DRAFT- Please do not quote without permission
This paper provides a summary of a much larger ongoing project
Most recent changes 20/03/06 Current Word count: 8,216

Dual-process and Cognitive Checking

"A dichotomy is only as valuable as the explanatory power of the hypothetical systems that it distinguishes".
Sloman (1996, page, 20)

In 1968 Wason published findings which appeared to demonstrate an alarming degree of human irrationality. Only 5-10% of participants gave the correct answer on his selection task, which was designed to test conditional reasoning about the falsification of a simple rule. Wason's selection task was to become the paradigm demonstration of poor human reasoning in a large and ever growing body of evidence showing that our reasoning is riddled with errors and biases. The findings in this area, which has become known as heuristics and biases research, have been cited as having "bleak implications for human rationality*"* (Nisbett & Borgida, 1975, cited in Botterill & Carruthers, page 105.). However, many philosophers (notably Cohen, 1981 & 1982) have argued that it is conceptually impossible for there to be wide spread human irrationality. This has led to an ongoing debate regarding the extent of human rationality. Those arguing that humans are not really so irrational have been quick to point out that although individuals make errors on such reasoning tasks they are capable of understanding why their answers were wrong and the reasons for giving the correct answer. Furthermore, the standard logical rules by which we judge human rationality are themselves a product of human rationality. This has led to a distinction between possible reasoning competence, which is our potential capacity for reasoning correctly, and actual reasoning performance, which is how we actually reason in a given situation. It is not the quality of our reasoning competence which has "bleak implications for human rationality" but the fact that our reasoning performance appears to display an abundance of systematic reasoning errors which suggests there could be widespread human irrationality.

This distinction between competence and performance opens the door for arguments claiming the data do not indicate human irrationality but actually indicate that certain unusual scenarios provoke reasoning errors. Further evidence for this claim comes from research using structurally the same reasoning tasks but set in more familiar terms or using normative rules. These deontic versions of the various reasoning tasks that had previously shown poor reasoning, elicited much better performance. However, only certain deontic versions improve performance whilst others continue to provide evidence of poor reasoning. This has led to a complicated debate over how to interpret the data and how to explain the psychological mechanisms that give rise to it. The interpretational issue is whether the data indicates systematic errors in human reasoning or that performance errors generated by the unusual nature of the task are interfering with our otherwise rational reasoning processes. This interpretational issue has now become central to the rationality debate. However, this issue cannot be settled without a detailed understanding of the psychological mechanisms which give rise to this data. Only once we have an understanding of the mechanisms underlying human reasoning can we judge if errors are systematic or triggered by certain cues. This is why an understanding of these mechanisms is the key step to understanding the extent and limitations of human rationality and thus advancing or even settling the rationality debate.

This will be the case even if we choose to adopt a different notion of rationality, although what needs comparing in order to make this assessment will differ. Stein (1996) argued against the ‘standard picture’ of rationality in which we assess rationality in terms of reasoning in accord with normative theories of logic based on probability theory, deductive logic, inductive principles, and decision theory. Stein argued that how humans ought to reason should be relative to how they can reason. This is similar to comparing actual reasoning performance with possible reasoning competence. Evans and Over (1996) offered another view of rationality with a distinction between ‘rationality 1’ aimed at trying to achieve practical goals and ‘rationality 2’ aimed at reasoning in accordance with normative theories. Even if we assess human reasoning in terms of achieving practical goals or against possible competence, gaining an understanding of the psychological mechanisms is a necessary step in assessing the extent and limitations of human reasoning. This is because there is still a need to gain an understanding of possible reasoning competence, to compare it with actual reasoning performance and to discover which factors give rise to differences between performance and competence. Whichever notion of rationality we accept, gaining a detailed understanding of the psychological mechanisms underlying human reasoning is the key to advancing the rationality debate.

Any satisfactory account of the psychological mechanisms underlying human reasoning must be able to explain the two central phenomena in heuristics and biases research. The first is the difference between possible reasoning competence and actual reasoning performance. The second is why different versions of structurally the same task give rise to differences in reasoning performance. This constraint has been central in the generation of accounts of the mechanisms underlying human reasoning. This has led to all the theories having a detailed explanation of how these phenomena occur, but often lacking sufficient detail in their accounts of general reasoning to be considered full and detailed theories of human reasoning. A satisfactory account of human reasoning must also be able to predict and explain everyday reasoning which is not generated in unusual scenarios or by unusual instructions. Furthermore, it is often overlooked that perhaps the most important application of an explanation of the mechanisms underlying human reasoning is to develop methods of teaching people reliable reasoning techniques which will enable them to improve their reasoning performance.

In their 1996 book *Rationality and Reasoning,* Evans and Over proposed a dichotomy/duality in processing to account for the difference between possible reasoning competence and actual reasoning performance. This dichotomy was loosely based on earlier dual-process accounts of reasoning developed by Evans (1977, 1982 & 1989). Their dual-process theory proposed a rule or model based reasoning system which generated high reasoning competence and another associative reasoning system which would produce errors if the nature of a task did not suit its mode of operation. This, coupled with a focus on the practical goals of rationality, not only gave an explanation of the central phenomena in heuristics and biases research, but also went a long way towards providing a complete account of the mechanisms underlying human reasoning. However, their theory lacked detail in a number of key areas. In particular there was no detailed or satisfactory account of the relationship between the two systems. This could hardly be expected given the scope of the 1996 project. However, since 1996 there has been relatively little focus on attempting to fill in this missing detail. Dual-process theory has been applied to other areas besides human reasoning[1] and a number of pieces of recent research claim to have generated further evidence for the theory[2]. Yet, despite this

[1] Notably, cognitive social psychology (Chaiken and Thorpe, 1999).
[2] Notably, Klaczynski and Daniel (2005), Evans and Curtis-Holmes (2005)

apparent success there has been no attempt to detail the commitments of dual-process theory and develop it from a detailed framework to a fully-fledged account of the mechanisms underlying human reasoning. This is particularly disappointing because dual-process theory is, to date, the account which is closest to providing a comprehensive theory of the mechanisms underlying human reasoning.

Section 1 of this essay will attempt to outline in detail the commitments of and evidence for, dual-process theory. It will also identify the areas in which the dual-process account lacks detail. I will argue that much of the explanatory power of dual-process theory is dependent on an account of the interaction between system 1 and system 2 processes. Section 2 will briefly assess relevance theory (Sperber et al, 1995) and the domain-specific reasoning module account (Cosmides and Tooby, 1992), which are the other leading explanations of the psychological mechanisms underlying human reasoning. I will argue that only Dual-process theory has the potential to give a full account of human reasoning and that for this reason it should be further developed. I will also consider where, if at all, dual-process theory conflicts with the other accounts of human reasoning. Section 3 examines in detail the possible models of interaction between system 1 and system 2 processes. I will argue for my 'Complex Cognitive Checking' model which is loosely based on Haidt's (2001) 'Social Intuitionist' model of moral judgement rather than the conflicting models advocated by Evans (2003). If we accept this model we should consider human cognition as one process which is made up of at least two kinds of processing rather than a dual-process akin to two minds competing within the same brain.

Section 1- Dual-process theory

Dual-process theory proposes a dichotomy underlying human reasoning, with two reasoning systems operating in totally distinct ways. This distinction is so clear and central to the account that Jonathan Evans (2003, page 454) claimed dual-process theories "essentially posit two minds in one brain".

System 1

System 1 is the evolutionarily older system characterised by associative cognition and rapid automatic parallel processing. It operates using a set of heuristics which evolved to overcome specific problems which we faced in our evolutionary history. System 1 is usually thought of as tacit, with only the outcome reaching conscious attention. This is thought to give rise to the phenomenal experience of intuition. Although the exact computational nature of system 1 is not detailed, the workings of the system are usually explained in terms of the associative learning process generated by neural networks (McLeod et al, 1998) or in terms of determining relevance. Evans and Over (1996) proposed the workings of system 1 are constrained by the first (cognitive) principle of Relevance: "Human cognitive processes are aimed at processing the most relevant information in the most relevant way" (Sperber et al, 1995, page 48). Relevance is determined by the cognitive effect of processing some information (generated by the association of new information with already existing information) minus the cognitive effort involved in processing that information. System 1 is not just one system but "…comprises a set of autonomous subsystems that include both innate input modules and domain-specific knowledge acquired by a domain-general learning mechanism" (Evans, 2003, page 454). This means that, although its primary function is to determine relevance by a simple system of weightings, system 1 processing is capable of competently performing this procedure in order to achieve a multitude of different goals. There is virtually no limit to system 1's processing capacity which enables its many sub-sections to run parallel processing, allowing many tasks to be performed simultaneously. However, system 1 processes are limited in respect of being automatic, habitual and determined by past learning. The way in which the system works is fixed and unaffected by instructions to aim to achieve goals other than those for which it is pre-programmed. System 1 is often thought of as 'quick and dirty' or 'rough and ready' processing that provides an answer quickly, which may not be completely accurate but hopefully accurate enough for the task in hand. Such processes are necessary because there are usually constraints on both the time and processing effort which can be devoted to reasoning.

System 2

System 2 is the evolutionarily more recent system characterised by rule-or model-based cognition and slow sequential processing, which is constrained by the limited capacity of working memory. This system is thought of as controlled explicit reasoning which gives rise to the phenomenal experience of thinking and our introspective reports of how we reason. Although language is not the only mode of operation attributed to system 2, its workings are often 'thought aloud', which Evans and Over (1996) attribute to the use of verbal working memory. They claim our possession of language is almost certainly a prerequisite for reflective consciousness and thus our capacity for explicit reasoning. However, Evans and Over (1996) do draw attention to the limitations of introspective knowledge when trying to investigate the workings of system 2. They emphasis that evidence shows introspective accounts generated by system 2 are individuals theorising in order to rationalise their own behaviour rather than an accurate insight into how they actually reason (Wason & Evans, 1975; Nisbett & Wilson, 1977). System 2 provides the ability to perform abstract thinking, allowing us to reason about and simulate hypothetical scenarios or future possibilities. This system is much more flexible, enabling it to change the goals of its processing in accordance with those ascribed by specific task instructions. It is this uniquely human system which allows us to anticipate the future and deal with novel situations for which we could not have learnt a strategy from experience. The exact computational workings of this system are yet to be detailed. Much of the processing of system 2 can be accounted for by the construction of mental models of possible outcomes as suggested by Johnson-Laird et al (1994), but some of the system's processing may be rule-based. Sloman (1996) suggested that there may be room to develop an account where system 2 is further divided into one rule-based and one model-based system.

There is evidence that this system is affected by the goals set by task instructions and the amount of emphasis placed on them (Evans, 2000). The fact that taking formal logic improves some students reasoning ability under exam conditions suggests their system 2 processes can be taught to reason in accordance with logical norms. This supported by evidence that using certain teaching methods can improve participants performance on the selection task (Barwise & Etchemendy,1994 cited in Bringsjord, Bringsford & Noel, 1998.). However, there is no guarantee these goals or logical rules are being internalised into the process. There is evidence that when people are taught the principles of logic which are required to give the correct answer on the selection task they revert back to normatively incorrect answers attributed to system 1 when the scenario varies slightly (Cox & Griggs, 1982 and Cheng et al, 1986). This suggests that at least the procedure for inhibiting system 1 is not being generalised. It may also indicate that these rules and goals are not internalised.

Interaction between the two systems

Since the first dual-process account of human reasoning the explanation of the relationship between the two systems has varied. Evans (1977 & 1982) proposed a conflict model in which conscious logical and unconscious non-logical processes were in direct competition. Later Evans (1989) proposed a sequential account in which system 2 processes take over where the system 1 processes leave off. When the term 'dual-process theory' was revived in Evans and Over's (1996) book *Rationality and Reasoning* they favoured an interactive model. The nature of this interaction is that system 2 processes are *"*shaped, directed and limited by tacit, pre-attentive processes*" (*page *146),* but they have a limited capacity to suppress system 1 processes. However, more recent explanations have favoured a 'conflict' account of the relationship between the two systems. Evans (2003, page 454) claimed there is "a range of experimental psychological evidence showing that the two systems compete for control of our inferences and actions". It is unclear whether this is a denial of the interaction posited in *Rationality and Reasoning*, or whether the reliance of system 2 on system 1 is acknowledged but downplayed because of the focus on competition between the two systems. However, the evidence which Evans cites is far from conclusive and does not rule out a cooperative model. This is a central and fundamental aspect of dual-process theory which lacks sufficient detail. This is something which Evans (2003, page 458) recognises in his comments that *"*an important challenge is to develop models to show how such two distinct systems interact in one brain and to consider specifically how the conflict and competition between the two systems might be resolved in the control of behaviour". However, this concession fails to grasp how much the whole theory rests on a good account of this interaction. Much of the explanatory power of the dual-process theory rests on the switch between the two reasoning systems. It is an account of how this switch takes place and what determines it, which will allow the dual-process theory to be used practically both to predict how people will reason and to teach them better reasoning strategies.

<u>The Evidence for Dual-Process theory</u>

The evidence which is most commonly cited in favour of the dual-processing theory is the explanation it offers of cognitive illusions like the selection task (Wason, 1968), the conjunction fallacy (Tversky & Kahneman, 1983) and belief bias effects (Evans, 1983 & Evans et al, 2002). Due to space constraints this essay will only examine one of these cognitive illusions (the belief bias effect) in detail and give a general account to cover the rest. There is also evidence supporting dual-process theory from the explanation of individual differences in reasoning ability (Stanovich, 2003, Stanovich and West, 2000 & 1998). Recently neuropsychological evidence has also been cited in support of dual-process theory (Geol & Dolan, 2003).

Generally speaking, the initial evidence for the dual-process account is the way in which positing two distinct reasoning systems explains why research consistently finds a difference between possible reasoning competence and actual reasoning performance. In cases where actual performance fails to match possible reasoning competence, such as the poor reasoning which is often displayed in heuristics and biases research, this is because system 1 generated the answer but the task did not suit its mode of operation. Performance which matches reasoning competence is generated in one of two ways. Either the task was suited to the heuristics system 1 employs; or system 2 processes superseded the system 1 processes. The former is usually used to explain why various versions of structurally the same task produce different results, although this might also be due to the latter. This second explanation also accounts for individual differences in reasoning ability, by claiming the individuals who display better reasoning performance are those who are more prone to system 2 processes overriding system 1 processes.

Dual-process theorists have often argued that system 1 processes are not related to general intelligence but system 2 processes, specifically the ability to override system 1 processes, should be strongly linked to measures of general intelligence. Stanovich and West (2003, 2000 &1998) found that the ability to give normatively correct answers to a range of reasoning tasks was consistently associated with those who had high cognitive ability (measured using SAT scores). Those with high cognitive ability were more able to resist the biases created by belief and contextual effects. Furthermore they found that the ability to solve deontic versions of tasks (such as the selection task) was not associated with cognitive ability, whereas the ability to solve the abstract versions was. This supports the dual-process account because solving deontic versions is attributed to the scenario giving cues that suit system 1's mode of operation.

Evans' (1983) investigation of belief-bias used syllogisms to create a conflict between logical rule-based reasoning and reasoning determined by the individual's prior beliefs about the truth of a given conclusion. These syllogisms are either conflicting, where the validity of the argument does not match expectations about the truth of the conclusion, or non-conflicting where the validity of the argument does match such expectations. An example of a conflicting syllogism (in this case valid, but the conclusion is expected to be false) is:

- No nutritional things are inexpensive
- Some vitamin tablets are inexpensive
- Therefore, some vitamin tablets are not nutritional

An example of a non-conflicting syllogism (in this case invalid and the conclusion is expected to be false) is:

- No millionaires are hard workers
- Some rich people are hard workers
- Therefore, some millionaires are not rich people

Participants are instructed to treat the task logically and support only the conclusions which necessarily follow from the premises. Participants accepted nearly 90% of the valid believable conclusions and over 50% of the valid unbelievable conclusions, whereas they accepted roughly 70% of the invalid believable conclusions and less than 10% of the invalid unbelievable conclusions. The results showed that both logic and the expected truth of the conclusion influence participants' decisions. Although reasoning was strongly influenced by the expected truth of a statement (this is indicated by the high number of invalid believable conclusions accepted), logical rules were still a factor in decision-making (this is most clearly indicated by the high number of normatively correct answers in the valid but unbelievable syllogisms). The dual-process theory proposes that although participants attempt to reason logically using system 2 they find it very difficult to override the belief bias generated by system 1 processing, which is competing for control over the responses. Evans (2000) found further evidence for this account when his results indicated that emphasising the instructions to reason logically significantly reduced the effects of belief bias.

Goel and Dolan (2003) used FMRI scans to investigate brain activity during reasoning about the syllogisms used in Evans' belief-bias research. They found that on trials where logically correct answers were given participants were using their right inferior prefrontal cortex; whereas when incorrect belief-biased responses were given they were using their ventral medial prefrontal cortex. They also cite a number of different studies to support their claim that the ventral medial prefrontal cortex is responsible for intuitive and heuristic responses which are a definitive feature of system 1 processing. They concluded that this provides evidence for the dual-process theory account of two different systems competing for control over decisions. They propose that the prefrontal cortex is vital in detecting and resolving such conflict.

Section 2- Other accounts of human reasoning

The major shortfall of dual-process theory is the lack of a detailed account of the interaction between system 1 and system 2 processes. Unfortunately this is a fundamental part of the dual-process account without which the theory has little explanatory power. However, a brief examination of the other accounts of human reasoning reveals that they are equally, if not further away from, providing a comprehensive account of human reasoning.

Domain-Specific Reasoning Modules

One of the most well known theories is Cosmides and Tooby's (1992) 'domain-specific reasoning modules' account. Their theory draws on evolutionary psychology taking into account the problems which our minds will have evolved to overcome. They argue that different evolutionary problems will have different optimal solutions, and therefore multiple specific solutions will be better than one general solution. Their theory proposes that the human mind evolved lots of domain-specific reasoning modules, each with algorithms for reasoning about solutions to certain problems which we faced throughout our evolutionary history. The main evidence they cite for this claim is that the deontic versions of the selection task which elicit better reasoning performance involve cheater detection, which can be attributed to a specific module for detecting cheaters in social exchanges. They also cite evidence that performance on probabilistic reasoning is dramatically improved when participants are given frequencies rather than percentages or a number between 1 and 0, which can again be attributed to a domain-specific module for reasoning about frequencies. Cosmides and Tooby's theory proposes massive modularity to account for variations in human reasoning in different domains and given different cues.

However, in cases where the evolutionary pressures were not significant enough or one problem only occurred infrequently then it seems reasonable to propose one module which does a reasonable job of solving two problems rather than two modules which solve each problem optimally. Furthermore, there may be no need to propose a domain-specific module where a process can be equally well accounted for by domain general reasoning, because this posits a more complicated explanation than is required. There is evidence that relevance theory can explain many of the findings attributed to domain-specific reasoning modules[3] (Sperber et al, 1995), although Sperber (in press) argues that relevance theory is compatible with a massively modular approach. The domain-specific reasoning module account conflicts with dual-process accounts because it largely ignores any role that system 2 processes could play in human reasoning. The dual-process theory attributes many cases of improved reasoning performance to the domain-general workings of system 2 whereas the domain-specific reasoning module account attributes the improved performance to specialised modules. However, the biggest problem with this theory of human reasoning is that it says little about the overall reasoning process. There is no explanation of the interaction between all these domain-specific modules or of reasoning for which there is no domain-specific module. The theory is designed to explain variations in reasoning performance, rather than to give a comprehensive account of the psychological mechanisms underlying human reasoning.

Relevance Theory

Relevance theory (Sperber & Wilson, 1986 and Sperber et al, 1995) proposes a set of principles which drive all human reasoning. The first (cognitive) principle of relevance theory is: "Human cognitive processes are aimed at processing the most relevant information in the most relevant way" *(*Sperber et al, 1995, page 48*)*. The way in which

[3] My own preliminary research suggests that improved performance on deontic versions of the selection task is caused by the scenario making the correct answers more salient (therefore more relevant) rather than instructions to look for cheaters.

relevance is maximised is perhaps best understood as a form of cognitive economy where relevance can be expressed as: ***effect*** minus ***effort***. Effect is generated by new information being processed in light of already available beliefs. If combining the new information with an existing belief generates a cognitive effect (e.g. creating a new belief, or contradicting an existing belief etc) then the new information is relevant to the context of the existing belief and the individual performing the processing. However effect comes at the cognitive cost of the processing effort involved in generating the effect in the first place. The more cognitive effort required to process the new information the less relevant it is. In order for there to be a large degree of relevance the information generating the effect must be easily accessible or considerations of cognitive effort will greatly reduce the degree of relevance. This accounts for variation in performance on various versions of reasoning tasks used in heuristics and biases research. The versions which elicit improved performance are ones where the instructions or scenario increase the relevance of the correct answer. There is certainly evidence that this is the case for the selection task (Sperber et al, 1995).

The second (communicative) principle is: "Every utterance conveys a presumption of its own relevance" (Sperber et al, 1995, page 48*)*. This means that when processing information which has been presented either verbally or in writing there is an assumption that there is a relevant effect. There is also an assumption that the effect must be sufficient to justify the effort required to process the information. This means that possible cognitive effects (relevant interpretations of the information) are considered in order of ease of comprehension and once an adequate level of relevance is reached no more interpretations are considered. In experimental settings the expectations of relevance are usually low. This is because the information is not usually relevant to participants in any way other than specifying a task they are willing to perform. Cognitive effort plays a central role in directing attention and retrieving background information when there are no strong expectations of effect. This means in most instances cognitive effort will dictate the interpretation of communicative attempts in experimental settings and the most obvious interpretation will be favoured. The communicative principle and considerations of cognitive effort can combine to explain the gap between possible reasoning competence and actual reasoning performance. If there are no strong expectations of effect then merely presenting participants with instructions that have no specific relevance to them is likely to induce a very shallow level of processing.

This fits the results displayed across heuristics and biases research where participants display poor reasoning unless the instructions make it absolutely clear what kind of strategy needs to be used to achieve the correct answer. Relevance theory is compatible with dual-process theory to some extent because the principles of relevance theory can be used to explain how system 1 processes work. However, it is not clear relevance theory should be applied to this level of processing rather than having a more central or domain general role.

Relevance theory seems like a plausible account but it is more like a set of principles or slogans that apply to reasoning than a fully fledged explanation of the psychological mechanisms underlying human reasoning. Although these principles can explain the central phenomena in heuristics and biases research they say little about how the processing is carried out. The theory gives an account of how considerations of cognitive effort will lead us to engage in ‘quick and dirty’ processing to get a rough answer; or how high levels of relevance lead to more detailed processing. But the theory says little or nothing about how this processing is carried out. It is also hard to know at what level to apply these principles. They could apply only to directing processing attention or they could apply to cognition at every level. The principles proposed by relevance theory may well apply to human reasoning, but these principles alone fall a long way short of being a comprehensive account of the mechanisms underlying human reasoning.

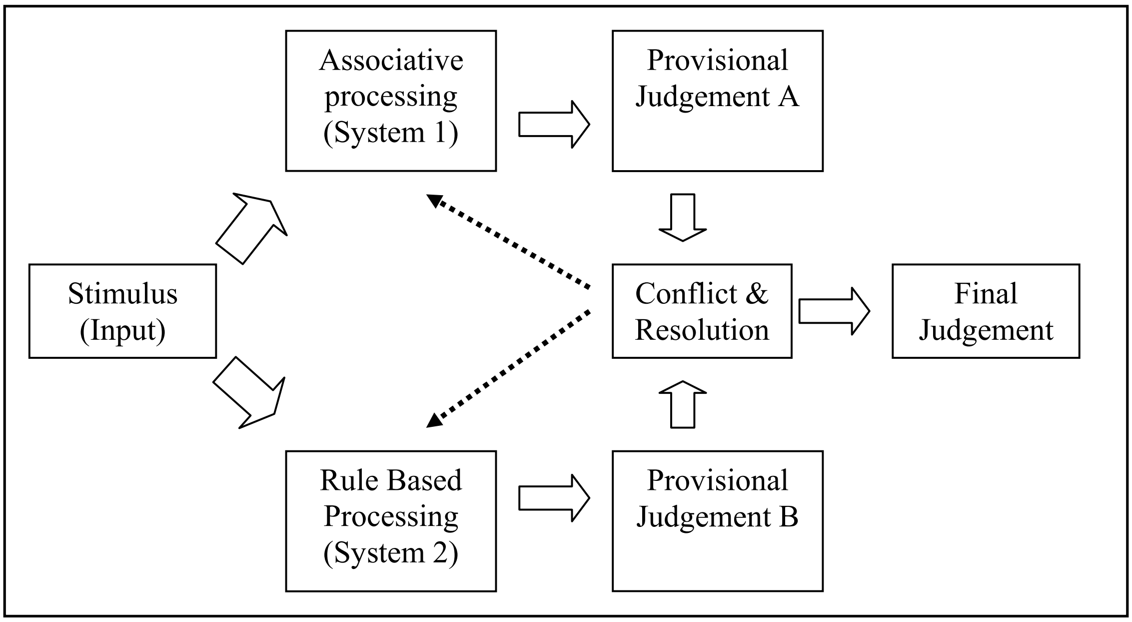
Neither domain-specific reasoning modules nor relevance theory can give a comprehensive account of the mechanisms underlying human reasoning. Dual-process theory is capable of explaining human reasoning across a wider range of situations than either of the two other accounts. It is clear that only dual-process theory has the potential to give a comprehensive account of the mechanisms underlying human reasoning. However, it is equally clear that in order to achieve this it needs to provide a properly articulated account of the interaction between system 1 and system 2 processes.

Section 3- *Models of interaction*

Conflicting Model

The switch between system 1 and system 2 is central to the dual-process account. However, we currently lack a detailed model of the relationship between the two systems. The account favoured by Evans (2003) asserts that the two systems are in conflict and system 2 processes have a limited capacity to inhibit system 1. There is virtually no explanation of the nature of the conflict between the two systems and the mechanisms involved in resolving it. Evans cites Goel and Dolan's (2003) findings that logical correct answers were associated with activity in a different brain area to belief-biased answers, and their suggestion that the prefrontal cortex is vital in detecting and resolving conflict between system 1 and 2 processes as evidence for this conflicting interaction. However, this adds little to our understanding of the interaction between the two systems. There is no detailed account of the factors affecting this switch beyond the evidence that the amount of time given for processing (Evans & Curtis-Holmes, 2005) and the emphasis on the type of reasoning which should be carried out (Evans 2003) both play a role in determining whether the switch from system 1 to system 2 processes occurs. Figure 1, (below) outlines a conflicting model of interaction between system 1 and system 2 processes based on the account offered by Evans (2003).

Figure 1: Dual-process (Conflicting systems) Model



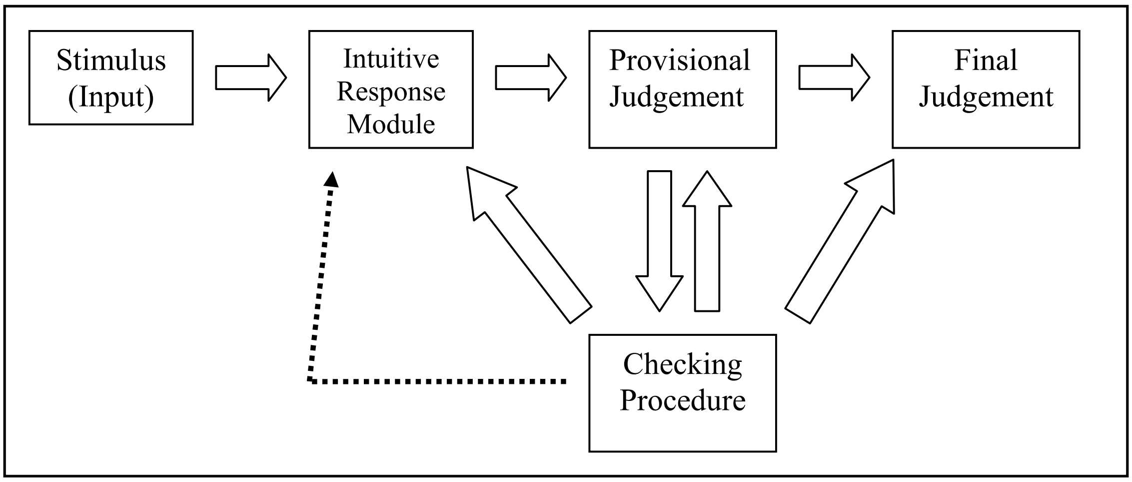
Both systems process the stimulus independently and reach their own provisional judgement. These two provisional judgements are then placed into conflict with each other, competing to become the final judgement and govern the individual's decision. The dashed arrows indicate a potential for learning from previous reasoning outcomes. This would not occur after one decision being reached but if information is processed in a certain way, numerous times, this may affect the way in which it is processed in the future. This would account for our ability to learn to make certain decisions- which used to require long complex calculations- quickly and easily if we are faced with it numerous times. This learning may be associative and does not guarantee that normative logical rules are being internalised. Often provisional judgement A becomes the final judgement without competing with a provisional judgement from system 2. This is when time constraints prevent system 2 reaching a provisional judgement or the limited processing capacity of system 2 is being used to process other information. Most of the explanatory power of this model comes from the conflict and resolution box but there is no detailed explanation of how the processing mechanisms postulated by this box resolve the conflict. It is a lack of explanation in this area that prevents any dual-process account based on this conflicting model of interaction being a complete theory of human reasoning.

Checking Model

The research supporting dual-process theory strongly suggests there are at least two clearly distinct types of processing underlying human reasoning. Dual-process theory makes the natural move of postulating two distinct encapsulated reasoning systems, each using one of these types of processing. However, when the interaction between the two systems is examined, it is not clear this move needs to be made or adds any explanatory power to the account of the mechanisms underlying reasoning. Postulating one reasoning system which incorporates both types of processing could provide an account of the mechanisms underlying human reasoning which fits the empirical data just as well, if not better.

The Complex Cognitive Checking model is my attempt at outlining how one reasoning system can accommodate both types of reasoning process and give an account of the interaction between them. This model is loosely based on Jonathan Haidt's (2001) 'Social Intuitionist' account of moral judgments. Haidt (2001) argued that accounts of moral reasoning have been obsessed with the importance of reason in reaching moral judgements. His research suggests that moral judgements are primarily driven by intuitive emotional responses rather than based on reasoning. Haidt proposed a 'Social Intuitionist' model of moral judgement in which people make an intuitive emotional decision and then engage in motivated reasoning in order to find justification for this judgement. There is also evidence that this kind of post-hoc justification is not just taking place in moral judgement but in all reasoning. Evidence indicates that introspective reports of reasoning are often individuals rationalising their own decisions rather than an accurate reflection of how they come to them (Wason & Evans, 1975; Nisbett & Wilson, 1977). This suggests that a model based on Haidt's explanation of moral judgements could be used to account for reasoning more generally. In such a model the intuitive system 1 processes would generate an initial judgement and the conscious system 2 processes would check for justification of this judgement. Perhaps the operation of system 1 processes should not be regarded as 'reasoning' at all. They are just judgemental responses lacking any justification or 'reasoning' behind them, just like the emotional judgemental responses in the early stage of Haidt's moral judgement system. Figure 2 (below) outlines the Complex Cognitive Checking model, which is a single system account of reasoning where the interaction between conscious model- or rule-based processes and tacit associative processes, is determined by relevance.

Figure 2: The Complex Cognitive Checking model



Information about the stimulus is fed into the relevant intuitive response module where a provisional judgement is made using associative processing. This judgement is then checked by a rule-or model-based system to assess if there are reasons justifying this judgement. A key difference between this model and the Dual-process conflicting systems model, is that in the Checking Model the output of the intuitive system (akin to system 1) is the input for the checking procedure (akin to system 2). If adequate and appropriate reasons can be found for the provisional judgement then it is reinforced and becomes the final judgement. If no justification for the provisional judegment is found then a new provisional judgement is made and checked. This either comes directly from the checking procedure, because the processing makes a certain judgement possible, or the judgement is referred back to a different intuitive response module. If there are time constraints, or the limited capacity of the checking procedure is occupied by other processing, then the provisional judgement becomes the final judgement without being checked. There are also times when the checking procedure is simply not engaged because the information being processed does not have significant relevance. The engagement of the checking procedure is accounted for in terms of relevance theory. This gives relevance theory the central role of explaining the interaction between the two types of processing. The amount of checking each judgement undergoes is determined by considerations of relevance. If the cognitive effort of processing the information is greater than the cognitive effect it produces, then the checking procedure will not be activated at all. Furthermore the rules or models used to check the intuitive judgement are those which are deemed most relevant. This accounts for the large effect that instructions can have on conscious rule- or model-based processing. The dashed arrow represents possible feedback from the checking procedure which over time could lead to learning different intuitive responses to certain stimulus. This accounts for our ability to learn to make rapid responses on complex tasks which we are faced with on numerous occasions. It is unclear exactly how this feedback would work or the number of times it must occur to facilitate rapid responses. However, evidence from variations of the selection task suggests that this feedback is not easily generalised to similar situations and may not constitute associative internalisation of the rules required to reach the normatively correct judgement (Cox & Griggs, 1982 and Cheng et al, 1986).

Comparing the two models

Many of the findings in heuristics and biases research do little to help discriminate between the two models and can be cited as support for both models. This is because it is possible to attribute system 1 reaching a judgement first to either system 2 being the slower system or system 1 providing provisional judgements for system 2 processes to check. However, in the future empirical research should be designed to discriminate between the two models. The key features of the checking model are that it provides an account of the interaction between the two systems based on the principles of relevance theory and that it posits intuitive reasoning (akin to system 1) as the input for the checking procedure (akin to system 2). This already provides some reasons to favour the checking model over the conflicting dual-process model.

The straightforward reason to prefer the complex cognitive checking model over the conflicting dual-process model is that it offers a fuller account of the interaction between the two types of processing. The checking model provides a set of principles that determine the interaction between the two processes, whereas the conflicting model does little to increase our understanding of this interaction despite the fact that it is a central part of the theory. The explanation of the gap between reasoning competence and reasoning performance offered by the conflicting dual-process model runs the risk of simply renaming the phenomena rather than actually giving an account of the mechanisms that give rise to it. This is because the explanation depends on the interaction between the two systems. Without a more detailed account of the mechanisms governing which system will give rise to the final judgement the conflicting dual-process model lacks explanatory power.

Rather than simply attributing variations in performance to different systems the checking account can give a more detailed explanation of why slight variations in the task instructions or surrounding scenario can elicit dramatically different results. This is done by postulating relevance as the cue for not only conscious processing but also the type of rule-or model-based reasoning which is used. The checking account helps solve the problem of how and at what level of processing to apply the principles of relevance theory. Nevertheless, it is still perfectly possible that these principles are in operation at many other levels of processing or even across all cognition.

It would be possible to modify the conflicting dual-process model so that relevance theory determined when system 2 processes inhibited system 1 processes. However such an account would postulate system 2 processes losing the conflict in cases of low relevance. However, there seems to be no reason to engage conscious rule-or model-based processing at all if it is not going to be used or even have any influence. Engaging evolutionarily more advanced processing that uses up valuable space in our working memory to no significant end is not adaptive, and a waste of both attention and cognitive energy. It seems unlikely that this is how the mechanisms underlying human reasoning would work.

Another reason to favour the checking model is that it accounts for the way in which tacit intuitive judgements can direct the attention of conscious processing. In the conflicting model the two are in constant competition, but the checking model gives the intuitive processes the role of guiding what the conscious processes should direct their attention towards. This is a link that early dual-process accounts (Evans & Over, 1996) acknowledged but was neglected by later accounts which focused on the conflict between the two systems. This fits Wason & Evans' (1975) findings that participant's conscious reasoning on the selection task was restricted to thinking about the matching cards, which suggests the focus of conscious processing is directed by pre-attentive tacit processes. Positing tacit processing as the input for conscious judgements also allows the checking model to explain how unconscious motivation can influence conscious decisions. Bargh et al, (2001, Experiment 2) showed that behavioural goals could be induced using priming. They found that activating the goal to co-operate outside of participants' awareness caused them to display the same level of co-operation as those given the conscious goal to co-operate. This is very hard to explain using the conflicting model, yet the checking model can simply attribute this to conscious processes being driven to check for justification of co-operative behaviour, and there being no obvious reason not to co-operate. Furthermore, the checking procedure also makes sense when considering cognitive economy because less processing effort is required to find justification for a certain answer than to consider all the evidence and then make a judgement as to what the answer is.

The checking model gives a better account of why much conscious processing is simply post-hoc justification of intuitive judgments. The conflicting model explains this in terms of system 2 processes being slower than system 1 processes. However, the checking model proposes that conscious processing is checking if there are good reasons behind a given intuitive judgement either to reinforce or replace that judgement. Both accounts are equally satisfactory in cases where an intuitive decision was reached and then individuals are asked for a verbal justification of the decision. However, in cases where there is motivated reasoning, or errors in conscious rule- or model-based reasoning the checking model provides a far more comprehensive explanation. The conflicting model can only explain these phenomena as a failure of system 2 to totally inhibit system 1 processes. Such an explanation is far from satisfactory as it suggests a judgement made with elements of both system 1 and 2 processes, which only further complicates an already under-explained interaction.

The checking model proposes that conscious rule-based reasoning checks for justification of intuitive judgements giving it both a biased starting point and biased motivation. Evidence suggests that much of our everyday cognition is driven by motivations of coherence with our existing intuitions or beliefs. People search exclusively for evidence which supports their view (Perkins et al, 1991) and often stop searching after finding even a single piece of supporting evidence (Perkins, Allen & Hanfer, 1983). This suggests that in many cases intuitive judgements will become actual judgements unless rule- or model-based reasoning comes across a reason to reject the judgement in the process of trying to find a justification for it.

In cases where there is strong motivation such as a firmly held conflicting belief, it is likely the requirements of what constitutes a reason for rejecting the intuitive judgement are made significantly harder to achieve. This fits the model of checking for justification of an inference rather than considering all the factors. It suggests that conscious processing is so heavily directed by receiving its input from tacit processing that it is not entirely free from the effects of a range of biases based on intuition, emotion, belief and ego-defence. This is supported by evidence that instructions to focus on logical necessity or giving participants unlimited time only reduces the prevalence of belief bias rather than eliminating it (Evans, 2000 & Evans & Curtis-Holmes, 2005), which should be the case if system 2 processes are free from the effects of such biases and the instructions cause system 2 processes to inhibit system 1 processes. Dual-process accounts have explained this in terms of individual differences in system 2 reasoning, or system 2 processes having only a limited capacity to inhibit system 1 processes. Given the confusion over the interaction between the two systems this does little to enhance our understanding of the process. However, the explanation offered by the checking model is a uniform and detailed account of why it is so hard to eliminate biases no matter how much time is devoted to learning normative theories of reasoning (Cheng et al, 1986).

If the syllogisms do not conflict with expected truth of the conclusion then either the checking procedure is not activated (causing participants to give the correct answer) or it is directed to find justification for the correct answer which should also result in the correct answer. This explains the high level of performance on the non-conflicting syllogism (Evans, 1983). When the syllogisms do conflict then either the checking procedure is not activated (causing participants to give the incorrect answer) or the checking procedure is directed to find justification for the incorrect answer. Unless the checking procedure is conducted in such a way that it comes across a reason to reject the incorrect answer before it finds even weak justification for it, then the incorrect answer will be given. This explains the prominence of belief bias and how difficult it is to overcome, which is indicated by the large proportion of incorrect answers on conflicting syllogisms in Evans' research (1983). It also explains how cueing certain reasoning strategies can enable some participants to overcome belief bias (Evans, 2000).

Based on the checking model account it is clear that when teaching normative theories of logical reasoning it is important to coach methods that will lead people to finding reasons to reject invalid intuitions before their reasoning comes across even tenuous support for them. Making individuals aware of the biases they are prone to when reasoning about such intuitions may also help reduce the prevalence of these biases.

Conclusion

Evidence from heuristics and biases research indicates a large difference between possible reasoning competence and actual reasoning performance. This has led to an ongoing debate over the nature and fragility of human rationality. Much of this debate turns on whether reasoning errors are caused by the unusual nature of many of the reasoning tasks, or are due to systematic errors in the mechanism underlying human reasoning. A comprehensive account of the mechanisms underlying human reasoning is required to advance the rationality debate. Such an account would also have practical benefits for both predicting, and developing teaching methods for improving, human reasoning.

The evidence from heuristics and biases research provides a strong indication that there are at least two types of processing which account for the difference between possible reasoning competence and actual reasoning performance. Given this the dual-process account is the only theory in a position to be developed into a comprehensive explanation of the psychological mechanisms underlying human reasoning. However, there are many areas in which it lacks sufficient detail, in particular the interaction between system 1 and system 2 processes. It is not clear how the currently favoured conflicting model offers any explanation of the interaction at all. The need for an account to incorporate two distinct types of reasoning process is not on its own justification for postulating two distinct reasoning mechanisms. Evidence from moral reasoning suggests a strong interaction between intuitive and rule- or model-based processes, where the former are directed to find justification for the later (Haidt, 2001). It seems likely that something similar could explain the mechanisms underlying our general reasoning.

This paper proposes the complex cognitive checking model of interaction between the two types of processing, which only postulates one reasoning system. This model gives a better account of some of the psychological research into human reasoning. However, further empirical evidence would be required to differentiate conclusively between the two models and provide a better understanding of the relationship between the two types of processing. There is also room for the development of further models of interaction between the two systems. It remains an open question as to whether this should be understood in terms of two distinct systems or one that incorporates both types of processing. The crucial methodological point is that we need to elaborate models which will enable empirical tests to give us a better articulated account of the complexity (dual or otherwise) of reasoning processes. The checking model is a detailed account of the complex processes underlying human reasoning which seems a plausible alternative to dual-process conflicting systems.

References

Bargh, J.A. et al., 2001, The automated will: Nonconscious activation and pursuit of behavioural goals, *Journal of Personality and Social Psychology*, Vol. 81, Pages: 1014-1027

Chaiken, S and Thorpe, 1999. Dual-process Theories in Social Psychology, Guilford Press.

Cheng, P. W., Holyoak, K. J., Nisbett, R. E., and Oliver, R. M. 1986. Pragmatic versus syntactic approaches to training deductive reasoning. *Cognitive Psychology*, 18, 293–328.

Cohen, L.J. 1981. Can human irrationality be experimentally demonstrated? *Behavioural and Brian Sciences*, 4.

Cohen, L.J. 1982. Are people programmed to commit fallacies? *Journal for the Theory of Social Behaviour, 12.*

Cosmides L. & Tooby J. 1992, Cognitive adaptation for social exchange. In Barkow et al, (1992), The Adapted Mind, Oxford University Press

Cox, J, R. and Griggs, R, A. 1982. The effects of experience on performance in Wason's selection task. *Memory & Cognition*, 10, 496-502

Evans, J. 1977. Towards a statistical theory of reasoning. *Quarterly Journal of Experimental Psychology*, 29, 297-306.

Evans, J. 1982. The Psychology of deductive reasoning. London: Routledge and Kegan Paul.

Evans, J. 1983. On the conflict between logic and belief in syllogistic reasoning. *Memory and Cognition*, 11, 295-306

Evans, J. 1989. Bias in human reasoning: Causes and Consequences. Hove, Uk: Lawrence Erbaum Associates Ltd.

Evans, J. and Over, D. 1996. Rationality and Reasoning. Psychological Press.

Evans, J. 2000. In Mental Models in Reasoning, Cited in Evans, J. 2003. In two minds: dual-process accounts of reasoning. *Trends in Cognitive Sciences*, 7, 10.

Evans, J. Handley S, J. Over D, E. and Perham N. 2002. Background beliefs in Bayesian inferences. *Memory and Cognition*, 30, 179-190

Evans, J. 2003. In two minds: dual-process accounts of reasoning. *Trends in Cognitive Sciences*, 7, 10.

Evans, J. and Curtis-Holmes, J. 2005, *Thinking & Reasoning*, 2005, 11, 4, 382-389.

Geol, V and Dolan, R, J. 2003. Explaining modulation of reasoning by belief. *Cognition*, 87, B11-B22

Haidt, J. 2001. The Emotional Dog and Its Rational Tail: A Social Intuitionist Approach to Moral Judgment. *Psychological Review*, 108, 4, 814-834.

Johnson-Laird, P.N. Bryne, R, M, J. and Schaeken, W. 1994. Why models rather than rules give a better account of propositional reasoning: A reply to Bonatti and O'Brien, Braine and Yang. *Psychological Review, 101, 734-739.*

Klaczynski, P, A. and Daneil, D, B. 2005. Individual differences in conditional reasoning: A dual-process account, *Thinking & Reasoning*, 11, 4, 305-325.

McLeod, P et al. 1998. Introduction To Connectionist Modelling Of Cognitive Processes, Oxford University Press.

Nisbett, R.E. and Borgida, E. 1975. Attribution and the Psychology of prediction. Cited in Botterill, G and Carruthers, P, 1999, *The philosophy of psychology*. Cambridge University Press, page 105.

Nisbett, R, E. and Wilson, T, D. 1977. Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, 84, 231-295

Perkins, D, N. Allen, R. and Hanfer, J. 1983. Difficulties in everyday reasoning. In W. Maxwell (Eds.), Thinking: The frontier expands (pp. 177-189). Hillsdale, NJ: Erlbaum.

Perkins, D, N. Farady, M and Bushey, B. 1991. Everyday reasoning and the roots of intelligence. In J.F. Voss, D.N. Perkins and J.W. Segal (Eds.) Informal reasoning and education (pp. 83-105). Hillsdale, NJ: Erbaum.

Sloman, S.A. 1996, The Empirical case for two systems of reasoning. *Psychological Bulletin*, 119, 3-22. Page 20.

Sperber, D. In Press. Modularity and relevance: How can a massively modular mind be flexible and context-sensitive? To appear in *The Innate Mind: Structure and Content.* Edited by Peter Carruthers, Stephen Laurence, & Stephen Stich .

Sperber, D. and Wilson, D. 1986. *Relevance: Communication and cognition*. (Oxford : Blackwell; Cambridge , Mass. : Harvard University Press).

Sperber D, Cara F & Girotto V. 1995, Relevance Theory Explains the Selection Task. *Cognition*, 57, 31-95.

Stanovich, K, E. and West, R, F. 1998. Cognitive ability and variation in selection task performance. *Thinking & Reasoning*, 4, 193-230.

Stanovich, K, E. and West, R, F. 200. Individual differences in reasoning: Implications for the rationality debate. *Behavioural and Brain Sciences*, 23, 645-726

Stanovich, K, E. 2003. Is probability matching smart? Associations between probabilistic choices and cognitive ability. *Memory and Cognition*, 31, 243-251

Stein, E. 1996. Without Good Reason: The Rationality Debate in Philosophy and Cognitive science. Oxford University Press.

Tversky, A and Kahneman, D, 1983. Extensional versus intuitive reasoning: the conjunction fallacy in probability judgement. *Psychological Review*, 904.

Wason, P, 1968, Reasoning about a Rule, *Quarterly Journal of Experimental Psychology*, 20, pp 273-81

Wason, P,C. and Evans J. 1975. Dual processes in reasoning? *Cognition*, 3, 141-154