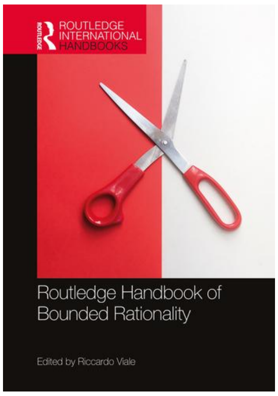

Routledge Handbook of Bounded Rationality

Riccardo Viale

Bounded rationality, reasoning and dual processing

Publication details
https://test.routledgehandbooks.com/doi/10.4324/9781315658353-12
Jonathan St B. T. Evans
Published online on: 29 Oct 2020

10

BOUNDED RATIONALITY, REASONING AND DUAL PROCESSING

Jonathan St. B. T. Evans

Introduction

Traditionally, philosophers distinguish between instrumental and epistemic rationality which may be defined roughly as follows:

> <u>Instrumental rationality</u>
> Acting in such a way as to achieve one's goals.
> <u>Epistemic rationality</u>
> Attaining and maintaining true beliefs.

Both concepts are slippery. Take, for example, my goals of being an Open Golf champion and the world's most famous cognitive psychologist. As I approach 70, the chance of the former had vanished more than 50 years ago, as several years of play and practice had resulted in only a middling 15 handicap. As to the latter, any realistic hope was extinguished perhaps 40 years ago. But I am surely not *irrational* for failing to achieve these goals, even if I was wildly optimistic in even contemplating them. This, of course, is what Herb Simon's *bounded rationality* is all about. We can only achieve goals that are within our capability, either as a species or as individuals. It does make the whole idea of instrumental rationality a bit messy though. What goals should we strive to attain? Only those that are fully realistic? Or should we set our sights high to reach the limits we can?

Let us take the game of chess, whose study fascinated Simon. I also love the game and in semi-retirement play a lot online and study the game, striving to improve my standard. The game is, of course intractably difficult due to the vast search space. No human can ever master it, nor will any machine we can imagine, even though they have now overtaken the best human players. Except in simplified endgame positions, or those allowing a forced win, no human or machine can be sure that it is picking the best move. None of us can optimise; so as Simon put it, in chess as in all complex decision making, we can only *satisfice* – choose moves which appear to be good enough. Unfortunately, the heuristic search methods promoted by Simon (e.g., Newell & Simon, 1972) fell by the wayside in computer AI which came to be dominated by brute force as computers acquired ever bigger memories and faster processors. Eventually, a program of this type, Deep Blue, came to beat the world champion Gary Kasparov in 1997, even if there were deep suspicions about human intervention by the IBM team at the time. Writing recently about this, Kasparov (2017) dismisses these suspicions on the grounds that the 'human-like' moves played by Deep Blue on an IBM supercomputer were replicated within a few years by chess engines available cheaply on personal computers. He had long understood that computers would eventually prevail – it was just a matter of *speed*. What is remarkable, however, is how well the best human players can compete with the current generation of programs that can search billions of chess positions before selecting a move. Significantly, Kasparov points out that this is because computers are weak in two areas where humans are strong: pattern recognition and strategic thinking.

The striking thing about the cognitive science of chess is just how *differently* computers had to be designed to play the game in order to match human ability. This shows us that the human brain, with its massively parallel architecture, has adapted to the solution of complex problems by means that do not require brute force computation. And of course, Simon was one of the first to understand the role of perception and pattern recognition in human expertise (Chase & Simon, 1973). Like machines, human players do calculate continuations and move sequences, but only a tiny fraction of what the machines do. The recognition of *which* moves deserve attention, cued by patterns learnt by long study, is what gives humans playing strength. This, of course, is an example of *dual processing* (Evans, 2008; Evans & Stanovich, 2013). These rapid recognition processes (Type 1, autonomous) greatly reduce the load of explicit calculation (Type 2, engaging working memory). If we think of rationality as depending simply on explicit calculation and reasoning, then we do our species a great disservice. Our working memories are small and slow but our capacity for implicit learning of patterns is vast. Thus, the bounds on our rationality are a lot less severe than one might imagine. Unfortunately for the human race, however, recent developments in artificial intelligence have shown that chess programs based on machine learning and pattern recognition are even better than the best of brute force engines. Reportedly, AlphaZero, originally developed to play Go but adapted for chess, taught itself to beat the world's best brute force program in just four hours (Klein, 2017)!

If instrumental rationality is a slippery concept, then so too is epistemic rationality. In fact, the psychology of reasoning has been radically reconstructed over the past 20 years, as the model handed by philosophers proved inadequate. The traditional emphasis was on the importance of logic as the basis of rational thinking (Henle, 1962; Inhelder & Piaget, 1958), which led to the foundation of the traditional study of deductive reasoning by psychologists. If you think epistemic rationality is about *truth*, then deductive reasoning is very important because it is truth-preserving. That is to say, any deductively valid conclusion is true, provided that its premises are true. How disappointing, then, when hundreds of laboratory studies showed that human adults, typically of undergraduate student level intelligence, made numerous logical errors and showed systematic cognitive biases (Evans, Newstead, & Byrne, 1993). Moreover, they were heavily influenced by the content or meaning of logical problems, which was irrelevant to their formal structure. Such findings led to a sharp rationality debate, when the philosopher Jonathan Cohen claimed that such experimental findings could not demonstrate that people were irrational (Cohen, 1981). Eventually, psychologists working this field confronted the dilemma before them: either people are inherently irrational or else there was something wrong with the paradigm they were using to study reasoning (Evans, 2002; Oaksford & Chater, 1998).

One problem is that nothing new can be learnt by deductive reasoning, which makes it a poor standard for rationality. We can form new beliefs by induction, but we cannot be sure that they are *true*. Strangely, induction has been the poor relation, with far fewer psychological studies of inductive inference carried out (Feeney & Heit, 2007). Reasoning from the particular to the general is immensely beneficial but not logically valid. Such inductions are not truth-preserving and their conclusions can only be held provisionally, or with a degree of probability pending further evidence. But how could we cope in the world without the ability to learn? Even if one takes a modular view of the human mind, dividing it into many specialised information processing systems, one must assume that many of these have the ability to learn and modify themselves with experience (Carruthers, 2006).

Those working within the 'new paradigm' in the psychology of reasoning (Elqayam & Over, 2012; Evans, 2012; Oaksford & Chater, 2013; Over, 2009) accept that people routinely use their beliefs when reasoning and allow them to draw conclusions with degrees of probability. Some authors have suggested that Bayesian theory should replace the role once played by logic (Oaksford & Chater, 2007) although this is not a necessary constraint. It is important to recognise, however, that the new paradigm does not mean that psychologists have switched from studying deduction to induction, which is still a separate paradigm. Because the new paradigm allows people to express degrees of belief in their conclusions, it would be easy to think that it was inductive inferences that were being studied. However, the field has rather moved towards the study of *uncertain* deduction, where uncertainty in premises is reflected by uncertainty in conclusions drawn from them (Evans & Over, 2012; Evans, Thompson, & Over, 2015). At one conference, I put up a slide with one of Sherlock Holmes' famous 'deductions' (see Box 9.1) and asked the audience of reasoning researchers whether the inference he drew was actually inductive, deductive, or abductive. The great majority voted for 'inductive' but I was able to show that Holmes' argument was in fact deductively valid. Their intuition, correctly, was that Holmes' inference was uncertain. However, the uncertainty lay entirely in the *premises* from which his deduction was made. Watson may have picked the mud up somewhere else, visited the Post Office to meet a friend, and so on. But if Holmes' premises were true, then so was his conclusion.

Box 9.1 Extract from Conan-Doyle's *The Sign of Four* (1890)

(HOLMES TO WATSON) "Observation shows me that you have been to the Wigmore Street Post-Office this morning, but deduction lets me know that when there you dispatched a telegram."

"Right!" said I. "Right on both points! But I confess that I don't see how you arrived at it. It was a sudden impulse upon my part, and I have mentioned it to no one."

"It is simplicity itself," he remarked, chuckling at my surprise, – "so absurdly simple that an explanation is superfluous; and yet it may serve to define the limits of observation and of deduction. Observation tells me that you have a little reddish mould adhering to your instep. Just opposite the Wigmore Street Office they have taken up the pavement and thrown up some earth which lies in such a way that it is difficult to avoid treading in it in entering. The earth is of this peculiar reddish tint which is found, as far as I know, nowhere else in the neighbourhood. So much is observation. The rest is deduction."

"How, then, did you deduce the telegram?"

"Why, of course I knew that you had not written a letter, since I sat opposite to you all morning. I see also in your open desk there that you have a sheet of stamps and a thick bundle of post-cards. What could you go into the post-office for, then, but to send a wire? Eliminate all other factors, and the one which remains must be the truth."

The one domain in which truth is valued above all else is science. Science strives for true explanations and the avoidance of contradictions. If two theories conflict in their proposals then we take it for granted that at least one will be false, even if there is no immediate means to prove which. The history of science teaches us that false beliefs, for example, about the motion of the planets, or the evolution of life will be eventually abandoned, no matter how much prior belief is operating in their favour. Important though science is, however, it is only one domain in which (some) people need to operate. Many scientists also hold religious and political beliefs, for example, which may be directly inconsistent with their scientific beliefs, or at least not provable by the scientific method. As another example, my scientific view of psychology is inconsistent with many of the tenets of folk psychology which I resist in my scientific thinking. In everyday life, however, I apply the native folk psychology which enables all of us to function socially. Our built-in folk psychology and theory of mind are good examples of fit-for-purpose beliefs. They do not stand up to scientifically rigorous examination, but they don't need to. They are fit for the purpose for which they evolved, which is communication, social interaction and the formation of social structures.

In discussing chess earlier, I pointed out that our pattern recognition and implicit learning abilities compensate to a great extent for the limited computational capacity of executive processing, requiring working memory. So one approach to bounded rationality is the idea that we can often rely heavily on gut feelings and quick and simple heuristics for our decision making (Dijksterhuis, Bos, Nordgren, & von Baaren, 2006; Gigerenzer, 2007; Gigerenzer & Todd, 1999; Gladwell, 2005). However, there are also vast amounts of research showing the importance of general intelligence, working memory and explicit forms of reasoning in the solution of a wide variety of cognitive tasks. This supports the proposals of dual-process theory (Evans & Stanovich, 2013), in which rapid and intuitive forms of processing (Type 1) are said to combine with slower and more effortful reasoning, the latter engaging working memory (Type 2). One of the dual-process theorists who most strongly advocates the importance of Type 2 processing in rational decision making is Keith Stanovich, who together with colleagues applied the theory to individual differences in reasoning and decision-making ability (Stanovich, 1999, 2011; Stanovich, West, & Toplak, 2016). The argument here is that while people often rely on heuristics and intuitions, this can lead them into serious errors and biases. Such errors can be avoided by successful intervention with high effort reasoning, provided that the individual has the cognitive capability and the relevant knowledge. So, at this point, I need to examine the dual-process case in more detail.

Default-interventionist dual-process theory and bounded rationality

Dual-process theory is something of an intellectual minefield for several reasons. One problem is that there are many authors in cognitive and social psychology advocating various form of dual processing which appear similar but may differ in detail. All broadly identify two forms of processing: Type 1 which is intuitive and typically quick, and Type 2 which is reflective and typically slow. One important structural difference is between theories that have a *parallel-competitive* approach and those which are *default-interventionist* (terms originally introduced by Evans (2007b). Parallel-competitive theories envisage that associative (Type 1) and rule-based (Type 2) processes proceed in parallel and compete for control of our behaviour (Sloman, 1996; Smith & DeCoster, 2000). The more popular form, however, is what I termed 'default-interventionist' (DI). Theorists of this persuasion include myself, Stanovich and Kahneman (Evans, 2007a; Evans & Stanovich, 2013; Kahneman, 2011; Stanovich, 2011). The common framework is the assumption that Type 1 processes, being relatively quick and automatic, provide

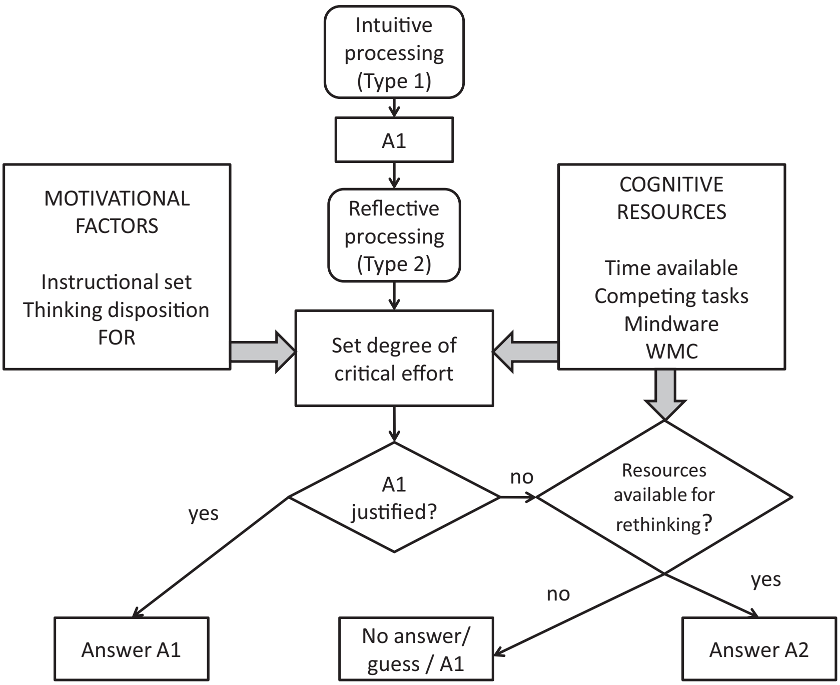

Figure 10.1 The default-interventionist process

default intuitions which are subject to scrutiny and possible intervention by Type 2 processing, which are slower and effortful. A great deal of research has been conducted by now showing that the answers people give to typical laboratory problems are affected by both experimental manipulations (e.g., time available, concurrent working memory load, instructions) and by individual differences in both cognitive ability and cognitive style. This work is reflected in the DI theory I proposed a few years ago (Evans, 2011) and shown in Figure 10.1.

The model assumes that Type 1 processing leads to a default intuitive answer A1 which appears in consciousness without effort or reflection. A1 is then subject to scrutiny by Type 2 processes which are effortful and require working memory. However, such scrutiny can be shallow or deep, which depends on a number of factors. Motivational factors include how people are instructed, whether they have a rational thinking disposition (which inclines people to check intuitions more carefully) and the feeling of rightness (FOR) or subjective confidence that they have in A1 (Thompson, Prowse Turner, & Pennycook, 2011). Cognitive resources include time available (speeded tasks will reduce Type 2 checking), individual differences in working memory capacity (highly correlated with IQ) and 'mindware', of which more shortly. When A1 is scrutinised, it may or may not be deemed satisfactory. Whether the intuition can successfully be replaced by an answer based on explicit reasoning and knowledge, however, depends on the cognitive resources available as Stanovich (2011) has argued. He has provided much empirical evidence (see also Stanovich et al., 2016) on how decision making is affected by individual differences in both cognitive ability and rational thinking dispositions. The latter measure the extent to which people will rely on intuitions or check them out by reasoning. On the basis of this work, Stanovich argues that a number of conditions must be met for someone to successfully engage in Type 2 reasoning and solve a demanding and novel experimental problem. First, people must be aware of the need for intervention by Type 2 reasoning, which is where rational thinking dispositions and feelings of rightness come into play. However, if they do intervene, then this will only be successful if both (1) they have sufficient cognitive capacity for the hypothetical reasoning involved; and (2) they have the relevant mindware – that is, formal or procedural knowledge required by the context.

The DI dual-process framework is clearly one of bounded rationality. Stanovich (2011) writes extensively about the *cognitive miser*, which is the idea that we evolved to expend as little mental effort as we can get away with. Most of our behaviour, most of the time is controlled by autonomous, Type 1 processes. This makes sense when you consider that Type 2 or executive processing engaged in working memory is a limited and essentially singular resource which can only perform one task at a time with any efficiency. Hence, this valuable resource, which corresponds with the focus of our conscious attention, must be used sparingly and for the most important tasks only. However, in laboratory tasks at least, the cognitive miser default can lead to cognitive biases and normative errors. Stanovich suggests this happens in three ways (see Stanovich, 2011, pp. 98–104). First, the system may detect no need for Type 2 intervention and default to autonomous processing. Second, a simplified and low effort form of Type 2 processing may be used to solve a problem and is equally likely to lead to a biased response. Finally, a genuine effort may be made at intervention but fail to change the default answer. As Elqayam (2012) points out, only those equipped with sufficient ability and mindware to solve the problems can fairly be accused of cognitive miserliness should they fail to do so. Stanovich (e.g., 2009) has a particular focus on the case of people with high IQs who nevertheless fail to make what he considers to be rational decisions.

There are some correspondences with Stanovich's arguments in my own writing and theoretical framework. For example, I have suggested (Evans, 2007a) that there is both a *fundamental heuristic bias* (Type 1) and a *fundamental analytic bias* (Type 2) and I apportioned approximately equal 'blame' to each in explaining the various biases in the literature. The Type 1 bias is to focus only on information which is preconsciously cued as relevant and the Type 2 bias is to focus on singular hypotheses and fail to consider alternatives. The latter corresponds to Stanovich's lazy and ineffective Type 2 processing. I also agree with Stanovich that the intervention with Type 2 reasoning does not necessarily lead to an improved response, as indicated in Figure 10.1. However, there are a couple of differences in our approach, also. First, in all my models, Type 2 processing is involved before a decision is made, even though the default answer may well continue unchanged. Second, a major reason why people persist with A1 in Figure 10.1 is that they seek justification for an initial intuition and often end up *rationalising* this response. In fact, my very earliest papers on dual processing, in collaboration with Pete Wason, were concerned entirely with this rationalising function (Evans & Wason, 1976; Wason & Evans, 1975). On the Wason selection task, with which these papers were concerned, it was many years before I could find clear evidence that Type 2 processing might lead to a change of the intuitive response at all, although I eventually did (Evans & Ball, 2010).

This rationalisation aspect of Type 2 processing does not fit as neatly with the cognitive miser hypothesis because it does require mental effort and working memory resources to carry out, even though the default intuition is maintained. However, it is supported by evidence that we spend most time thinking about responses that are intuitively cued (Ball, Lucas, Miles, & Gale, 2003; Evans, 1996). It also fits with evolutionary accounts of explicit reasoning that link this facility to argumentation and social discourse (Mercier & Sperber, 2011). Entertaining this argument for a moment, one way in which an evolved argumentation function could become adapted to the assistance of problem solving would be precisely in the kind of model I am proposing here. Instead of simply making an intuitive response, we first seek arguments and justifications, by an internal mental process, to support them. While this may bias the system towards rationalisation and confabulation, it does at least allow the possibility that error will be detected and corrected. This is more likely to happen when more time and effort are expended on examining the intuitive response, which reflects the individual differences in *rational thinking dispositions* studied by Stanovich and others. These measures, even when based on self-report, do predict accuracy on cognitive tasks when effects of cognitive ability are statistically removed, as many studies in cognitive and social psychology have shown.

Metacognitive feelings

Metacognition is the knowledge or feeling that we have about our own cognitive processes. When we make judgements or predictions, for example, these may come with feelings of confidence which we can report upon if asked. A well-established literature in the study of judgement and decision-making reports a general phenomenon of overconfidence. When we are more confident, we are in fact more likely to be right, but not as likely as we think we are (Koehler, Brenner, & Griffin, 2002). There has been a lot of research into feeling of knowing (FOK) in memory research, which again shows imperfect correlation with actual accuracy of memories (Koriat, 1993). Of more relevance to this chapter and the bounded rationality argument is the suggestion that these metacognitive feelings have a regulatory purpose, controlling the amount of effort that is expended (Ackerman, 2014).

Valerie Thompson (2009) was the first to introduce the issue of metacognitive feelings to the psychology of reasoning. It was an attempt to solve a fundamental issue for dual process theories at the time of writing, namely, what determines whether an intervention with Type 2 processing will occur. She suggested that the default intuitive answer that comes to mind carries with it a feeling of rightness (FOR) which determines the amount of effort that will be made. When FOR is high, we will tend to accept our intuitions with little effort at reasoning and when it is low, we will intervene more. A number of empirical studies followed which supported these claims. Thompson invented a two-response paradigm, in which people are asked to give a quick intuitive answer together with the FOR rating. They are then allowed to think further for as long as they like and change the answer if they wish. On a number of different cognitive tasks, Thompson and colleagues showed that low FOR leads to greater rethinking time and more likelihood that initial answers will be changed (Thompson et al., 2011).

On the face of it, the FOR research seems to fit a neat story of evolution and bounded rationality. Our intuitions come helpfully packaged with feelings of confidence which tell us whether we need to expend effort or not in checking them out. However, there is a snag. None of this research – on reasoning and decision making – has provided evidence for even a partial accuracy of the FOR judgement, as observed in other fields where metacognition is studied. The answers which people rethink and change are as likely to be right as wrong and low FOR can lead to abandonment of an initially correct judgement. Some cues which lead to errors are particularly convincing, such as the *matching bias* which dominates choices on the abstract selection task (Thompson, Evans, & Campbell, 2013) and the compelling but wrong intuitions on the cognitive reflection task (Frederick, 2005). Belief bias is a major cause of error in syllogistic reasoning (Evans, Barston, & Pollard, 1983; Klauer, Musch, & Naumer, 2000) and yet believable conclusions inspire positive feelings of liking and confidence (Morsanyi & Handley, 2012; Shynkarkuk & Thompson, 2006). None of which evidence supports the claim that we evolved these metacognitive feelings because they are of assistance to the cognitive miser.

There are two problems with all of these metacognitive measures. First, they are introspective reports whose validity can only be established by relating them to other behavioural measures. Second, these relations are correlational rather than causal. What we know about FOR is that strong feelings are associated with answers that are given quickly and generally held on to with little effort to rethink them. Putting these two problems together, we can see that there is no clear reason to think that FOR corresponds to a stage of cognitive processing in the manner that researchers in the field assume. They could simply be conscious feelings that are by-products of the processes that determine the behaviour measures. In other words, a brain process that leads to a quick and firmly held decision, also generates conscious feelings of rightness and confidence, which in themselves are of no functional significance.

Recent studies with the Thompson two-response paradigm have suggested than on some simpler forms of syllogistic reasoning, most of the people who get it right after reflection, also got it right with their initial intuitive response (Bago & De Neys, 2017). So, it appears that no slow reflective Type 2 process was apparently involved. Connected to this is another recent finding that people of higher cognitive ability or with better mindware for the numerical problems studied have stronger intuitions leading them to the right answers, whereas those of lower ability are more likely to be influenced intuitively by prior belief (Thompson, Pennycook, Trippas, & Evans, 2018). There are actually a number of recent studies showing that people may have 'logical intuitions' (De Neys, 2012, 2017). While some authors have taken such findings as evidence against the traditional default-interventionist account, the case is not that clear-cut (Evans, 2018). First, most of these recent studies use relatively simple tasks where cues for logically correct response might well be present without need for working memory engagement. Second, those of higher cognitive ability may have acquired mindware that has become automated with practice. Although laboratory tasks that are both novel and requiring a degree of complexity for their solution typically require Type 2 processing, it is a fallacy to assume that Type 1 processes are necessarily biased or that Type 2 processing always leads to correct answers (Evans & Stanovich, 2013).

Conclusion

In the early part of this chapter, I showed that Simon's notion of bounded rationality applies both to how we can achieve our goals and to how we can organise and validate our knowledge about the world. I also argued, using the domain of chess, that while our capabilities of explicit calculation are sharply limited compared with modern computers, the brain has evolved rapid, pattern recognition systems which compensate to a large extent for this limitation. When we examine the field of dual processing in the study of reasoning and decision making, however, the dice have been loaded heavily in favour of tasks that require Type 2 processing (explicit calculation and reasoning loading on working memory) for their solution. This is because the tasks presented tend to be novel and difficult to solve on the basis of prior experience. For the same reason, this literature has given relatively bad press to intuitive or Type 1 processes which operate quickly and autonomously. In fact, the same kinds of process that enable the best human chess players to approach the standard of brute force computer programs! And yet researchers are somehow surprised when they simplify the standard laboratory tasks and find a lot of accurate Type 1 processing is involved in their solution.

I have shown that the default-interventionist dual process theories that have dominated in these fields are essentially within the bounded rationality tradition. That is to say, it is assumed that Type 1 processing will dominate by default and high effort Type 2 processing will be selectively applied. There is much accumulated evidence that this is the case (Evans, 2007a; Evans & Stanovich, 2013; Stanovich, 2011) although recent studies have somewhat muddied the waters

(De Neys, 2017). The studies by Thompson and colleagues of feelings of rightness are of much interest but produce a mixed picture. On the one hand, FOR most definitely correlates with the amount of time and effort people expend on reasoning, as opposed to reliance on initial intuitions. On the other hand, there is no evidence that these feelings are related to actual accuracy, and some major cognitive biases (e.g., matching bias, belief bias) are supported by strong and false feelings of confidence. It is at least possible that these feelings are no more than a by-product of the brain processes that lead to strongly or weakly held decisions.

While DI dual-process theory is under pressure from those inside as well as outside of the paradigm, we should not lose sight of the broader picture. First, there is no doubt that the great bulk of our behaviour is controlled by fast and automatic processing. Outside of the laboratory, such processing is often very effective and compensates for the limitations of our explicit computational capacity. There is also no doubt that we do have a central working memory system which is implicated in the successful execution of a very large range of cognitive functions (Baddeley, 2007). So the way in which our brains provide our bounded rationality simply *has* to be some kind of dual process story, whether we call these Type 1 and 2 or just autonomous and executive processes. Although some evolutionary psychologists have argued for a massively modular brain that does not depend on any general-purpose reasoning system (Cosmides & Tooby, 1992) and other prominent researchers state that we can safely rely on intuitions and gut feelings for complex decisions (Gigerenzer, 2007), these argument simply do not hold up. More than 100 years of research on general intelligence and 50 plus years of research on working memory attest to the fact that there is a central cognitive system allowing behaviour to be based on reasoning and hypothetical thought and which operates distinctly from autonomous processing. The intellectual success of our own species quite clearly depends upon it.

Acknowledgements

The author is grateful to Shira Elqayam for a critical reading of an earlier draft of this chapter.

References

Ackerman, R. (2014). The diminishing criterion role for metacognitive regulation of time investment. *Journal of Experimental Psychology General*, 143, 1349–1368.

Baddeley, A. (2007). *Working memory, thought and action*. Oxford: Oxford University Press.

Bago, B., & De Neys, W. (2017). Fast logic?: Examining the time course assumption of dual process theory. *Cognition*, 158, 90–109. doi:10.1016/j.cognition.2016.10.014

Ball, L. J., Lucas, E. J., Miles, J. N. V., & Gale, A. G. (2003). Inspection times and the selection task: What do eye-movements reveal about relevance effects? *Quarterly Journal of Experimental Psychology*, 56(6), 1053–1077.

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

Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology*, 4, 55–81.

Cohen, L. J. (1981). Can human irrationality be experimentally demonstrated? *Behavioral and Brain Sciences*, 4, 317–370.

Cosmides, L., & Tooby, J. (1992). Cognitive adaptations for social exchange. In J. H. Barkow, L. Cosmides & J. Tooby (Eds.), *The adapted mind: Evolutionary psychology and the generation of culture* (pp. 163–228). New York: Oxford University Press.

De Neys, W. (2012). Bias and conflict: A case for logical intuitions. *Perspectives on Psychological Science*, 7, 28–38.

De Neys, W. (Ed.) (2017). *Dual process theory 2.0*. London: Routledge.

Dijksterhuis, A., Bos, M. W., Nordgren, L. F., & von Baaren, R. B. (2006). On making the right choice: The deliberation-without-attention effect. *Science*, 311, 1005–1007.

Elqayam, S. (2012). Grounded rationality: Descriptivism in epistemic context. *Synthese*, 189, 39–49.

Elqayam, S. & Over, D. E. (2012). New paradigm in psychology of reasoning: Probabilities, beliefs and dual processing. *Mind & Society*, 11, 27–40.

Evans, J. St. B. T. (1996). Deciding before you think: Relevance and reasoning in the selection task. *British Journal of Psychology*, 87, 223–240.

Evans, J. St. B. T. (2002). Logic and human reasoning: An assessment of the deduction paradigm. *Psychological Bulletin*, 128, 978–996.

Evans, J. St. B. T. (2007a). *Hypothetical thinking: Dual processes in reasoning and judgement*. Hove: Psychology Press.

Evans, J. St. B. T. (2007b). On the resolution of conflict in dual-process theories of reasoning. *Thinking & Reasoning*, 13, 321–329.

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

Evans, J. St. B. T. (2010). *Thinking twice: Two minds in one brain*. Oxford: Oxford University Press.

Evans, J. St. B. T. (2011). Dual-process theories of reasoning: Contemporary issues and developmental applications. *Developmental Review*, 31, 86–102.

Evans, J. St. B. T. (2012). Questions and challenges for the new psychology of reasoning. *Thinking & Reasoning*, 18, 5–31.

Evans, J. St. B. T. (2018). Dual-process theory: Perspectives and problems. In W. De Neys (Ed.), *Dual process theory 2.0* (pp. 137–155). London: Routledge.

Evans, J. St. B. T., & Ball, L. J. (2010). Do people reason on the Wason selection task?: A new look at the data of Ball et al. (2003). *Quarterly Journal of Experimental Psychology*, 63(3), 434–441.

Evans, J. St. B. T., Barston, J. L., & Pollard, P. (1983). On the conflict between logic and belief in syllogistic reasoning. *Memory & Cognition*, 11, 295–306.

Evans, J. St. B. T., Newstead, S. E., & Byrne, R. M. J. (1993). *Human reasoning: The psychology of deduction*. Hove: Lawrence Erlbaum.

Evans, J. St. B. T., & Over, D. E. (2012). Reasoning to and from belief: Deduction and induction are still distinct. *Thinking & Reasoning*, 19(3–4), 267–283. doi:10.1080/13546783.2012.745450.

Evans, J. St. B. T., & Stanovich, K. E. (2013). Dual process theories of higher cognition: Advancing the debate. *Perspectives on Psychological Science*, 8, 223–241.

Evans, J. St. B. T., Thompson, V. A., & Over, D. E. (2015). Uncertain deduction and conditional reasoning. *Frontiers in Psychology*, 6. doi:10.3389/fpsyg.2015.00398.

Evans, J. St. B. T., & Wason, P.C. (1976). Rationalisation in a reasoning task. *British Journal of Psychology*, 63, 205–212.

Feeney, A., & Heit, E. (2007). *Inductive reasoning*. Cambridge: Cambridge University Press.

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives*, 19(4), 25–42.

Gigerenzer, G. (2007). *Gut feelings: The intelligence of the unconscious*. London: Penguin.

Gigerenzer, G., & Todd, P. M. (1999). *Simple heuristics that make us smart*. Oxford: Oxford University Press.

Gladwell, M. (2005). *Blink*. London: Penguin.

Henle, M. (1962). On the relation between logic and thinking. *Psychological Review*, 69, 366–378.

Inhelder, B., & Piaget, J. (1958). *The growth of logical thinking*. New York: Basic Books.

Kahneman, D. (2011). *Thinking, fast and slow*. New York: Farrar, Straus and Giroux.

Kasparov, G. (2017). *Thinking deeply: Where machine intelligence ends*. London: John Murray.

Klauer, K. C., Musch, J., & Naumer, B. (2000). On belief bias in syllogistic reasoning. *Psychological Review*, 107, 852–884.

Klein, M. (2017). Google's AlphaZero destroys Stockfish in 100 game match. Available at: www.chess.com/news/view/google-s-alphazero-destroys-stockfish-in-100-game-match.

Koehler, D. J., Brenner, L., & Griffin, D. (2002). The calibration of expert judgment: heuristics and biases beyond the laboratory. In T. Gilovich, D. Griffin, & A. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment*. Cambridge: Cambridge University Press.

Koriat, A. (1993). How do we know what we know? The accessibility model of the feeling of knowing. *Psychological Review*, 100, 609–639.

Mercier, H., & Sperber, D. (2011). Why do humans reason? Arguments for an argumentative theory. *Behavioral and Brain Sciences*, 34, 57–111.

Morsanyi, K., & Handley, S. (2012). Logic feels so good – I like it. Evidence for an implicit detection of logicality in syllogistic reasoning. *Journal of Experimental Psychology-Learning Memory and Cognition*, 38, 596–616.

Newell, A., & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice-Hall.

Oaksford, M., & Chater, N. (1998). *Rationality in an uncertain world*. Hove: Psychology Press.

Oaksford, M., & Chater, N. (2007). *Bayesian rationality: The probabilistic approach to human reasoning*. Oxford: Oxford University Press.

Oaksford, M., & Chater, N. (2013). Dynamic inference and everyday conditional reasoning in the new paradigm. *Thinking & Reasoning*, 19(3–4), 346–379. doi:10.1080/13546783.2013.808163.

Over, D. E. (2009). New paradigm psychology of reasoning. *Thinking & Reasoning*, 15, 431–438.

Shynkarkuk, J. M., & Thompson, V. A. (2006). Confidence and accuracy in deductive reasoning. *Memory & Cognition*, 34, 619–632.

Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin*, 119, 3–22.

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*. Mahwah, NJ: Lawrence Erlbaum Associates.

Stanovich, K. E. (2009). The thinking that IQ tests miss. *Scientific American*, November/December, 33–39.

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

Stanovich, K. E., West, C., & Toplak, M. E. (2016). *The rationality quotient: Towards a test of rational thinking*. Cambridge, MA: MIT Press.

Thompson, V. A. (2009). Dual-process theories: A metacognitive perspective. In J. S. B. T. Evans & K. Frankish (Eds.), *In two minds: Dual processes and beyond* (pp. 171–196). Oxford: Oxford University Press.

Thompson, V. A., Evans, J. S. T., & Campbell, J. I. D. (2013). Matching bias on the selection task: It's fast and feels good. *Thinking & Reasoning*, 19(3–4), 431–452. doi:10.1080/13546783.2013.820220.

Thompson, V. A., Pennycook, G., Trippas, D., & Evans, J. St. B. T. (2018). Do smart people have better intuitions? *Journal of Experimental Psychology General*. 147(7), 945–961.

Thompson, V. A., Prowse Turner, J. A., & Pennycook, G. (2011). Intuition, reason, and metacognition. *Cognitive Psychology*, 63(3), 107–140.

Wason, P. C., & Evans, J. St. B. T. (1975). Dual processes in reasoning? *Cognition*, 3, 141–154.