention as processes that are considered 'controlled' can operate non-consciously. (This also has negative implications for establishing what counts as meta-cognitive control, as we'll see below.) This means that DPT needs to determine a threshold for the qualitative distinction between processing types. Moreover, proponents of DPT would then need to

argue why *this* particular threshold justifies viewing Type 1 and Type 2 as distinct types (De Neys, 2021).[8]

So how does this relate to the issue of reification? From a conceptual standpoint, the above analysis shows that DPT lacks an unambiguous rationale for individuating exactly two types of processes. Neither autonomy nor working memory provide us with a justification for treating Type 1 and Type 2 as qualitatively discrete and conceptually opposed.

4.2 Metacognitive control

One way to respond to the above conceptual challenges is to point toward the rich history of behavioral research into dual processing (Evans, 2018; Evans & Stanovich, 2013b). When all else fails, what does the evidence tell us? Much ink has been spilled over how to interpret existing research, with many skeptics remaining unconvinced by 'supporting' evidence (De Neys, 2021; Gigerenzer & Regier, 1996; Keren & Schul, 2009; Mugg, 2013, 2016; Osman, 2004). However, we will not get into these debates. Instead, we want to consider *what it would take* for some data to count as evidence in support of Type 1 and Type 2 as discrete reasoning types.[9] Therefore, we now turn to Dewey (2023), who provides a remarkably original take on reasoning task behaviors and their analysis across dual-process and rivaling frameworks.

According to Dewey, some dual-process theorists and their opponents agree that reasoning task evidence is best depicted via non-linear, monotonic models, they disagree as to whether monotonicity is indicative of a single, continuous reasoning process or two qualitatively distinct reasoning processes. Dewey's concern is with the interpretation of monotonic analysis; her goal is to show that, whereas monotonicity can tell us something about whether two sets of behavioral responses were created by a single or multiple types of cognitive processes, "it can't tell us whether those processes *have the defining properties* of the single or dual types of reasoning posited by [single process theories] and DPT" (2023, 2). This is in response to Stephens et al. (2018) and Stephens et al. (2019) who use *signal detection theory* and *state-trace analysis*, among other methodologies, to argue that the presence of monotonicity in behavioral experiments is better explained—or rather, better accommodated—by single process models of reasoning.[10]

[8] Although it is well-understood that deliberative processes tend to require more attentional resources, working memory and controlled attention seem to operate on a continuum. In other words, all cognitive processing requires some working memory or attentional resources (Thompson & Newman, 2018).

[9] An additional reason is that we are not necessarily interested in defending single-process models as explanatory alternatives in this paper.

[10] Dewey (2023, 2-13) explains in great detail how exactly state-trace analysis and signal detection theory are used (or mis-used) for monotonic analysis of single-process and dual-process theoretic evidence. Signal detection theory is commonly used to differentiate between information-bearing or stimulus-bound behaviors from random or noisy behavioral patterns in uncertain decision-making. State-trace analysis refers to the method for determining the number of underlying parameters or latent variables that are varying across two or more tasks.

Dewey's chief complaint with Stephens' analyses is that they lack a "necessary relation" between the "dissociating conditions in reasoning task experiments" that *should* distinguish Type 1 from Type 2 processing. That is to say, attempts by opponents of DPT to replicate the latent variables of reasoning tasks do not adequately capture the qualitative opposition that is defining of dual-process theory. Dewey concludes that such caveats undermine any negative claims concerning the falsity of DPT. In other words, DPT could be *true* even if the evidence fits models assuming only a single process; and conversely, DPT could be *false* even if the evidence fits dual-process models (Dewey, 2023, 14).

This is a crucial insight for the debate over single-versus dual-process modeling as it suggests that the degree to which either theory is supported by behavioral evidence depends upon (a) how—and how many—latent variables are posited, (b) how those variables are formally modeled and analyzed, and (c) how the data is re-interpreted with respect to the fit between theory and model.

However, unlike De Neys (2021), who maintains that "there is currently no good evidence that would allow one to decide between the single-process and dual-process model view" (1422), Dewey advocates for a new criterion distinguishing Type 1 and Type 2 processes, one that would avoid the evidential ambiguity seen above—this criterion is referred to as *metacognitive control*. Metacognitive control is described as the operation by which individuals switch from 'default reasoning' (i.e., Type 1 processing) to 'intervening reasoning' (i.e., Type 2 processing).

However, we hold that Dewey's argument for metacognitive control would not add anything new to the existing conceptual distinctions we've argued against thus far if her aim were not to provide a strategy for creating task-specific models and developing task-specific analyses for testing those models.[11] So, in the remaining part of the section we will explain why this is the case and, thereby, motivate the importance of understanding Dewey's proposal in a task-specific way. More precisely, we think there are two dimensions along which metacognitive control does not resolve debates over the interpretation of behavioral evidence unless we operationalize it in task-specific ways. The first dimension concerns how we are to distinguish 'controlled' from 'non-controlled' action; the second concerns what counts as 'metacognition' in principle. The next section will then focus on the issues that arise under Dewey's task-specific proposals.

In Section 4.1 we reviewed a few reasons why the conceptual distinction between automatic and controlled behavior has been a historically problematic. There are additional reasons to be wary of attempts to draw a principled distinction between controlled and non-controlled action. Recent research on agentive control in the philosophy of action and philosophical psychology reveals growing rifts on the nature and role that self-control plays in human agency, with special focus on factors such as *perception* and *attention* (Wu, 2014), *emotion* (Scarantino & De Sousa, 2018), and *intentionality* (Pacherie & Haggard, 2010), among many other factors. This has led not only to highly sophisticated accounts of *agentive control* which defy strict dichotomies like those assumed by DPT, but it indicates that control is *not* a

[11] We thank an anonymous reviewer for highlighting that this in indeed a fair construal of Dewey's aims.

one-size-fits all concept: control comes in degrees, may be variable with respect to when and how it is manifested (or frustrated) for the same individual, and may differ in how it is studied and defined between first-person and third-person accounts (Shepherd, 2014). Furthermore, as we discussed in detail Section 3.1, the mechanisms responsible for automatic versus controlled behavior are not located in a single neural module or network. Both the default mode network and executive control networks have been repeatedly shown to play active roles in both automatic and controlled behaviors simultaneously in a range of neural imagining studies.

This brings us to the theory of metacognition. Dewey's understanding of metacognitive control follows Ackerman and Thompson's (2017) account of 'meta-reasoning': their account suggests that individuals utilize metacognitive 'monitoring' and 'heuristics' functions to recognize the "rightness" of judgments and subsequent actions (Dewey, 2023, 16). However, one worry we have with this account is that it does not specify what makes a cognitive achievement a 'meta' cognitive achievement. Dewey offers no explanation for this, whereas Ackerman & Thompson only offer hints of what this may involve—e.g., they state that: "[M]etacognitive processes are assumed to have a control function over the initiation or cessation of mental effort. In other words, if we are confident in our answer, we will act on it" (2017, 607). This doesn't get us very far in understanding what metacognition is in relation to agential control. It can be read in one of two ways: either that metacognitive processes are those we are in control of, or that metacognitive processes are those that we are confident in. The former option is not helpful as it presupposes agential control in determining when 'meta-reasoning' has occurred, but this merely begs the question. The latter option, by contrast, is more helpful, but it too runs into problems.

Presumably, appealing to one's confidence in their judgments and reasoning implies that metacognition must involve a high degree of self-awareness or introspective access. One cannot be sure of their judgments if they cannot self-reflect. But again, recent work in philosophical psychology indicates that self-awareness and introspective access may not be the best ways to distinguish control from non-control. There are two simple reasons for this. First, many controlled actions are not performed with full awareness (for examples, see Shepherd, 2014). Second, just because we have first-person self-awareness over some act does not mean we are in control of it. Much of the research on confabulation and self-deception in both philosophy of action and cognitive psychology indicate that often we reason backward, justifying our actions *as if* we selected them (McGeer, 1996; Metcalf et al., 2007). So, appealing to confidence or self-assuredness as a means for grounding metacognition in agentive control also runs into unresolved conceptual problems.

In sum, the criterion of metacognitive control relies on two highly contentious assumptions: (i) that we can reliably (conceptually and empirically) distinguish controlled from non-controlled action, and (ii) that metacognition is an intrinsic and well understood component of agentive control.

Yet, here is where the contextual nature of Dewey's methodological analysis becomes important. In order to avoid the criticisms mentioned above metacognitive control needs to be operationalized at a *task-specific* level. We agree that a shift to a task-specific analysis of Type 1 and Type 2 processing may allow one to circumvent

some of the classical problems associated with DPT (such as the mapping problem). This would mean that the precise operationalization of a demarcation criterion should be based on features of individual reasoning tasks. We take this to be an important position that is neatly exemplified by Dewey. However, in the remainder of our paper, we will argue that going task-specific appears less promising than proponents of DPT may have hoped.

5 DPT and the two-mind trope: implications for psychological science

Above we argued that Dewey's proposal highlights the possibility that DPT *could be* defended on a task-specific basis that allows us to meet the challenges of qualitative discreteness and conceptual opposition. However, in this section we want to highlight some of the tradeoffs and problems one encounters when going down this route.

5.1 From task-specificity to task-neutrality

As our analysis of the reificationist interpretation of Type 1 and Type 2 shows, debates over dual-process theories of reasoning-task behavior have become further complicated by the question of whether cognitive processing types ought to be defined task-specifically or task-neutrally. Dewey's (2023, 199) strategy builds on the assumption that identifying defining Type 1 and Type 2 properties is done by determining the task-specific condition that can isolate those properties. Modelers must use formal monotonic analysis to determine if a two-process model explains response rates for a specific task best. In turn, we can generalize across experimental contexts to determine general types of reasoning (Dewey, 2023, 206). Such a generalization is possible because Dewey assumes that the relationship between task-specific models and a general task-neutral DPT is a relationship of operationalization. However, De Neys (2021, 2023), has argued that debates on task-specific or task-neutral definitions have not only grown stagnant, but that mounting differences in reasoning behaviors at the task-specific level frustrate attempts to draw any meaningful conclusions with respect to how many cognitive types or systems there are.

Moreover, although task-specific operationalized definitions of Type 1 and Type 2 processes might serve to circumvent some of the problems of DPT, it also introduces new ones. For instance, it shifts the burden of formulating task-neutral features to formulating valid comparisons *between* the different task-specific features. That is, once we commit to only task-specific criteria, we can no longer know whether or how task-specific features of experiments co-refer, and, in this way, something like the mapping problem re-emerges at the task-specific level. Of course, this does not mean that DPT must necessarily be defined through entirely task-neutral features; but it does raise the question as to how important task-neutral ontological criteria are for defining (or refining) the characteristics used to assess task-specific accounts of Type 1 and Type 2. This question *should* matter to psychologists as, without some generalizing features, it would seem to be prohibitively difficult to compare models, extrapolate data, and

ultimately draw meaningful conclusions about differences between apparently similar reasoning tasks. Of course, in principle, psychologists might simply have different epistemic aims (implying different scientific values in practice) and might give up on all generalizations or even between task-comparisons, but then they need to be explicit about that. In terms of Dewey's operationalization strategy, without criteria for co-reference, there cannot be any valid cross-experimental generalization, which prevents the move "from modeling back to theory" (Dewey, 2023, 206).

5.2 DPT as *meta*-theory

To address the worries about finding a task-neutral operational definition of Type 1 and 2 processing, or their interactions, proponents of DPT have alternatively argued that DPT is better thought of as an encompassing meta-theory. But what is a meta-theory? Evans & Stanovich offer the following:

> [T]he word theory is applied in a most ambiguous way in psychology. It can mean anything from a broad set of proposals of a particular kind, more accurately termed a metatheory, to a specific model of task level behavior. Metatheories are not directly testable, but they are interesting and important as the history of psychology shows beyond doubt. However, when critics refer to the dual-process theory at the program level, they somehow expect it to have the directly testable properties of a task-level account. (Evans & Stanovich, 2013b, 268)

According to Evans & Stanovich, a meta-theory is apparently a productive piece of metaphysics that allows us to organize and structure our research programs (cf. Lakatos, 1978). Although no universal operational definition of both types can be formulated, they can be specified and tested in domain- or task-specific environments (see, recently, Tinghög et al., 2023). Hence, it could be argued that the meta-theory interpretation sits quite well with the task-specific defense we analyzed above.

We find additional problems with this interpretation. One might hold that the meta-theory is what rectifies some of the worries about the epistemic aims that we have to give up on when pursuing entirely task-specific formulations of DPT. The idea is that, despite differences in individual reasoning-task behaviors, there seems to be a unifying dualistic structure to all of the underlying processes (or at least, *descriptions* of such processes lend themselves to such an approximation). This seems to be the minimal commitment of the meta-theory interpretation. Yet, this commitment is susceptible to the very criticisms we've raised in this paper, namely, that there is no clear rationale for distinguishing Type 1 and Type 2 as reifications. Cast in the language of Lakatosian research programs, this constitutes a major flaw in the *hardcore* of the DPT framework.

We take it that Evans & Stanovich will not be convinced by this as they claim: "The theory that critics are attacking is in fact a construction: an abstraction of salient features from many different dual-processing proposals, that we term the *received* view" (2013b, 263). It is, however, hard to see how criticisms of what they label the 'received view' is relevantly different from what they call the meta-theory.

The point here seems to be that meta-theories need to be productive as *organizing principles*. What would it take, then, to show that an organizing principle is not productive? Either we show that every single model fails to predict or explain what it claims to predict or explain—this is clearly an unrealistic task. Or, as we do, argue that we lack a convincing rationale for demarcating Type 1 and Type 2 processes that generalizes across applications of the framework.

6 Conclusion

We have argued that individuation via reification is the only sensible option for DPT. Yet, a closer look at how dual-process theorists have attempted to justify reifying exactly two types of processes reveals the lack of a clear rationale for doing so. In this regard, we have argued that criteria for individuating the two types rely on highly contentious assumptions. Moreover, we also pointed out that tasks-specific operationalizations do not provide a neat solution that would salvage these criteria. Our conclusion is that if these criteria fail (or remain contentious), then construing Type 1 and Type 2 as reifications cannot deliver the vital functional foundations that DPT requires. Although the perspective defended in this paper does problematize DPT in its current state, it also points a way forward in an entrenched debate. Proponents of DPT need to formulate a strategy for explicating the functionalist foundations of DPT in terms of reification.

Author's contribution As stated in the submission portal.

Funding Open access funding provided by FCT|FCCN (b-on). Funded by the Fundacao para a Ciencia e a Tecnologia, # 2021.03201.CEECIND.

Data availability Not applicable.

Declarations

Ethical approval Not applicable.

Informed consent Not applicable.

Competing interests Not applicable.

Research involving human participants and/or animals Not applicable

References

Ackerman, R., & Thompson, V. A. (2017). Meta-Reasoning: Monitoring and Control of Thinking and Reasoning. *Trends in Cognitive Sciences,21*(8), 607–617. https://doi.org/10.1016/j.tics.2017.05.004

Bargh, J. A. (1992). The ecology of automaticity: Toward establishing the conditions needed to produce automatic processing effects. *The American Journal of Psychology,105*, 181–199.

Barrett, L. F., Tugade, M. M., & Engle, R. W. (2004). Individual differences in working memory capacity and dual-process theories of the mind. *Psychological Bulletin, 130*, 553–573.

Beck, L., & Grayot, J. D. (2021). New functionalism and the social and behavioral sciences. *European Journal for Philosophy of Science, 11*(4), 1–28.

Berg, N., & Gigerenzer, G. (2010). As-if behavioral economics: Neoclassical economics in disguise? *History of Economic Ideas, 18*(1), 133–166.

Blaise, C., Harris, M. B., Guerrero, J. V., & Bunge, S. A. (2012). Rethinking the role of automaticity in cognitive control. *The Quarterly Journal of Experimental Psychology, 65*(2), 268–276.

Botvinick, M. M., & Braver, T. (2015). Motivation and cognitive control: From behavior to neural mechanism. *Annual Review of Psychology, 66*, 83–113.

Buckner, C. (2015). Functional kinds: A skeptical look. *Synthese, 192*(12), 3915–3942.

Buckner, R. L., Andrews-Hanna, J. R., & Schacter, D. L. (2008). The brain's default network: Anatomy, function, and relevance to disease. *Annals of the New York Academy of Sciences, 1124*, 1–38. https://doi.org/10.1196/annals.1440.011

De Neys, W. (Ed.). (2018). *Dual Process Theory 2.0* (1st ed.). Routledge. https://doi.org/10.4324/9781315204550

De Neys, W. (2021). On dual-and single-process models of thinking. *Perspectives on Psychological Science, 16*(6), 1412–1427.

De Neys, W. (2023). Advancing theorizing about fast-and-slow thinking. *The Behavioral and Brain Sciences, 46*, e111. https://doi.org/10.1017/S0140525X2200142X

Dewey, A. R. (2023). Metacognitive control in single-vs. dual-process theory. *Thinking and Reasoning, 29*(2), 177–212.

Evans, J. S. (2006). Dual system theories of cognition: Some Issues. *Proceedings of the Annual Meeting of the Cognitive Science Society, 28*, 202–207.

Evans, J. S. B. T. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Annual Review of Psychology, 59*, 255–278.

Evans, J. S. B. T. (2010). Intuition and Reasoning: A Dual-Process Perspective. *Psychological Inquiry, 21*, 313–326. https://doi.org/10.1080/1047840X.2010.521057

Evans, J. S. B. T. (2018). Dual process theory: Perspectives and problems. In W. De Neys (Ed.), *Dual process theory 2.0* (pp. 145–164). Routledge.

Evans, J. S. B., & Frankish, K. (Eds.). (2009). *In two minds: Dual processes and beyond* (Vol. 10). Oxford University Press.

Evans, J. S. B. T., & Stanovich, K. E. (2013a). Dual-process theories of higher cognition advancing the debate. *Perspectives on Psychological Science, 8*(3), 223–241.

Evans, J. S. B. T., & Stanovich, K. E. (2013b). Theory and metatheory in the study of dual processing: Reply to comments. *Perspectives on Psychological Science, 8*(3), 263–271.

Gigerenzer, G. (2010). Personal reflections on theory and psychology. *Theory & Psychology, 20*(6), 733–743.

Gigerenzer, G. (2015). On the supposed evidence for libertarian paternalism. *Review of Philosophy and Psychology, 6*(3), 361–383.

Gigerenzer, G., & Regier, T. (1996). How do we tell an association from a rule? Comment on Sloman (1996). *Psychological Bulletin, 119*, 23–26.

Goel, V. (2008). Anatomy of deductive reasoning. *Trends in Cognitive Sciences, 11*, 435–441.

Gomez-Lavin, J. (2021). Working memory is not a natural kind and cannot explain central cognition. *Review of Philosophy and Psychology, 12*(2), 199–225.

Grayot, J. D. (2020). Dual process theories in behavioral economics and neuroeconomics: A critical review. *Review of Philosophy and Psychology, 11*(1), 105–136.

Kahneman, D. (2011). *Thinking, fast and slow*. Macmillan.

Kahneman, D., & Frederick, S. (2002). Representativeness revisited: Attribute substitution in intuitive judgment. *Heuristics and Biases: The Psychology of Intuitive Judgment, 49*(49–81), 74.

Kahneman, D., & Frederick, S. (2005). A model of heuristic judgment. In K. Holyoak & R. Morrison (Eds.), *The Cambridge handbook of thinking and reasoning* (pp. 267–293). Cambridge University Press.
Keren, G. (2013). A tale of two systems: A scientific advance or a theoretical stone soup? Commentary on Evans & Stanovich (2013). *Perspectives on Psychological Science, 8*(3), 257–262.
Keren, G., & Schul, Y. (2009). Two is not always better than one: A critical evaluation of two-system theories. *Perspectives on Psychological Science, 4*(6), 533–550.
Lakatos, I. (1978). Science and pseudoscience. *Philosophical Papers, 1*, 1–7.
Lieberman, M. D. (2003). Reflective and reflexive judgment processes: a social cognitive neuroscience approach. In J. P. Forgas, K. R. Williams, and W. von Hippel (Eds.), *Social Judgments: Implicit and Explicit Processes* (pp. 44–67). Cambridge University Press.
Marr, D. (1982). *Vision: A compuzarional approach*. Freeman & Co.
McGeer, V. (1996). Is" Self-Knowledge" an empirical problem? Renegotiating the space of philosophical explanation. *The Journal of Philosophy, 93*(10), 483–515.
Metcalf, K., Langdon, R., & Coltheart, M. (2007). Models of confabulation: A critical review and a new framework. *Cognitive Neuropsychology, 24*(1), 23–47.
Moors, A. (2014). Examining the mapping problem in dual process models. In J. W. Sherman, B. Gawronski, & Y. Trope (Eds.), *Dual process theories of the social mind* (pp. 20–34). New York, NY, USA: Guilford Press.
Moors, A., & De Houwer, J. (2006). Automaticity: A theoretical and conceptual analysis. *Psychological Bulletin, 132*(2), 297.
Mugg, J. (2013). Simultaneous contradictory belief and the two-system hypothesis. *Proceedings of the Annual Meeting of the Cognitive Science Society, 35*, 1044–1048.
Mugg, J. (2016). The dual-process turn: How recent defenses of dual-process theories of reasoning fail. *Philosophical Psychology, 29*(2), 300–309.
Murphy, C., Jefferies, E., Rueschemeyer, S. A., Sormaz, M., Wang, H. T., Margulies, D. S., & Smallwood, J. (2018). Distant from input: Evidence of regions within the default mode network supporting perceptually-decoupled and conceptually-guided cognition. *NeuroImage, 171*, 393–401. https://doi.org/10.1016/j.neuroimage.2018.01.017.PMC5883322.PMID29339310
Neander, K. (2017). *A mark of the mental: In defense of informational teleosemantics*. MIT Press.
O'Reilly, R. C., Herd, S. A., & Pauli, W. M. (2010). Computational models of cognitive control. *Current Opinion in Neurobiology, 20*, 257–261.
Osman, M. (2004). An evaluation of dual-process theories of reasoning. *Psychonomic Bulletin & Review, 11*(6), 988–1010.
Pacherie, E., & Haggard, P. (2010). What are intentions. *Conscious will and responsibility. A Tribute to Benjamin Libet, 7*, 70–84.
Piccinini, G., & Craver, C. (2011). Integrating psychology and neuroscience: Functional analyses as mechanism sketches. *Synthese, 183*(3), 283–311.
Power, J. D., & Petersen, S. E. (2013). Control-related systems in the human brain. *Current Opinion in Neurobiology, 23*, 223–228.
Raichle, M. E. (2015). The brain's default mode network. *Annual Review of Neuroscience, 38*, 433–447. https://doi.org/10.1146/annurev-neuro-071013-014030
Samuels, R. (2009). The magical number two, plus or minus: Dual-process theory as a theory of cognitive kinds. In J. S. B. Evans & K. Frankish (Eds.), In two minds: Dual processes and beyond (pp. 129–146). Oxford University Press.
Scarantino, A., & De Sousa, R. (2018). Emotion. *The Stanford Encyclopedia of Philosophy* (Summer 2021 Edition), Edward N. Zalta (Ed.), URL = <https://www.plato.stanford.edu/archives/sum2021/entries/emotion/>
Shepherd, J. (2014). The contours of control. *Philosophical Studies, 170*, 395–411.
Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin, 119*(1), 3.
Smith, E. R., & DeCoster, J. (2000). Dual-process models in social and cognitive psychology: Conceptual integration and links to underlying memory systems. *Personality and Social Psychology Review, 4*(2), 108–131.
Stanovich, K. E. (1999). *Who is rational? Studies of individual differences in reasoning*. Psychology Press.
Stanovich, K. E. (2004). *The Robot's rebellion: Finding meaning in the age of Darwin*. University of Chicago Press.

Stanovich, K.E. (2009). Distinguishing the reflective, algorithmic, and autonomous minds: Is it time for a triprocess theory. In eds. Evans, J.S.B.T., & Frankish, K. (Eds.), *In two minds: Dual processes and Beyond* (pp. 55–88). Oxford University Press.

Stanovich, K. E. (2011). *Rationality and the reflective mind*. Oxford University Press.

Stephens, R. G., Dunn, J. C., & Hayes, B. K. (2018). Are there two processes in reasoning? The dimensionality of inductive and deductive inferences. *Psychological Review, 125*(2), 218.

Stephens, R. G., Matzke, D., & Hayes, B. K. (2019). Disappearing dissociations in experimental psychology: Using state-trace analysis to test for multiple processes. *Journal of Mathematical Psychology, 90*, 3–22.

Thompson, V., & Newman, I. (2018). Logical intuitions and other conundra for dual process theories. In W. De Neys (Ed.), *Dual Process Theory 2.0*. Routledge.

Tinghög, G., Koppel, L., & Västfjäll, D. (2023). Dual-process theory is Barbapapa. *Behavioral and Brain Sciences, 46*, E144. https://doi.org/10.1017/S0140525X22003211

Vatansever, D., Menon, D. K., & Stamatakis, E. A. (2017). Default mode contributions to automated information processing. *Proceedings of the National Academy of Sciences of the United States of America, 2017*, 10521. https://doi.org/10.1073/pnas.1710521114

Weiskopf, D. A. (2011a). Models and mechanisms in psychological explanation. *Synthese, 183*(3), 313.

Weiskopf, D. A. (2011b). The functional unity of special science kinds. *The British Journal for the Philosophy of Science, 62*(2), 233–258.

Wu, W. (2014). *Attention*. Routledge.

Zahnoun, F. (2020). Explaining the reified notion of representation from a linguistic perspective. *Phenomenology and the Cognitive Sciences, 19*(1), 79–96.