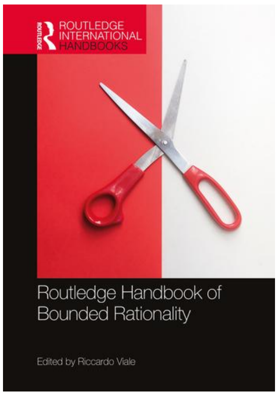

Routledge Handbook of Bounded Rationality

Riccardo Viale

Bounded rationality and dual systems

Publication details
https://test.routledgehandbooks.com/doi/10.4324/9781315658353-14

Samuel C. Bellini-Leite, Keith Frankish

How to cite :- Samuel C. Bellini-Leite, Keith Frankish. 29 Oct 2020, *Bounded rationality and dual systems from:* Routledge Handbook of Bounded Rationality Routledge
Accessed on: 22 Mar 2023
https://test.routledgehandbooks.com/doi/10.4324/9781315658353-14

12

BOUNDED RATIONALITY AND DUAL SYSTEMS

Samuel C. Bellini-Leite and Keith Frankish

Introduction

The notion of bounded rationality was initially developed by Herbert Simon as a corrective to the ideal models of rationality used in economics, psychology, and philosophy (Simon, 1957). Bounded rationality moderates the requirements on rational agents to reflect the limitations of human cognitive capacity and the conditions of the task environment (Simon, 1955, 1989). In a similar spirit, Christopher Cherniak has argued that traditional conceptions of rationality have no relevance for cognitive science. To adopt ideal general rationality conditions, Cherniak claims, is to deny or ignore our 'finitary predicament' – the basic fact that human reasoning is limited by constraints of time and cognitive capacity (Cherniak, 1986).

The history of the cognitive psychology of reasoning is to a large extent the story of accumulating evidence for human failure to meet standards of ideal rationality. Peter Wason started the modern study of reasoning with tasks, such as the Thog problem and the Wason selection task, which challenged the notion that human cognitive development culminates in a formal-logical stage (e.g., Wason, 1966). In the study of judgement and decision making, Daniel Kahneman and collaborators showed that humans routinely violate the axioms of probability theory and rational choice theory (Kahneman et al., 1982; Tversky, 1969). These pioneering advances led to the identification of various inherent biases in human cognition, and it is now common for researchers on reasoning and decision-making to adopt limited standards of rationality, though there is continued dispute over the precise nature of the limitations dictated (Gigerenzer, 1996; Kahneman, 2003).

A popular way to explain cognitive biases is to appeal to some form of *dual-process theory* (DP theory or DPT), developed by cognitive and social psychologists since the late 1970s (for reviews, see Frankish and Evans, 2009; Evans and Stanovich, 2013). Such theories hold that human reasoning and decision making are supported by two different processing systems: System 1, which is fast, autonomous, effortless, and nonconscious, and System 2, which is slow, controlled, effortful, and conscious.[1] It is common to ascribe biases to System 1 processing (which is seen as geared to norms of evolutionary rationality), and normative responses to System 2 processing. Individual differences in the activation and use of System 2 are often cited to explain why some individuals produce the modal, biased response in reasoning tasks, whereas others respond in line with standards of normative rationality (Stanovich, 1999).

But is the DP approach compatible with the tenets of bounded rationality? Does DPT really abandon the notion of ideal rationality, or does it merely treat it as a feature of System 2 rather than of cognition as a whole? Certainly, some writers in the bounded rationality tradition are suspicious of DPT. They deny that there is any fundamental division of mental systems and account for 'higher cognitive judgements' by positing a range of heuristics, from which a selection is made as needed (Gigerenzer and Regier, 1996; Gigerenzer and Selten, 2001; Kruglanski and Gigerenzer, 2011). In part at least, this negative attitude towards DPT may reflect a view of System 2 as some sort of ideal central system. For example, Gigerenzer writes:

> Heuristics are sometimes subsumed into a 'System 1' that is supposedly responsible for associations and making errors and is contrasted with a 'System 2' that embodies the laws of logic and probability, again without specifying models of the processes in either system.
>
> *2008, p. 21*

Now, one thing that is meant by saying that System 2 is normatively rational is that by engaging in effortful, conscious, System 2 thinking we can apply learned rules of inference, follow task instructions, and correct biased intuitive responses.[2] This has always been a central claim of DPT. It does not follow, however, that System 2 processes *must* be normatively rational in this sense. For we can also go wrong in our effortful, conscious thinking – applying unsound rules of inference, misunderstanding task instructions, and overriding intuitive responses that were in fact normatively correct. We may also consciously employ learned heuristics and rules of thumb, which give us quick and dirty solutions that may be inaccurate or biased. What System 2 gives us is a new level of *control*, which can be used to arrive at normatively correct responses but can also be misused to produce errors. It is not *definitional* of System 2 that it follows normatively correct principles. This point was perhaps not sufficiently stressed by early DP theorists, but it is now clearly acknowledged in the DP literature (see, e.g., Evans and Stanovich, 2013). To this extent, DPT contains no commitment to ideal rationality and there is no conflict with bounded rationality.

There is another worry about DPT, however, which concerns the role of System 2 in abductive inference and belief fixation. The worry has its roots in an earlier form of DPT proposed by Jerry Fodor, which divides the mind into encapsulated input systems and unencapsulated central systems (Fodor, 1983, 2001). This early form of DPT does involve a commitment to unbounded central processing, which it represents as sensitive to global assessments of context and relevance, and Fodor himself drew a bleak moral for cognitive science. Since later forms of DPT have been heavily influenced by Fodor, it is appropriate to ask whether this commitment carries over to them. If it does, then DPT not only conflicts with bounded rationality but also threatens to render part of the mind inaccessible to cognitive science.

We shall set out this worry in some detail and then go on to outline a reply to it. We shall show that there are various ways in which System 2 processing is limited and that contextual processing can be done by System 1. We shall conclude that DPT need not include a commitment to ideal rationality and can incorporate the insights of bounded rationality.

Fodor's DPT and the limits of cognitive science

The mental division proposed by Fodor is between input systems, which interpret perceptual stimuli, and central systems, which are involved in belief fixation and general problem-solving

(Fodor, 1983). Fodor argues that input systems are *modular*, and he develops a detailed account of the features of modular systems. One key property of modules is that they are *domain-specific*. They are dedicated to processing stimuli of a specific type, such as colours, faces, voices, or uttered sentences, and each is structured to deal with its own specific domains, perhaps incorporating information about it. A second core property of modular systems is that they are *informationally encapsulated* – insensitive to information stored elsewhere in the cognitive system. Thus, a module for colour perception would not draw on general knowledge about the colours of particular things. Fodor does concede that contextual information can bias perception, but he argues that this is because perception is not limited to input analysis, so the bias can be applied later by central systems (Fodor, 1983, p. 73). Similarly, he argues that priming effects in language (where experience of one word facilitates the recognition of another, semantically related one) may be due to superficial associations between lexical items *inside* the language module, rather than to the top-down influence of background knowledge about the world. Informational encapsulation of this kind is the core feature of modules, as Fodor conceives of them, and he suggests that their other features can be explained as resulting from it (Fodor, 1983, pp. 79–82).

Central systems, in Fodor's view, have contrasting features to input systems. They are domain neutral (not tailored to any specific type of task), and they are unencapsulated (and therefore nonmodular). The latter feature, Fodor argues, is a consequence of the fact that belief fixation typically involves *abductive inference* – finding the best explanation of the information available – and such inference is a global process, which is *isotropic* and *Quineian* (Fodor, 1983, 2001). To say that it is isotropic is to say that it is open-ended: any item of knowledge could in principle be relevant to the confirmation of any belief (knowledge about astronomy could be relevant to problems in subatomic physics; knowledge of economics could be relevant to evolutionary theory, and so on). To say that abductive inference is Quineian is to say that it is a holistic process: the degree of confirmation a belief receives depends on considerations such as simplicity and conservatism, which are determined globally, by the belief's relations to the rest of the belief system.

Fodor argues that because central processes have this isotropic and Quineian character, cognitive science cannot get any explanatory purchase on them. The argument (developed at length in Fodor, 2001) is at heart simple. Fodor argues that the only serious approach in contemporary cognitive science is the computational theory of mind, which identifies mental processes with formal operations upon mental representations. But such operations are sensitive only to local properties of representations, not to global properties of the belief system. Hence, the more global a process is, the less contemporary cognitive science understands it. Fodor calls this 'Fodor's First Law of the Nonexistence of Cognitive Science' (1983, p. 107).

This problem for computationalism, Fodor notes, manifests itself in the field of artificial intelligence as the notorious *frame problem* (McCarthy and Hayes, 1969; Pylyshyn, 1987). In traditional computational AI ('Good-Old-Fashioned AI' or GOFAI), a robot guides its behaviour by reference to an internal model of the world, and the programmer must provide it with a procedure for updating this model each time it acts, determining what will and will not change as result of each action. The problem is to find a tractable way of doing this, since, given the right context, any action could change anything. What will change depends on the background state of the world, and working out what revisions to make thus involves holistic, abductive inference. Fodor concludes that cognitive science can deal only with modular, informationally encapsulated systems, the rest being inexplicable by contemporary cognitive science and resistant to artificial modelling.

Now, if we look at Simon's description of the ideally rational 'economic man' posited by traditional economic theory, we find a similarity to Fodor's conception of central processing. Like Fodorian central systems, the economic man is able to make rational choices because he has

> knowledge of the relevant aspects of his environment which, if not absolutely complete, is at least impressively clear and voluminous ... a well-organized and stable system of preferences, and a skill in computation that enables him to calculate, for the alternative courses of actions that are available to him, which will permit him to reach the highest attainable point on his preference scale.
>
> *Simon, 1955, p. 99*

By analogy, Fodorian central processing can be thought of as a 'cognitive' economic man – a powerful homunculus with a voluminous knowledge base, a coherent system of preferences, and the computational skill to determine the rational response in the light of both. Simon notes that this idealized conception provides a poor foundation for economic theory, and we think that Fodor's conception of central systems provides a poor foundation for cognitive theory.

If DPT is committed to a similarly idealized conception of System 2, then DP theorists have reason to be worried.

From Fodor to contemporary DPT

How does Fodor's division between input systems and central systems relate to the two systems posited by contemporary DPT? There are similarities. In DPT, System 1 includes input systems of the sort Fodor discusses, while System 2 processes are all central ones. However, the correspondence is not straightforward. System 1 *also* includes many processes of reasoning, problem solving, and belief fixation, which are all central ones in Fodor's scheme. In fact, in DPT, the two sorts of processing are distinguished not by their functions, but by their computational characteristics (speed, effort, seriality, consciousness, and so on).

Another similarity with Fodor's account is that many DP theorists regard System 1 processes as modular. Stanovich, for example, conceives of System 1 as a set of adaptative problem-solving sub-systems, mostly modular in character (he refers to it as The Autonomous Set of Systems, or TASS) (Stanovich, 2004). However, the notion of modularity in play here is weaker than Fodor's. Stanovich does not treat encapsulation or domain-specificity as defining features of TASS. He notes that encapsulation is a problematic feature, which may be a matter of degree, and he holds that TASS include domain-general processes of learning and emotion-mediated behavioural regulation. In addition, he does not assume that TASS are innately specified with fixed neural structures, and he allows that modules can be developed through practice, as a controlled process becomes autonomous (Stanovich, 2004, pp. 38–40).

For Stanovich, the defining feature of TASS is that they are *autonomous* or *mandatory* – that is, they are not subject to higher-level control and, once triggered by their proprietary stimuli, they run automatically to completion. This is why they are fast and do not load on central resources, such as working memory. This does not mean that TASS/System 1 processes cannot be overridden, since higher-level processes may intervene to prevent their outputs from influencing behaviour.[3]

Despite these differences with Fodor, Stanovich's model (and modern DPT generally) still retain a sharp division between autonomous, System 1 processes and central, System 2 ones. Stanovich suggests that System 2 (he also refers to it as the *analytic system*) can be roughly characterized as exhibiting 'serial processing, central, executive control, conscious awareness, capacity-demanding operations, and domain generality in the information recruited to aid computation' (Stanovich, 2004, p. 45).

Thus, though DP theorists do not treat System 2 as comprising *all* central processes, they agree with Fodor that there are *some* genuinely nonmodular, domain-general central processes, namely, the conscious, controlled ones. The challenge, then, for DP theories is to explain how System 2 can be flexible and domain-general while still being a limited, bounded system. We shall not attempt to provide an answer here. (For a detailed proposal in the spirit of DPT, see Carruthers, 2006.) Instead, we shall review some evidence relevant to a solution, showing that the division between autonomous, nonconscious, System 1 processes and controlled, conscious, System 2 processes should not be understood in Fodorian fashion as a contrast between rigidly encapsulated systems and unbounded, ideally rational ones. We shall highlight evidence that conscious, controlled cognition is computationally limited, and thus certainly not ideally rational, and then indicate how contextual processing may be done by autonomous, nonconscious processes. This points to an explanation of how the two systems together can approximate to flexible, general-purpose cognition by balancing the computational load between them.

The limitations of System 2

Cognitive psychology shows decisively that our conscious cognitive capacities are limited. This is evidenced by work on short-term and working memory (e.g., Baddeley, 1992; Miller, 1956), attention (e.g., Simon and Chabris, 1999), competing tasks with divided attention (e.g., Schneider and Shiffrin, 1977), and cognitive theories of consciousness (e.g., Baars, 1988). When anything remotely like conscious central systems are involved, what we observe is a competition for *limited* resources. Although we can perform several automatic tasks simultaneously (for example, singing while driving), when conscious, controlled cognition is required, we have a hard time multitasking. We can see, feel, hear, and move at the same time, but we cannot attend to a lecture, do maths problems, and talk about politics at the same time.

These limitations show up in various ways. Selective attention experiments reveal that people can be conscious of only a few items in a rich perceptual scene. Subjects asked to monitor the flow of certain information in a perceptual scene, such as details in a basketball game, become unconscious of other, unrelated information in the scene.[4] Other stimuli are certainly being processed, since we react to meaningful stimuli outside a monitored flow of events, such as someone calling our name. But if we switch focus to the new stimulus, we miss aspects of the scene being monitored, revealing competition for limited capacity.

In dual-task experiments, subjects are asked to complete two tasks at the same time. Initially, task interference reduces performance, but if one of the tasks becomes predictable, performance increases. This is attributed to a process of the automatization – a transfer from controlled, System 2 processing to autonomous, System 1 processing, with the consequent freeing up of the limited resources available for the former (Schneider and Shiffrin, 1977).

The limits of conscious processing are closely bound up with those of working memory (see Baars and Franklin, 2003). (Indeed, some DP theorists define System 2 processes as the ones that load on working memory; Evans and Stanovich, 2013.) Working memory is believed to comprise three short-term storage capacities, or *loops*, for visuo-spatial, phonological, and episodic information, each of which is limited in the number of items it can retain and the time for which it can retain them (e.g., Baddeley, 1992). For example, in the Corsi block-tapping test, a visuo-spatial task that requires participants to track the order of cubes the experimenter points to, people can only keep up with a sequence of about five. Likewise, the phonological loop, which can be used to mentally rehearse words and numbers for a few minutes, has a limit of about seven items (Miller, 1956).

Could limited-capacity systems like these be isotropic and Quineian? We shall return to this later, but it is obvious at the outset that they could not be Quineian, since that would require that the confirmation of a single belief be influenced by many others (in principle, all of them), and certainly by many more than the seven-item limit of phonological working memory. As Carruthers notes: 'it has traditionally been assumed by philosophers that any candidate new belief should be checked for consistency with existing beliefs before being accepted. But in fact consistency-checking is demonstrably intractable, if attempted on an exhaustive basis' (2006, p. 52).Of course, a rational agent must have some sort of coherence in their belief set, but they cannot be conscious of all the important confirmation relations in their whole web of beliefs.

We suspect that the problem here stems from the fact that Fodor explicitly models the psychological processes of abductive inference and belief fixation on the processes of theory construction and confirmation in science (Fodor, 1983, pp. 104–105). There are certainly parallels between them, but there are also big differences – not least the fact that science is a social enterprise, which is carried out over an extended period of time and with large amounts of external scaffolding in the form of records and notes. At any rate, while the truth-value of a belief or the consistency of a preference may depend on its links to many other mental states, no more than a few of these links can be checked consciously. This is one thing that condemns us to bounded rationality.

If the conscious mind is so limited, why should we be tempted to assign ideal rationality to it? Perhaps the answer is that we identify our conscious processes with *us* – with the self that is supposed to be responsive to norms of rationality. From a phenomenological perspective, this is tempting, and it may lead us to think of System 2 as a powerful homunculus – an executive system or 'central meaner' (Dennett, 1991), which accesses all our knowledge and is the locus for rational decision making. This would be a serious mistake, however. Our conscious, controlled processes are *part* of our minds, just as our automatic, nonconscious ones are, and they must depend heavily on nonconscious, autonomous processing.[5]

Where does this leave us? Fodor supposed that central processes had to be highly sensitive to context and background information precisely because autonomous, peripheral systems were not (Fodor, 2001). Within the context of DPT, we can reverse this line of reasoning. Since System 2 processes are in fact severely limited, it follows that System 1 must do the work of supplying relevant, contextualized information for conscious central systems to work with (Evans, 2009). As we shall see in the next section, this conclusion accords well with new models of brain functioning, which locate contextual processing in nonconscious, probabilistic systems that are not limited by the capacity of working memory or other central resources.

The flexibility of System 1

If System 1 processing is flexible and context-sensitive, then System 2 can be freed from the burden of ideal rationality. It need not work on complex webs of content but merely on those items that have been pre-selected for relevance by System 1. And there are, in fact, good reasons for thinking that this is indeed the case.

First, as we have already noted, DP theories generally employ a weaker notion of modularity than Fodor's, placing less stress on features such as encapsulation. Carruthers, for example, notes that although modules must be *frugal*, in the sense that they draw on only a limited amount of information in their processing, it does not follow that they each have access only to a *specific* subset of information. A module might have access in principle to all the information in the system but draw on only a tiny fraction of it on each occasion, perhaps using simple search heuristics to make the selection (Carruthers, 2006, pp. 58–59).

Second, it is now widely accepted that modules needn't be persisting, innately specified structures. Rather, they may be 'soft modules' – temporary constructions, assembled from pre-existing components to deal with specific environmental problems. This point is stressed by Michael Anderson (Anderson, 2014). Although a critic of rigid modularity, Anderson does not deny that the cortex contains functionally differentiated regions, which are biased towards different response profiles (Anderson, 2014, p. 52). However, he argues that these regions can interact with each other and form coalitions to perform other tasks. When existing functionally differentiated regions are unable to solve a task as they stand, coalitions between them can be formed 'on the fly' to attempt a local resolution. Anderson dubs such constructions, composed of coalitions of different networks, 'Transiently Assembled Local Neural Systems' or TALoNS. These coalitions generate new possibilities of interaction but are still constrained by the available abilities of the participating regions. Being formed for specific tasks, these coalitions can generate responses in a rapid, effortless manner, without engaging resource-hungry, conscious control processes.

Third, recent models of brain functioning undermine the rigid Fodorian distinction between strictly encapsulated systems and free unencapsulated systems, even within perceptual processing itself. According to currently popular *predictive processing* (PP) theories, the brain's goal is to predict proximal stimuli rather than simply to process past input (e.g., Friston, 2005; Clark, 2013, 2016). These theories see the brain as having a hierarchical structure, with lower levels being close to perceptual input mechanisms and higher levels receiving information from diverse multimodal regions. Signals flow both up and down the hierarchy, with top-down signals predicting lower-level activity and bottom-up signals flagging errors in those predictions. Predictions are based on probabilistic *generative models* distributed through the hierarchy. These models monitor statistical patterns in the layer below, generating nonconscious hypotheses in an attempt to accommodate incoming data.[6]

Within this architecture, the influence of different neural regions is modulated to reflect their success in prediction, through a mechanism known as *precision weighting*. Neural regions that do badly must accept input from other regions, whereas ones that constantly do well are granted more autonomy, allowing a form of encapsulation. Thus, the informational closure of a region reflects its success in prediction, generating a soft module, like one of Anderson's TALoNS. As Clark puts it: 'Distinctive, objectively identifiable, local processing organizations … emerge and operate within a larger, more integrative, framework in which functionally differentiated populations and sub-populations are engaged and nuanced in different ways so as to serve different tasks …' (2016, p. 150). Coalition formation and shifts in control are determined contextually, by success in accommodating new stimuli.

If this approach is on the right track, then Fodor is quite wrong to characterize input systems as strongly encapsulated. For, as Clark emphasizes, context sensitivity is pervasive in the PP account of perception, with top-down predictions guiding how stimuli are interpreted and signal distinguished from noise. For example, if someone says to you, 'Oedipus married his own nother', the currently dominant top-down prediction might constrain interpretation so that the /n/ sound is nonconsciously interpreted as /m/, and the word is consciously heard as 'mother' (Clark, 2013).

Such contextually sensitive processing can be done rapidly since stimuli that can be well accommodated by current predictions are ignored. Instead of letting all worldly information move up to higher layers, lower layers pass on only stimuli that generate high prediction error. This selectional effect is further enhanced by precision weighting. As Clark puts it, 'very low-precision prediction errors will have little or no influence upon ongoing processing and will fail to recruit or nuance higher level representations' (2016, p. 148). By these means, relevance is selected for right from the start, facilitating rapid responding.

A central feature of the PP framework is that it draws no sharp boundary between perception and cognition (or, indeed, between perception, cognition, and action). Judgement-like top-down predictions are continually modifying and being modified by perception-like bottom-up error signals. Thus, what DPT theorists would think of System 1 judgements may be more akin to perceptual processes than to explicit System 2 judgements, sharing in the context-sensitivity of perception. If one's previously calm companion starts to show signs of anger, bottom-up error signals will cause one's generative models to adapt, yielding a new prediction with a judgement-like form ('He is angry and may become violent').

Although few DP theorists have explicitly adopted PP approaches, some have stressed the perception-like character of System 1 judgements. For example, Kahneman and Frederick observe that the perception/judgement boundary 'is fuzzy and permeable: the *perception* of a stranger as menacing is inseparable from a *prediction* of future harm' (Kahneman and Frederick, 2002, p. 50). They note, moreover, that the heuristics and biases research programme (which was a major inspiration for dual-process theories) was from the start 'guided by the idea that intuitive judgments occupy a position between the automatic parallel operations of perception and the controlled serial operations of reasoning' (Kahneman and Frederick, 2002). There are good reasons, then, for thinking that the dual-process conception of System 1 and PP make a natural partnership.

The moral of this is that the representations that are made available for explicit System 2 will have already gone through considerable prior processing and filtering. Only high error (unexpected and thus informative) stimuli reach higher levels of processing and become available for explicit System 2 reasoning. Thus, on the view we have outlined, System 2 reasoning would be neither isotropic nor Quineian. It would not be isotropic because it would not be true that any element could influence it (it receives only a small subset of the data, selected as relevant) and it would not be Quineian since it would not have access to all the information required to check for holistic consistency. In fact, it would be limited in just the ways the evidence suggests it in fact is. Thus, it could meet only bounded standards of rationality.

Conclusion

It is tempting to think of conscious thought as open-ended and unconstrained, and of nonconscious processes as inflexible and encapsulated. Cognitive psychology and computational neuroscience show that this is wrong. Conscious cognition is in fact severely limited in capacity, while nonconscious processes are tuned to the heavy demands of contextual and relevance processing. It is thus vital that modern dual-process theories do not carry over the conception of rationality implicit in Fodor's precursor theory. In thinking about the functions and capacities of the posited dual systems, the perspective of bounded rationality is essential. System 2, if it exists, is not an idealized reasoner, a cognitive analogue of Simon's 'economic man'.

Acknowledgements

The authors thank Riccardo Viale for his detailed and helpful comments on an earlier draft of this chapter.

Notes

1 More recently, there has been a tendency to speak of two *types of processing* rather than two systems (see e.g., Evans and Stanovich, 2013; Bellini-Leite, 2018). For present purposes, however, we shall use 'systems' terminology, since it better expresses the concerns we wish to address about architectural interpretations of dual-process theory.

2 Of course, the fact that System 2 can be used to override intuitive responses does not mean that it will be. System 2 thinking is effortful and demanding of cognitive resources, and people engage in it only when they have sufficient time, motivation, and cognitive capacity. And even when they do engage in it, they may still fail to override an intuitive System 1 response. For discussion, see Viale (2018).

3 Another similarity between Stanovich's DPT and Fodor's architecture is that Stanovich sees System 2 representations as decoupled from online perceptual processing, so that they can be used for offline, hypothetical thinking (e.g., Stanovich and Toplak, 2013). However, this decoupling need not bring with it the problematic features of Fodorian central processing. Decoupled representations are detached from the world, but this does not mean that they are integrated with the rest of the system's knowledge in reasoning processes that are Quineian and isotropic. Hypothetical thinking may involve only a limited number of representations held in working memory.

4 The most effective illustration is Simons and Chabris' (1999) experiment, in which participants who are required to the count passes in a video of a basketball game completely fail to notice when a person in a gorilla costume appears.

5 In fact, those who see System 2 as a personal-level system – in the sense that its processes involve intentional actions – are also those who argue against a 'central meaner' and for an illusionist view of consciousness (Dennett, 1991; Frankish 2015, 2016).

6 The claim that System 1 is supported by probabilistic calculations does not entail that the system's overt responses will respect probabilistic principles. The calculations govern the system's internal workings, and in principle its overt responses could conform to different principles, such as those of classical logic. This is not to say that we cannot plausibly infer internal computational principles from behaviour, but we must consider a wider range of evidence, including the speed, accuracy, and range of overt responses.

References

Anderson, M. (2014). *After phrenology: Neural reuse and the interactive brain*. Cambridge, MA: MIT Press.

Baddeley, A. (1992). Working memory. *Science*, 255(5044): 556–559.

Baars, B. J. (1988). *A cognitive theory of consciousness*. Cambridge: Cambridge University Press.

Baars, B., and Franklin, S. (2003). How conscious experience and working memory interact. *Trends in Cognitive Science*, 7(4): 166–172.

Bellini-Leite, S. (2018). Dual process theory: Systems, types, minds, modes, kinds or metaphors? A critical review. *Review of Philosophy and Psychology*, 9(2): 213–225.

Carruthers, P. (2006). *The architecture of the mind*. Oxford: Oxford University Press.

Cherniak, C. (1986). *Minimal rationality*. Cambridge, MA: MIT Press.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, 36(3): 181–204.

Clark, A. (2016). *Surfing uncertainty: Prediction, action, and the embodied mind*. Oxford: Oxford University Press.

Dennett, D. (1991). *Consciousness explained*. New York: Little, Brown.

Evans, J. (2009). How many dual-process theories do we need? One, two, or many? In J. Evans and K. Frankish (Eds.), *In two minds: Dual processes and beyond* (pp. 33–54). Oxford: Oxford University Press.

Evans, J., and Stanovich, K. (2013). Dual-process theories of higher cognition: Advancing the debate. *Perspectives on Psychological Science*, 8(3): 223–241.

Fodor, J. (1983). *The modularity of mind*. Cambridge, MA: MIT Press.

Fodor, J. (2001). *The mind doesn't work that way*. Cambridge, MA: MIT Press.

Frankish, K. (2015). Dennett's dual-process theory of reasoning. In C. Muñoz-Suárez and F. De Brigard (Eds.), *Content and consciousness revisited* (pp. 73–92). Cham: Springer.

Frankish, K. (2016). Illusionism as a theory of consciousness. *Journal of Consciousness Studies*, 23: 11–39.

Frankish, K., and Evans, J. (2009). The duality of mind: An historical perspective. In J. Evans and K. Frankish (Eds.), *In two minds: Dual processes and beyond* (pp. 1–29). Oxford: Oxford University Press.

Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 360(1456): 815–836.

Gigerenzer, G. (1996). On narrow norms and vague heuristics: A reply to Kahneman and Tversky. *Psychological Review*, 103(3): 592–596.

Gigerenzer, G. (2008). Why heuristics work. *Perspectives on Psychological Science*, 3(1): 20–29.

Gigerenzer, G., and Regier, T. (1996). How do we tell an association from a rule? Comment on Sloman (1996). *Psychological Bulletin*, 119(1): 23–26.

Gigerenzer, G., and Selten, R. (Eds.) (2001). *Bounded rationality: The adaptive toolbox*. Cambridge, MA: MIT Press.

Kahneman, D. (2003). Maps of bounded rationality: Psychology for behavioral economics. *The American Economic Review*, 93(5): 1449–1475.

Kahneman, D., and Frederick, S. (2002). Representativeness revisited: Attribute substitution in intuitive judgment. In T. Gilovich, D. Griffin, and D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 49–81). Cambridge: Cambridge University Press.

Kahneman, D., Slovic, P., and Tversky, A. (Eds.) (1982). *Judgment under uncertainty: Heuristics and biases*. Cambridge: Cambridge University Press.

Kruglanski, A. W., and Gigerenzer, G. (2011). Intuitive and deliberative judgments are based on common principles. *Psychological Review*, 118(1): 97–109.

Miller, G. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 63(2): 81–97.

McCarthy, J., and Hayes P. (1969). Some philosophical problems from the standpoint of artificial intelligence. In B. Meltzer and D. Michie (Eds.), *Machine intelligence* 4 (pp. 463–502). Edinburgh: Edinburgh University Press.

Pylyshyn, Z. (Ed.) (1987). *The robot's dilemma: The frame problem in artificial intelligence*. Norwood, NJ: Ablex.

Schneider, W., and Shiffrin, R. (1977). Controlled and automatic human information processing I: Detection, search and attention. *Psychological Review*, 84(1): 1–66.

Simon, H. A. (1955). A behavioral model of rational choice. *The Quarterly Journal of Economics*, 69(1): 99–118.

Simon, H. A. (1957). *Models of man: Social and rational – Mathematical essays on rational human behavior in a social setting*. New York: Wiley.

Simon, H. A. (1989). Cognitive architectures and rational analysis: Comments. Paper presented at 21st Annual Symposium on Cognition, Carnegie-Mellon University.

Simons, D., and Chabris, C. (1999). Gorillas in our midst: Sustained inattentional blindness for dynamic events. *Perception*, 28(9): 1059–1074.

Stanovich, K. (1999). *Who is rational? Studies of individual differences in reasoning*. Mahwah, NJ: Lawrence Erlbaum.

Stanovich, K. (2004). *The robot's rebellion: Finding meaning in the age of Darwin*. Chicago, IL: University of Chicago Press.

Stanovich, K. E., and Toplak, M. E. (2012). Defining features versus incidental correlates of Type 1 and Type 2 processing. *Mind & Society*, 11(1): 3–13.

Tversky, A. (1969). Intransitivity of preferences. *Psychological Review*, 76(1): 31–48.

Viale, R. (2018). The normative and descriptive weaknesses of behavioral economics-informed nudge: Depowered paternalism and unjustified libertarianism. *Mind & Society*, 17(1–2): 53–69.

Wason, P. (1966). Reasoning, in B. Foss (Ed.), *New horizons in psychology* (pp. 135–151). Harmondsworth: Penguin.