Reasoning and moral judgment: A common experimental toolbox

Bastien Trémolière

Université du Québec à Trois-Rivières, Québec, Canada

Wim De Neys

LaPsyDE (CNRS Unit 8240)

Sorbonne - Université Paris Descartes, Paris, France

Jean-François Bonnefon

Center for Research in Management (CNRS Unit 5303)

Université Toulouse 1 Capitole, Toulouse, France

Abstract

The dual-process model offers a unified framework for the study of reasoning and moral judgment, making it easier than ever for reasoning specialists to branch out to moral judgment, of for moral judgment specialists to incorporate findings and methods from the psychology of reasoning. In this chapter, we draw a parallel between investigations of belief bias and investigations of moral dilemmas, to show how the same experimental toolbox (time pressure, cognitive interference, individual differences, mortality salience) was applied to both phenomena, within the theoretical context of the dual-process model.

Reasoning and moral judgment: A common experimental toolbox

The table of contents of the present volume suggests that the psychology of reasoning is expanding at a fast pace, covering grounds which used to be considered the purchase of decision-making, theory of mind, or moral judgment. One reason for this rapid expansion is that the psychology of reasoning has now largely adopted the new lingua franca of higher order cognition: probability, utility, and the dual-process model (Bonnefon, 2013). Couching problems in terms of probability and utility allows the psychology of reasoning to tackle decisional and moral issues, from bets and lotteries to the moral dilemmas we will consider in this chapter. In parallel, relying on the manipulation and measures of the dual-process model insures that results obtained in the field of reasoning are compatible with and relevant for the numerous other fields which have adopted this model as their leading theoretical framework.

Other fields, though, have not always taken full stock of the new shape of the psychology of reasoning. In the field of moral judgment in particular, reasoning is (at worst) considered irrelevant, or (at best) given a definition that is not quite the same as in the psychology of reasoning. Jonathan Haidt, for example, famously asserted that our moral positions do not usually stem from reasoning. Moral positions, according to Haidt (2001, 2007, 2012), are first and foremost driven by intuitions, and reasoning is only used later, to find reasons that might sway others' intuitions in favor of our position.

The dual-process model of moral judgment, developed by Joshua Greene and collaborators (starting with Greene, Sommerville, Nystrom, Darley, & Cohen, 2001), gives a larger role to reasoning, albeit one that depends on a restricted definition. In the dual-process model of moral judgment, intuition and reasoning concurrently drive moral positions. Reasoning corresponds to the conscious application of a decision rule (Greene, 2013), and moral reasoning compels one to a conclusion on pain of inconsistency with one's commitments to one or several moral decision rules (Paxton & Greene, 2010).

Accordingly, it would appear that reasoning in the dual-process model of moral judgment is solely a conscious, rule-based affair. In other words, the dual-process model of

moral judgment defines reasoning as its deliberative component. This can be a source of confusion, because specialists of reasoning tend to adopt a broader definition (Evans, 2008; Evans & Stanovich, 2013; Sloman, 1996). Specialists of reasoning agree that the conscious application of rules, and the commitment to rational axioms such as consistency, are indeed a hallmark of deliberate, reflective reasoning, also known as System 2 reasoning. However, they typically retain the term 'reasoning' (only qualified as intuitive, heuristic, or System 1), for the unconscious processing of beliefs driven by affective and heuristic shortcuts.

With this caveat in mind, the widespread adoption of the dual-process model by the fields of reasoning and moral-judgment offers novel opportunities for them to connect. In particular, the experimental protocols used in the two fields are now largely comparable, because they utilize the same basic toolbox provided by the dual-process model. In the rest of this chapter, we will illustrate this novel interoperability by showing how the dual-process toolbox was applied in parallel to the flagship problems of the two fields: belief bias and utilitarian dilemmas. We begin with a quick overview of the dual-process toolbox, that is, the list of manipulations and measures that researchers typically use to test predictions derived from the dual process model. We then introduce belief bias and utilitarian dilemmas, and show that they delivered very similar findings in experiments that used five dual-process tools (time pressure, cognitive interference, individual differences – and mortality salience, which will require some additional explaining). Note that due to space limitations, we focus on behavioral tools in the dual-process toolbox, to the exclusion of methods aimed at identifying the neural correlates of belief bias (De Neys, Vartanian, & Goel, 2008; Goel, 2007; Goel & Dolan, 2003; Rotello & Heit, in press; Tsujii, Masuda, Akiyama, & Watanabe, 2010; Tsujii, Okada, & Watanabe, 2010; Tsujii & Watanabe, 2009, 2010) and utilitarian dilemmas (Greene, 2007; Greene, Nystrom, Engell, Darley, & Cohen, 2004; Greene et al., 2001; Kahane et al., 2012; Koenigs et al., 2007; Shenhav & Greene, 2014; Thomas, Croft, & Tranel, 2011).

The dual-process toolbox

Although there are different versions of the dual-process model (De Neys & Bonnefon, 2013; Evans, 2008; Evans & Stanovich, 2013; Kahneman, 2011), all variants appear to share the same assumption, broadly formulated: System 2 processing requires access to central executive resources in order to suspend or inhibit the intuitive response to a problem, for the time it takes to produce a reflective response. This assumption is critically important because it applies the same whatever the field one is investigating, and allows for predictions which are tested the same, whatever the field one is investigating.

Let us consider a problem which can elicits one of two responses, r and R. The problem can be one of logical reasoning, in which participants must decide whether a conclusion is valid or invalid – or the problem can be one of moral judgment, in which participants must decide whether an action is morally acceptable or unacceptable. Let us now imagine that we want to apply the dual-process model to the problem, and test the hypothesis that r and R are intuitive and reflective responses, respectively.

Because the reflective response is elaborated after the intuitive response is generated and inhibited, a first testable prediction is that R will take longer to produce than r. Even better, we can run a causal test by limiting the time that some participants have to solve the problem, and check whether this **time pressure** decreases the frequency of R and increases the frequency of r. Additionally, because the elaboration of the reflective response requires central executive resources which cannot be deployed for several tasks at once, another testable prediction is that **cognitive interference,** in the form of a concurrent task, should decrease the frequency of R and increase the frequency of r.

Given that the elaboration of the reflective response requires to mobilize and expend central executive resources, a third testable prediction is that individuals who either possess these resources in larger amounts, or have a greater individual propensity to mobilize them, should be more likely to produce the reflective response. Accordingly, these **individual differences** should correlate positively with the likelihood of producing R, and negatively with

the likelihood of producing r. Finally, any experimental manipulation whose effect is to limit, divert or discourage the use of central executive resources should by the same logic decrease the frequency of R and increase the frequency of r. For example, participants may receive instructions encouraging them to trust or not trust their gut feelings, or they may be subjected to priming manipulations to the same effect. In this chapter, we will consider the effect of the **mortality salience** manipulation, which consists of asking subjects to write a very short essay on their future death. Importantly, this reminder of mortality is believed to temporarily decrease the availability of central executive resources. As a consequence, conducting the mortality salience manipulation before the target problem should decrease the frequency of R and increase the frequency of r.

In the rest of this chapter, we will review the use of these four dual-process tools, as applied to one problem of logical reasoning (belief bias) and one problem of moral judgment (utilitarian dilemmas). Our aim is to show that the reasoning and the moral literature are much closer than one might think, in the sense that the two fields have applied this exact same toolbox to the two problems, in independent and parallel fashion, but to mostly comparable results. Before we move on to this review, we briefly introduce our two problems in the next section.

Belief bias and moral dilemmas

Reasoners display **belief bias** when their judgment about the validity of a conclusion is unduly influenced by the believability of this conclusion (Evans, Barston, & Pollard, 1983; Klauer, Musch, & Naumer, 2000; Thompson & Evans, 2012). For example, the fact that we know roses to be flowers can make it hard to spot that the conclusion of syllogism (1) is logically invalid:

(1) a. All flowers need water.
    b. Roses need water.
    c. Therefore, roses are flowers.

To accept (1-c) as a valid conclusion is to fall prey to belief bias. Likewise, to reject (2-c) as an invalid conclusion is to fall prey to belief bias, since this conclusion is actually valid, however inconsistent it is with our common beliefs about astronauts:

(2) a. People in good health are happy.
b. Some astronauts are unhappy.
c. Therefore, some astronauts are not in good health.

In examples (1) and (2) above, the believability of the conclusion is incongruent with its logical validity. We will call these the 'conflict' syllogisms, as opposed to the 'control' syllogisms, in which a response based on believability would be the same as a response based on logical validity. In conflict syllogisms, one gives a different response as a function of whether one relies on believability or validity. Let us call the former response r (based on believability), and the latter response R (based on validity). The data we will review in the rest of this chapter will test the hypothesis that r is the intuitive, System 1 response; and R is the reflective, System 2 response.

We are well aware that specialists of belief bias would use more complex indices (Heit & Rotello, 2014; Trippas, Handley, & Verde, 2014), more sophisticated analyses (Dube, Rotello, & Heit, 2010), and more creative protocols (Trippas, Verde, & Handley, 2014), than that we just introduced. This sophistication has allowed for a much deeper understanding of belief bias, but it comes at a cost, as it makes it harder for researchers in other fields to capture the gist of belief bias research and its relevance for their own work. In this chapter, we strive to speak to moral researchers as well as reasoning researchers, and we limit our exposition of belief bias accordingly. This is also true of our introduction of moral dilemmas, which we attempt to keep reasonably free of complications, in order to speak to reasoning researchers as well as moral researchers.

The **moral dilemmas** that have generated the most research, from a dual-process perspective, describe situations where rights and duties are in conflict with the greater good.

In these situations, one can adopt a deontic response (giving priority to personal rights over the greater good) or a utilitarian response (giving priority to the greater good over personal rights). Most typically, the dilemmas describe fictional circumstances in which one has the possibility to save several lives by killing a single person. The deontic response is to find this course of action morally unacceptable (the personal right to not be killed trumps the greater good, i.e., the net number of survivors), whereas the utilitarian response is to find this course of action acceptable (the greater good, i.e., the net number of survivors, trumps the personal right to not be killed). As an illustration, we quote in full three dilemmas used in many research papers with small variations (especially in the exact phrasing of the question asked to participants at the end of the dilemma):

(3) [Footbridge.] A runaway trolley is heading down the tracks toward 10 workmen who will be killed if the trolley proceeds on its present course. You are on a footbridge over the tracks, in between the approaching trolley and the 10 workmen. Next to you on this footbridge is a stranger who happens to be very large. The only way to save the lives of the 10 workmen is to push this stranger off the bridge and onto the tracks below where his large body will stop the trolley. The stranger will die if you do this, but the 10 workmen will be saved. Would it be wrong for you to push the stranger on to the tracks in order to save the 10 workmen?

(4) [Sophie's choice.] It is wartime and you and your two children, ages eight and five, are living in a territory that has been occupied by the enemy. At the enemy's headquarters is a doctor who performs painful experiments on humans that inevitably lead to death. He intends to perform experiments on one of your children, but he will allow you to choose which of your children will be experimented upon. You have 24 hours to bring one of your children to his laboratory. If you refuse to bring one of your children to his laboratory he will find them both and experiment on both of them. Would it be wrong for you to bring one of your children to the laboratory in order to avoid having them

both die?

(5) [Crying baby.] Enemy soldiers have taken over your village. They have orders to kill all remaining civilians. You and some of your townspeople have sought refuge in the cellar of a large house. Outside you hear the voices of soldiers who have come to search the house for valuables. Your baby begins to cry loudly. You cover his mouth to block the sound. If you remove your hand from his mouth his crying will summon the attention of the soldiers who will kill you, your child, and the others hiding out in the cellar. To save yourself and the others you must smother your child to death. Would it be wrong for you to smother your child in order to save yourself and the other townspeople?

Agreeing to push the stranger on to the tracks, to bring one's own children to the laboratory, and to smother one's own baby – or at least not finding these actions wrong – are utilitarian responses to the dilemmas. Let us call R the utilitarian response, and r the deontic response. The data we will review in the rest of this chapter will test the hypothesis that r (deontic) is the intuitive, System 1 response, whereas R (utilitarian) is the reflective, System 2 response.

We are now in a position to review the experiments that attempted (on the one hand) to test the hypothesis that responding on the basis of believability in the 'conflict' type of belief bias problems is an intuitive, System 1 response; or (on the other hand) to test the hypothesis that giving a deontic response to utilitarian dilemmas is an intuitive, System 1 response. More importantly, we will show that these two hypotheses were tested by utilizing the exact same dual-process toolbox based on time pressure, cognitive interference, individual differences, and (for something a bit different) mortality salience. We now address these tools one by one.

Time pressure

Our focus in this section will be on time pressure manipulations (i.e., the effect of imposing a time limit to respond to logical syllogisms or moral dilemmas). Before we consider the results of time pressure studies, though, we review some results related to the processing time of syllogisms and moral dilemmas. Indeed, the dual-process model makes two predictions about processing time. First, people should take longer to solve a conflict problem (in which the intuitive response is supposedly different from the reflective response) than a control problem (in which the intuitive and reflective responses are presumed to be the same). This prediction assumes that, for conflict problems, at least some individuals will take the necessary time to inhibit their intuitive response and generate a reflective response. Second, individuals who give what is presumed to be the reflective response should take longer than individuals who settle for the intuitive response. As it turns out, studies of belief bias have focused on the first prediction, whereas studies on moral dilemma have focused on the second.

Several studies on belief bias have shown that reasoners took longer to inspect or to process conflict syllogisms, in which the conclusion based on believability is different from the response based on logical validity (Ball, Phillips, Wade, & Quayle, 2006; De Neys & Franssens, 2009; Stupple & Ball, 2008; Svedholm-Häkkinen, 2015). These studies have quite consistently recorded a 15 to 30% increase in processing time for conflict syllogisms. These results are consistent with the dual-process model, but they cannot on their own tell us which of the believability or validity response is the reflective one. In that regard it would be informative to compare processing time for individuals who eventually give the believability response, to processing time for individuals who eventually give the validity response (and ideally contrast this with their processing time on no-conflict problems, e.g., De Neys & Dieussaert, 2005). Interestingly, this is the strategy that researchers have adopted with respect to the processing of utilitarian moral dilemmas.

In a typical utilitarian moral dilemma, one can generate a deontic response (i.e., it is

morally unacceptable to kill, even if killing one person would save several lives) or a utilitarian response (i.e., it is morally acceptable to kill if the death of one would save several other lives). If one of these two responses is intuitive and the other is reflective, we should expect a difference in processing time as a function of whether an individual eventually gives one or the other. Data are hardly conclusive in that regard, though. Greene et al. (2001) observed that processing time was greater for utilitarian responses, but a re-analysis revealed that this effect was driven by a subset of scenarios, not all of which had desirable properties (McGuire, Langdon, Coltheart, & Mackenzie, 2009). The response time effect was shown again to be fragile in two other articles. In one case, it was limited to a condition in which participants experienced cognitive interference (Greene, Morelli, Lowenberg, Nystrom, & Cohen, 2008); and in the other case, it was again contingent on a limited subset of scenarios (Koop, 2013). For the time being, we must conclude that no solid data exist that would show a difference in processing time between utilitarian and deontic responses.

The lack of correlational data between type of response and processing time should not be counted as a strong argument against the dual-process model, though. Solving syllogisms or moral dilemmas takes a long time, and it is entirely possible that people do not respond as soon as they can. Accordingly, response times in such situations must be considered a noisy measure at best. A much better test of the dual-process model consists of using time pressure, that is, giving two groups of participants a different time limit to give their response. The logic here is that participants working under severe time constraints are much less likely to have the time to generate a reflective response – therefore, imposing a sever time constraints should decrease the frequency of that of the two responses which requires reflective processing.

In studies of belief bias, the time limit imposed to participants under time pressure is commonly set at 10 seconds. That is, participants under time pressure have 10 seconds to read the syllogism and assess whether the proposed conclusion is logically valid. Depending on the study, participants in the control group are given either unlimited time, or about 20

seconds, which corresponds to the average amount of time people take to respond under free-time conditions. In the first application of the time pressure protocol to belief bias, Evans and Curtis-Holmes (2005) observed that the frequency of the validity response (as opposed to the believability response) decreased by about 20 percentage points when participants were given 10 seconds to solve the syllogisms. This result was replicated, for a similar effect size, by Tsujii and Watanabe (2010), who used a 10-second limit for the time pressure group and a 20-second limit for the control group. The same effect size was observed again by Trippas, Handley, and Verde (2013), albeit for only one of the two types of conflict syllogisms. The frequency of the validity response decreased by about 20 percentage points under time pressure for syllogisms in which the conclusion was unbelievable and yet valid. For the other type of conflict syllogism (invalid yet believable conclusion), the effect was weaker (less than 10 percentage points), and only observed for participants in the lower range of cognitive ability. With this proviso, it seems reasonably clear that time pressure decreases the frequency of the validity response to conflict syllogism, which the dual-process model counts as evidence that the validity response is the reflective, System 1 response.

Turning now to moral dilemmas, we would expect that a time pressure manipulation would decrease the frequency of utilitarian responses, if the utilitarian response is indeed the reflective, System 2 response. Just as with belief bias, this prediction was tested and validated in several studies.

Suter and Hertwig (2011) conducted the first study applying time pressure to moral dilemmas. In their study, participants first read the contents of a dilemma (e.g., Crying Baby, see above) for a maximum of 35 seconds, and then moved on to a second screen displaying the moral acceptability question (e.g., ‘Is it appropriate for you to smother your child in order to save yourself and the other townspeople?’). Participants in the time pressure condition had 8 seconds to respond, whereas participants in the control condition had to wait 180 seconds before responding. The time pressure manipulation had the expected effect on the frequency of utilitarian responses to three high-conflict dilemmas such as Crying Baby, decreasing it by

about 20 percentage points. Trémolière and Bonnefon (2014) replicated this effect on a new selection of four dilemmas, in which the frequency of utilitarian responses decrease by an average of 15 to 20 percentage points under time pressure – at least for versions of the dilemmas in which a relatively small number of lives could be saved (5 to 50). The time pressure effect disappeared, though, when a large number of lives could be saved (500–5000) suggesting that utilitarian killings were no longer counterintuitive in these situations.

These finding were consistent with the hypothesis that utilitarian responses were reflective, System 2 outputs (at least when the number of lives saved by a utilitarian killings was relatively small). Later studies consolidated this conclusion by using more tightly controlled protocols and a greater range of dilemmas. In the original study of Suter and Hertwig (2011), giving the utilitarian response always amounted to choose the 'yes' option. Furthermore, it was plausible that participants used some of their 35s reading time to start thinking about their decision. To control for these two potential biases, Cummins and Cummins (2012) conducted a study in which the utilitarian response alternated between 'yes' and 'no', participants had to scroll down to read the scenario, and decision time in the pressure condition was set at 200ms for each word in the decision question. The frequency of utilitarian responses dropped by 10 to 15 percentage points on average, for all dilemmas but those few in which extremely few people make the utilitarian decision in the first place.

In sum, time pressure manipulations have been used in parallel in investigation of belief bias and moral dilemmas. In both fields, the disappearance of one type of response under time pressure (the validity response for belief bias, the utilitarian response for moral dilemmas) was taken as evidence that this response was the product of reflective, System 2 processing. This time pressure criterion is only one tool in the dual-process toolbox, though – and we will now turn to a second tool, the use of cognitive interference protocols.

Cognitive interference

The dual-process model assumes that reflective, System 2 processing requires the mobilization of central executive resources. When two tasks simultaneously require these resources, performance is expected to decrease on at least one of the tasks. Accordingly, the dual-process model predicts that cognitive interference, in the form of a concurrent task, will decrease the frequency of the reflective, System 2 response to a problem. This prediction applies the same to belief bias and moral dilemmas – cognitive interference should decrease the frequency of the validity response to conflict syllogisms, and decrease the frequency of the utilitarian response to moral dilemmas. And indeed, specialists of belief bias and moral dilemmas have applied this logic in parallel to their respective fields, to very comparable results.

Unlike manipulations of time pressure, which are highly similar across studies and across the two fields, manipulations of cognitive interference are seldom similar from one study to another. The principle is always the same: participants are required to devote the lion's share of their attention to an interfering task which is meant to keep their working memory busy, while processing logical syllogisms or moral dilemmas. As we will see, though, the interfering task can be as varied as counting target words while listening to a song, memorizing patterns of dots, taking a 2-back task, or monitoring strings of numbers scrolling across a computer screen.

In the first article to apply a cognitive interference manipulation to belief bias (De Neys, 2006), participants were assigned to one of three conditions. In the control condition (no interference), participants simply solved conflict syllogisms. In the low and high interference conditions, a 3×3 matrix appeared for 850ms before each syllogism. Some cells of the matrix contained dots, whose location participants had to memorize. Once they had solved the syllogism, participants had to correctly place the dots in an empty matrix. The low interference condition utilized easy configurations of dots (horizontal or vertical), whereas the high interference condition utilized complex patterns of dots (Figure 1). Keeping in mind these

complex patterns put a high load on executive resources (Miyake, Friedman, Rettinger, Shah, & Hegarty, 2001), whereas vertical or horizontal patterns puts only a low load on executive resources.

As expected, cognitive interference decreased the frequency of the validity response to conflict syllogisms, from about 70% under no interference, to about 60% under low interference, to about 50% under high interference. This finding was replicated by Trémolière, De Neys, and Bonnefon (2014), who observed the same drop from about 60% validity responses to conflict problems under low interference, to 50% under high interference.

Other studies consolidated this finding using other manipulations of cognitive interference. In one case, cognitive interference was manipulated by having participants listen to a song while solving the problems, and counting how many times the word ‘time’ occurred in the song (DeWall, Baumeister, & Masicampo, 2008). The word ‘times’ occurred 16 times in 5 minutes and 37 seconds, or once every 21 second in average, ensuring that this interfering task required sustained attention. In one other case (Tsujii & Watanabe, 2009), cognitive interference was manipulated by contrasting a 0-back task in the control condition to a 2-back task in the interference condition.

Convergent findings thus suggest that the validity response to a conflict syllogism decreases in frequency when reasoners experience cognitive interference. Under the dual-process framework, this counts as evidence that the validity response requires reflective, System 2 processing. In parallel, moral scientists used comparable manipulations to establish that the utilitarian response to a moral dilemma was similarly impaired by cognitive interference.

The first published attempt to demonstrate an effect of cognitive interference on moral dilemmas led to mitigated results (Greene et al., 2008). While participants in the control condition simply solve moral dilemmas such as Crying Baby, participants in the interference condition had to simultaneously monitor a stream of digits scrolling at the bottom of the screen, and hit a button every time they detected the number 5 (which accounted for 20% of

the digits). This manipulation had no effect on the frequency of utilitarian responses – however, participants in the interference condition took longer to deliver utilitarian responses, compared to participants in the control condition.

The fact that the frequency of utilitarian responses was not impacted by the interference manipulation was unexpected under the dual-process model. Greene et al. (2008) speculated that participants might be keenly aware of the interference manipulation and decided to push through it by making extra cognitive effort. Accordingly, they would reach utilitarian response the same, only after further effort, which would account for their longer processing time.

Another possibility is that the interference manipulation was simply not powerful enough – accordingly, it would have delayed but not prevented participants to reach utilitarian responses. Trémolière, De Neys, and Bonnefon (2012) tested this hypothesis by using several versions of the dot matrix task (low load, high load, and extreme load, see Figure 1) as their interference manipulation. They observed that the frequency of utilitarian responses was the same (about 55%) in the low and high load conditions, but decreased to less than 40% in the extreme load condition. This result was replicated in another study (Trémolière & Bonnefon, 2014), in which the frequency of utilitarian responses dropped from about 55% in a low load condition, to about 45% in an extreme load condition (see also Conway & Gawronski, 2013).

In sum, utilitarian responses do disappear under cognitive interference, but the level of cognitive load required to demonstrate this effect is larger than what it is in belief bias studies. The fact that the dot matrix task has been used in both domains allows for a direct comparison of the cognitive load required to affect belief bias, and the cognitive load required to affect moral dilemmas. As it turns out, giving an utilitarian response to a moral dilemma requires executive resources, but to a lesser degree than giving a validity response to a conflict syllogism.

Individual differences

The dual-process model assumes that the elaboration of reflective, System 2 responses requires reasoners to mobilize central executive resources. Accordingly, the model predicts that reflective, System 2 responses are more likely to be elaborated by reasoners who possess these resources in greater amount, or have a greater disposition to mobilize them. The model therefore gives an important role to cognitive ability and thinking dispositions as individual moderators of the likelihood to elaborate a reflective response. Interestingly, studies of moral dilemmas have mostly investigated the role of thinking dispositions, whereas studies of belief bias have given equal measures of interest to cognitive ability and thinking dispositions.

In studies of belief bias, cognitive ability has been indexed through performance on a broad array of tasks – for example, working memory span (De Neys, 2006), inhibitory control (Handley, Capon, Beveridge, Dennis, & Evans, 2004), standardized tests of intelligence (Kokis, MacPherson, Toplak, West, & Stanovich, 2002; MacPherson & Stanovich, 2007; Newstead, Handley, Harley, Wright, & Farelly, 2004; Sa, West, & Stanovich, 1999), or the SAT score for samples of American students (Stanovich & West, 1998; West, Toplak, & Stanovich, 2008). These measures commonly account for 15–20% of the variance in the likelihood of giving validity responses to conflict syllogisms.

Studies of moral dilemmas are far behind studies of belief bias with respect to investigating the moderating role of cognitive ability. In fact, we are only aware of one published article in which this question was addressed, to inconclusive results (Moore, Clark, & Kane, 2008). In this study, cognitive ability was indexed by performance on three working memory span measures. Participants with higher working memory capacity found utilitarian killings more appropriate than participants with lower working memory capacity (in line with the dual process model), but only for the subset of problems in which the individual to be killed would die in any case (think of the Sophie's Choice problem above, in which both children will die if no action is taken – compared to the Footbridge problem, in which the

stranger will not die if no action is taken). It is intriguing that the issue was not pushed any further, especially in comparison to the vast body of work on cognitive ability in the belief bias domain. Maybe the effect of cognitive ability is simply harder to capture in the moral domain. We have already pointed to some evidence that utilitarian dilemmas might place lower demands on working memory than conflict syllogisms do (see the Cognitive Interference section). If this is correct, then perhaps differences in cognitive ability account for a much smaller share of variance in the moral domain, as compared to the belief bias domain.

Both fields, though, have given about equal attention to the moderating role of thinking dispositions (i.e., the propensity to mobilize executive resources, rather than the amount in which they are available). A closer look will nevertheless reveal different practices in how these disposition are measured. In studies of belief bias, thinking dispositions are commonly measured through self-reports. The two most frequent tools for doing so are the Actively Open Minded Thinking scale (Stanovich & West, 1997), either on is own or as part of a broader thinking disposition composite, and the Need for Cognition scale (Cacioppo & Petty, 1982), either on its own or as part of the Rational-Experiential Inventory (Paccini & Epstein, 1999). In contrast, studies of moral dilemmas most commonly use the performance measure known as the Cognitive Reflection Test (Fredericks, 2005).

The Actively Open Minded Thinking scale tracks a distinctive style of cognitive regulation, aimed at not systematically strengthening whatever beliefs one already holds. High scores on the scale correspond to high agreement with statements such as ‘People should always take into consideration evidence that goes against their beliefs,’ and low agreement with statements such as ‘No one can talk me out of something I know is right.’ Higher scores on this scale correlate with a higher propensity to give validity responses to logical syllogisms, accounting for 5–15% of the variance (Kokis et al., 2002; MacPherson & Stanovich, 2007; Sa et al., 1999; Stanovich & West, 1998). Need for Cognition is less consistently linked to validity responses. The correlation seems to be .20 at the highest (MacPherson & Stanovich, 2007), and at least two studies found no correlation at all between

validity responses and Need for Cognition (Kokis et al., 2002; Newstead et al., 2004).

Need for Cognition seems to do slightly better in the context of moral dilemmas, with reported correlations of .20 and .40 (Bartels, 2008; Conway & Gawronski, 2013). Need for Cognition is not, however, the most common measure of thinking dispositions in this context. Moral researchers focus instead on a performance (rather than self-reported) measure of thinking disposition, the Cognitive Reflection Test (CRT; Fredericks, 2005). The CRT consists of three brain teasers with a very compelling incorrect response, and a very simple correct response (in the sense that it only requires the most basic arithmetic skills). Performance on the CRT requires one to ignore the compelling but incorrect response that comes first to mind, in order to do the simple calculation that will deliver the correct response. For example, the lily pad item reads: 'In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake?'. The correct response (47 days) can be computed simply enough, if only one ignores for a moment the compelling but incorrect response to the problem (24 days).

There seems to be a detectable correlation between the performance on the CRT and the likelihood of giving a utilitarian response to a moral dilemma, although the value of this correlation seems to be quite different if the CRT is given before (.25–.60) or after (.00–.25) the moral dilemma (Paxton, Ungar, & Greene, 2012; Paxton, Bruni, & Greene, 2014). The link between CRT and utilitarian judgment is not quite robust in any case. Some authors have recently hypothesized that high CRT performers may actually consider that utilitarian killings are no longer a moral issue, but a personal choice – which would explain why they may or may not endorse them, depending on the occasion (Royzman, Landy, & Leeman, in press). Another intriguing possibility, especially in light of the belief bias studies we just reviewed, is that the CRT might predict utilitarian judgments only to the extent that it overlaps with Actively Open Minded Thinking (Baron, Scott, Fincher, & Metz, in press). If this claim is correct, then Actively Open Minded Thinking might be the only genuine and robust individual moderator

identified so far as being at work in the two fields.

Mortality salience

Most generally, the dual-process model predicts that any experimental manipulation that encourages or discourages people to mobilize executive resources, will translate to higher or lower frequencies of reflective System 2 responses, respectively. Many such manipulations exist, which range from explicit instructions to reflect, to implicit priming of reflection. We conclude our survey with one such manipulation, mortality salience. The mortality salience manipulation consists of asking participants to write a few lines in response to the following prompt, before the main task of the experiment:

(6) Briefly describe the emotions that the thought of your own death arouse in you. Jot down, as specifically as you can, what you think will happen as you physically die and once you are physically dead.

Terror management Theory (Greenberg, Pyszczynski, & Solomon, 1986) assumes that such reminders of our future, inescapable death trigger two waves of cognitive defense mechanisms aimed at protecting us against debilitating anxiety (Burke, Martens, & Faucher, 2010; Pyszczynski, Greenberg, & Solomon, 1999). The first wave is the most interesting for our current purposes. Indeed, Terror Management Theory assumes that immediately after the mortality salience manipulation, participants enter a 5–10 minute phase during which their executive resources are mobilized in order to displace thoughts of death outside of conscious attention. Accordingly, this manipulation should decrease the frequency of reflective System 2 responses to both conflict syllogisms and moral dilemmas, at least for a window of 5 to 10 minutes.

These two predictions were tested and validated in two research articles. In a study of belief bias (Trémolière et al., 2014), participants were assigned either to the mortality salience group, or to a control group in which participants wrote a few lines following a slightly

different prompt:

(7) Briefly describe the emotions that the thought of extreme pain arouse in you. Jot down, as specifically as you can, what you think will happen as you endure extreme pain and once you have endured extreme pain.

This is a typical control condition for mortality salience studies, whose purpose is to ensure that the effect of mortality salience is not simply due to negative affect. After this initial writing stage, participants solved the syllogisms typical of belief bias research. As predicted by the dual-process model, the frequency of validity responses dropped by 13 percentage points for participants in the mortality salience condition, compared to participants in the pain condition.

In another study (Trémolière et al., 2012), participants were similarly assigned to either one of the mortality salience or the pain group, and moved on to solving moral dilemmas. Once more, in line with the prediction of the dual process model, the frequency of utilitarian responses dropped by 23 percentage points in the mortality salience condition. This is a huge drop, compared to the effect of common cognitive interference manipulations – in fact, a second experiment showed the effect of mortality salience to be higher than that of the extreme load manipulation displayed in Figure 1.

In sum, the mortality salience manipulation (which was assumed to mobilize executive resources and impair reflective thinking) was shown to decrease the frequency of reflective responses in both the belief bias and the moral domain. The effect sizes observed in these studies suggested that thinking about death was severely impairing reflective thinking. Indeed, a drop of 13 percentage point for belief bias is higher than the usual effect of cognitive interference (commonly about 10 percentage points), and a drop of 23 percentage points for moral dilemmas is unheard of, even under extreme load conditions.

Conclusion

The dual-process model has become a central theoretical framework both for the psychology of reasoning and for the psychology of moral judgment. As a consequence, it has become far easier to make connections between two fields that grew mostly separate for several decades. Most importantly, the two fields now share a common experimental toolbox, inspired by the dual-process model. In particular, both fields rely on manipulations of time pressure and cognitive interference, and both fields investigate broadly the same array of individual moderators. As a result, findings obtained within one field can serve to refine and inspire work in the other field.

Consider for example the case of cognitive interference manipulations. Even though these manipulations can utilize many different protocols, enough studies in the two fields are available to conclude that manipulations which are sufficient to impact belief bias, are not powerful enough to impact moral judgment. Based on such results, we can gain an incrementally parametric perspective on the amount of cognitive resources which are required in the two domains.

Consider as another example the case of individual moderators. Whereas a vast number of studies identified the role of cognitive ability in belief bias, the field of moral judgment is still very much in a state of infancy in that respect. Thinking dispositions, on the other hand, have been largely studied in both fields, albeit with a different approach to measurement. Whereas reasoning studies have largely focused on self-reports of thinking dispositions such as Actively open Minded Thinking, moral studies focused on performance measures such as the Cognitive Reflection Test. The two fields have much to gain in pooling results and systematically comparing the share of variance explained by these different moderators, in their tasks of interest.

Finally, we wish to emphasize that the toolbox is still opened and waiting for new tools. Every time a new manipulation is introduced in either one of the two fields, it becomes relevant to researchers in the other field. Once more, as we illustrated with the case of the

mortality salience manipulation, the two fields have much to gain in monitoring the novel manipulations developed in either one, in order to systematically compare and contrast their effects in the other. The unification of the two fields, once barely conceivable, is now a very reasonable prospect, within the framework of the dual-process model.

References

Ball, L. J., Phillips, P., Wade, C. N., & Quayle, J. D. (2006). Effects of belief and logic on syllogistic reasoning: Eye-movement evidence for selective processing models. *Experimental Psychology*, *53*, 77–86.

Baron, J., Scott, S., Fincher, K., & Metz, S. E. (in press). Why does the Cognitive Reflection Test (sometimes) predict utilitarian moral judgment (and other things)? *Journal of Applied Research in Memory and Cognition*.

Bartels, D. M. (2008). Principled moral sentiment and the flexibility of moral judgment and decision making. *Cognition*, *108*, 381–417.

Bonnefon, J. F. (2013). New ambitions for a new paradigm: Putting the psychology of reasoning at the service of humanity. *Thinking & Reasoning*, *19*, 381–398.

Burke, B. L., Martens, A., & Faucher, E. H. (2010). Two decades of terror management theory: a meta-analysis of mortality salience research. *Personality and Social Psychology Review*, *14*, 155–195.

Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition. *Journal of Personality and Social Psychology*, *42*, 116–131.

Conway, P., & Gawronski, B. (2013). Deontological and utilitarian inclinations in moral decision making: A process dissociation approach. *Journal of Personality and Social Psychology*, *104*, 216–235.

Cummins, D. D., & Cummins, R. C. (2012). Emotion and deliberative reasoning in moral judgment. *Frontiers in Psychology*, *3*, 328.

De Neys, W. (2006). Dual processing in reasoning: Two systems but one reasoner. *Psychological Science*, *17*, 428–433.

De Neys, W., & Bonnefon, J. F. (2013). The whys and whens of individual differences in thinking biases. *Trends in Cognitive Sciences*, *17*, 172-178.

De Neys, W., & Dieussaert, K. (2005). Individual differences in rational thinking time. In *Proceedings of the annual conference of the cognitive science society* (Vol. 27, pp.

577–582).

De Neys, W., & Franssens, S. (2009). Belief inhibition during thinking: Not always winning but at least taking part. *Cognition*, *113*, 45–61.

De Neys, W., Vartanian, W., & Goel, V. (2008). Smarter than we think: When our brains detect that we are biased. *Psychological Science*, *19*, 483–489.

DeWall, C. N., Baumeister, R. F., & Masicampo, E. J. (2008). Evidence that logical reasoning depends on conscious processing. *Consciousness and Cognition*, *17*, 628–645.

Dube, C., Rotello, C. M., & Heit, E. (2010). Assessing the belief bias effect with ROCs: It's a response bias effect. *Psychological Review*, *117*(831–863).

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

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

Evans, J. S. B. T., & Curtis-Holmes, J. (2005). Rapid responding increases belief bias: Evidence for the dual-process theory of reasoning. *Thinking and Reasoning*, *11*, 382–389.

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

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

Goel, V. (2007). Anatomy of deductive reasoning. *Trends in Cognitive Science*, *11*, 435–441.

Goel, V., & Dolan, R. J. (2003). Explaining modulation of reasoning by belief. *Cognition*, *87*, B11–B22.

Greenberg, J., Pyszczynski, T., & Solomon, S. (1986). The causes and consequences of a need for self-esteem: A terror management theory. In R. F. Baumeister (Ed.), *Public self and private self* (pp. 189–212). NY: Springer-Verlag.

Greene, J. D. (2007). Why are VMPFC patients more utilitarian? A dual-process theory of

moral judgment explains. *Trends in Cognitive Sciences*, *17*, 322–323.

Greene, J. D. (2013). *Moral tribes: Emotion, reason, and the gap between us and them*. Penguin Press.

Greene, J. D., Morelli, S. A., Lowenberg, K., Nystrom, L. E., & Cohen, J. D. (2008). Cognitive load selectively interferes with utilitarian moral judgment. *Cognition*, *107*, 1144–1154.

Greene, J. D., Nystrom, L. E., Engell, A. D., Darley, J. M., & Cohen, J. D. (2004). The neural bases of cognitive conflict and control in moral judgment. *Neuron*, *44*, 389–400.

Greene, J. D., Sommerville, R. B., Nystrom, L. E., Darley, J. M., & Cohen, J. D. (2001). An fMRI investigation of emotional engagement in moral judgment. *Science*, *293*, 2105–2108.

Haidt, J. (2001). The emotional dog and its rational tail: a social intuitionist approach to moral judgment. *Psychological Review*, *108*, 814–834.

Haidt, J. (2007). The new synthesis in moral psychology. *Science*, *316*, 998–1002.

Haidt, J. (2012). *The righteous mind: Why good people are divided by politics and religion*. New York: Pantheon.

Handley, S. J., Capon, A., Beveridge, M., Dennis, I., & Evans, J. S. B. T. (2004). Working memory, inhibitory control, and the development of children's reasoning. *Thinking and Reasoning*, *10*, 175–195.

Heit, E., & Rotello, C. M. (2014). Traditional difference-score analyses of reasoning are flawed. *Cognition*, *131*, 75–91.

Kahane, G., Wiech, K., Shackel, N., Farias, M., Savlescu, J., & Tracey, I. (2012). The neural basis of intuitive and counterintuitive moral judgments. *Social, Cognitive, and Affective Neuroscience*, *7*, 393–402.

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

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

Koenigs, M., Young, L., Adolphs, R., Tranel, D., Cushman, F., Hauser, M., & Damasio, A.

(2007). Damage to the prefrontal cortex increases utilitarian moral judgements. *Nature*, *446*, 908–911.

Kokis, J. V., MacPherson, R., Toplak, M. E., West, R. F., & Stanovich, K. E. (2002). Heuristic and analytic processing: Age trends and associations with cognitive ability and cognitive styles. *Journal of Experimental Child Psychology*, *83*, 26–52.

Koop, G. J. (2013). An assessment of the temporal dynamics of moral decisions. *Judgment and Decision Making*, *8*, 527–539.

MacPherson, R., & Stanovich, K. E. (2007). Cognitive ability, thinking dispositions, and instructional set as predictors of critical thinking. *Learning and Individual Differences*, *17*, 115–127.

McGuire, J., Langdon, R., Coltheart, M., & Mackenzie, C. (2009). A reanalysis of the personal/impersonal distinction in moral psychology research. *Journal of Experimental Social Psychology*, *45*, 577–580.

Miyake, A., Friedman, N. P., Rettinger, D. A., Shah, P., & Hegarty, M. (2001). How are visuospatial working memory, executive functioning, and spatial abilities related? A latent-variable analysis. *Journal of Experimental Psychology: General*, *130*, 621–640.

Moore, A. B., Clark, B. A., & Kane, M. J. (2008). Who shalt not kill? Individual differences in working memory capacity, executive control, and moral judgment. *Psychgological Science*, *19*, 549–557.

Newstead, S. E., Handley, S. J., Harley, C., Wright, H., & Farelly, D. (2004). Individual differences in deductive reasoning. *Quarterly Journal of Experimental Psychology*, *57A*, 33–60.

Paccini, R., & Epstein, S. (1999). The relation of rational and experiential information processing styles to personality, basic beliefs, and the ratio-bias phenomenon. *Journal of Personality and Social Psychology*, *76*, 972–987.

Paxton, J. M., Bruni, T., & Greene, J. D. (2014). Are counter-intuitive deontological judgments really counter-intuitive? An empirical reply to Kahane et al. (2012). *Social, Cognitive,*

and Affective Neuroscience, *9*, 1368–1371.

Paxton, J. M., & Greene, J. D. (2010). Moral reasoning: Hints and allegations. *Topics in Cognitive Science*, *2*, 511–527.

Paxton, J. M., Ungar, L., & Greene, J. D. (2012). Reflection and reasoning in moral judgment. *Cognitive Science*, *36*, 163–177.

Pyszczynski, T., Greenberg, J., & Solomon, S. (1999). A dual-process model of defense against conscious and unconscious death-related thoughts: An extension of terror management theory. *Psychological Review*, *106*, 835–845.

Rotello, C. M., & Heit, E. (in press). The neural correlates of belief bias: activation in inferior frontal cortex reflects response rate differences. *Frontiers in Human Neuroscience*.

Royzman, E. B., Landy, J. F., & Leeman, R. F. (in press). Are thoughtful people more utilitarian? CRT as a unique predictor of moral minimalism in the dilemmatic context. *Cognitive Science*.

Sa, W. C., West, R. F., & Stanovich, K. E. (1999). The domain specificity and generality of belief bias: Searching for a generalizable critical thinking skill. *Journal of Educational Psychology*, *91*, 497–510.

Shenhav, A., & Greene, J. D. (2014). Integrative moral judgment: Dissociating the roles of the amygdala and the ventromedial prefrontal cortex. *Journal of Neuroscience*, *34*, 4741–4749.

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

Stanovich, K. E., & West, R. F. (1997). Reasoning independently of prior belief and individual differences in actively open-minded thinking. *Journal of Educational Psychology*, *89*, 342–357.

Stanovich, K. E., & West, R. F. (1998). Individual differences in rational thought. *Journal of Experimental Psychology: General*, *127*, 161–188.

Stupple, E. J., & Ball, L. J. (2008). Belief–logic conflict resolution in syllogistic reasoning:

Inspection-time evidence for a parallel-process model. *Thinking & Reasoning*, *14*, 168–181.

Suter, R., & Hertwig, R. (2011). Time and moral judgment. *Cognition*, *119*, 454–458.

Svedholm-Häkkinen, A. M. (2015). Highly reflective reasoners show no signs of belief inhibition. *Acta Psychologica*, *154*, 69–76.

Thomas, B. C., Croft, K. E., & Tranel, D. (2011). Harming kin to save strangers: Further evidence for abnormally utilitarian moral judgments after ventromedial prefrontal damage. *Journal of Cognitive Neuroscience*, *23*, 2186–2196.

Thompson, V. A., & Evans, J. S. B. T. (2012). Belief bias in informal reasoning. *Thinking and Reasoning*, *18*, 278–310.

Trémolière, B., & Bonnefon, J. F. (2014). Efficient kill-save ratios ease up the cognitive demands on counterintuitive moral utilitarianism. *Personality and Social Psychology Bulletin*, *40*, 333-351.

Trémolière, B., De Neys, W., & Bonnefon, J. F. (2012). Mortality salience and morality: Thinking about death makes people less utilitarian. *Cognition*, *124*, 379–384.

Trémolière, B., De Neys, W., & Bonnefon, J. F. (2014). The grim reasoner: Analytical reasoning under mortality salience. *Thinking & Reasoningn*, *20*, 333-351.

Trippas, D., Handley, S. J., & Verde, M. (2014). Fluency and belief bias in deductive reasoning: New indices for old effects. *Frontiers in Psychology*, *5*, 631.

Trippas, D., Handley, S. J., & Verde, M. F. (2013). The SDT model of belief bias: complexity, time and cognitive ability mediate the effects of believability. *Journal of Experimental Psychology: Learning, Memory & Cognition*, *39*, 1393–1402.

Trippas, D., Verde, M., & Handley, S. J. (2014). Using forced choice to test belief bias in syllogistic reasoning. *Cognition*, *133*, 586–600.

Tsujii, T., Masuda, S., Akiyama, T., & Watanabe, S. (2010). The role of inferior frontal cortex in belief-bias reasoning: an rTMS study. *Neuropsychologia*, *48*, 2005–2008.

Tsujii, T., Okada, M., & Watanabe, S. (2010). Effects of aging on hemispheric asymmetry in

inferior frontal cortex activity during belief-bias syllogistic reasoning: A near-infrared spectroscopy study. *Behavioural Brain Research*, *210*, 178–183.

Tsujii, T., & Watanabe, S. (2009). Neural correlates of dual-task effect on belief-bias syllogistic reasoning: a near-infrared spectroscopy study. *Brain Research*, *1287*, 118–125.

Tsujii, T., & Watanabe, S. (2010). Neural correlates of belief-bias reasoning under time pressure: A near-infrared spectroscopy study. *Neuroimage*, *50*, 1320–1326.

West, R. F., Toplak, M. E., & Stanovich, K. E. (2008). Heuristic and biases as measures of critical thinking: Associations with cognitive ability and thinking dispositions. *Journal of Educational Psychology*, *100*, 930–941.

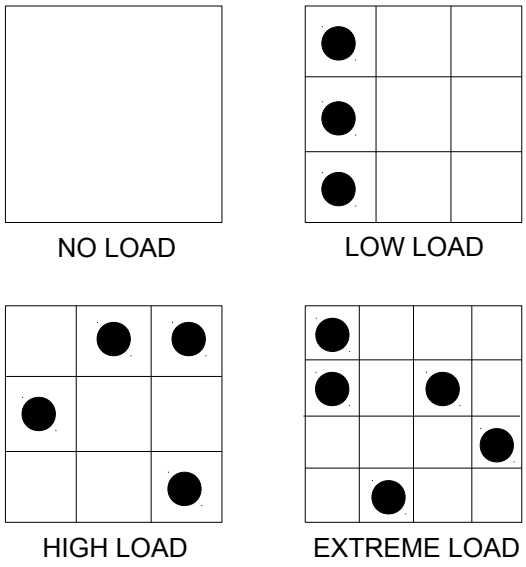

Figure 1. Examples of dot matrices used to induce no cognitive load (i.e., not matrix to memorize), low load, high load, and extreme load.