The Metacognitive Advantage of Deliberative Thinkers: A Dual-Process Perspective on Overconfidence

André Mata
University of Heidelberg

Mário B. Ferreira
University of Lisbon

Steven J. Sherman
Indiana University, Bloomington

We explored whether the thinking mode—deliberative versus intuitive—that people use to solve a problem or make a judgment influences their awareness of their own and others' performance. The results of 7 studies support the hypothesis that deliberative thinkers have a metacognitive advantage over intuitive thinkers: Deliberative thinkers are aware of both the deliberative solution and the intuitive alternative; realizing that the deliberative solution is better, they are likely to feel more confident and be more accurate in how they assess their performance and that of others. Intuitive thinkers, on the other hand, are aware only of the intuitive solution; whenever this solution is incorrect, they are unaware of how poor their performance was and how they rank in comparison to others. Implications of this metacognitive advantage are discussed.

Keywords: overconfidence, metacognition, reasoning, intuition, dual-process models

One of the prevailing perspectives on reasoning and on judgment and decision making (JDM) is a dual-process framework according to which the cognitive processes underlying our judgments, decisions, and problem solving fall into one of two categories: fast and largely automatic processes that enable us to take mental shortcuts to problem solving, referred to as the intuitive, heuristic, or System 1 processing mode, or slower analytical processes, referred to as the deliberative, rational, rule-based, or System 2 mode (e.g., Chaiken & Trope, 1999; Epstein, 1994; Evans, 2006; Ferreira, Garcia-Marques, Sherman, & Sherman, 2006; Kahneman, 2011; Kahneman & Frederick, 2002; Sloman, 1996; Stanovich & West, 2000). Very often, deliberation and intuition concur as to the best solution to a problem. However, there are circumstances in which these modes of thinking are in conflict and suggest different solutions. In this area of research, several problems have been created that present such situations of conflict and thus enable researchers to diagnose the operation of deliberative versus intuitive thought processes on the basis of the responses that people give. Consider this problem taken from the Cognitive Reflection Test (CRT; Frederick, 2005):

> A bat and a ball together cost 110 cents.
> The bat costs 100 cents more than the ball.
> How much does the ball cost?

The intuitive answer that comes immediately to mind is 10 cents, but this is incorrect. If the ball costs 10 cents and the bat costs 100 cents more, then the bat alone would cost 110 cents, which would make the combined price of bat and ball 120 cents. The correct answer is that the ball costs 5 cents and the bat 105 cents, but this solution requires reasoning deliberatively beyond the initial intuition. This follows from *default-interventionist* (Evans, 2007) models of reasoning and JDM (e.g., Evans, 2006; Kahneman & Frederick, 2002; Stanovich & West, 2000; Thompson, 2009; Thompson, Prowse Turner, & Pennycook, 2011). According to these models, the intuitive response is the default. It always comes to mind because it results from fast and effortless automatic processing. Slower and more effortful deliberative reasoning may or may not intervene, depending on cognitive resources, time, and motivation, although people often rely on intuition alone. But even if deliberation takes over, it does so only after the intuitive response has emerged. This is shown in research where people are asked to make judgments or solve reasoning problems quickly (e.g., De Neys, 2006; Evans & Curtis-Holmes, 2005; Finucane, Alhakami, Slovic, & Johnson, 2000; Roberts & Newton, 2001; Schroyens, Schaeken, & Handley, 2003; Tsujii &

This article was published Online First July 29, 2013.

André Mata, Department of Psychology, University of Heidelberg, Heidelberg, Germany; Mário B. Ferreira, Department of Psychology and Education, University of Lisbon, Lisbon, Portugal; Steven J. Sherman, Department of Psychology, Indiana University, Bloomington.

This research was supported by grants from the Fundação para a Ciência e a Tecnologia (Portugal) and from the Deutsche Forschungsgemeinschaft (Germany). We are thankful to Joana Reis for her assistance in data collection.

Correspondence concerning this article should be addressed to André Mata, Department of Psychology, University of Heidelberg, Hauptstrasse 47-51, 69117 Heidelberg, Germany. E-mail: andre.mata@psychologie.uni-heidelberg.de

Journal of Personality and Social Psychology, 2013, Vol. 105, No. 3, 353–373
 0022-3514/13/$12.00 DOI: 10.1037/a0033640

Watanabe, 2010). These speeded responses are usually more consistent with more intuitive, heuristic processes, and only as the time to answer is increased are the responses more consistent with a deliberative analysis (although giving people extra time to think does not always guarantee that they will come to the correct solution; e.g., Thompson et al., 2011).

In this paper, we explore a consequence that the temporal dynamics of intuition and deliberation have for metacognition: Whenever the two modes of thinking suggest different answers, people who rely on intuition alone will not detect the conflict between the intuitive and the deliberative solutions; they will be aware only of the intuitive solution (but see De Neys, 2012). However, because deliberative reasoning intervenes only after one has thought of the intuitive solution, people who go beyond their initial intuition and respond deliberatively are likely to be aware of both the deliberative and the intuitive solutions (i.e., in the bat-and-ball problem, a person who gives the correct answer, 5 cents, thought of the incorrect response, 10 cents, first, and is therefore aware of both alternatives).

We suggest that this double awareness of the intuitive and the deliberative solutions puts deliberative problem solvers at a clear metacognitive advantage over intuitive problem solvers. First, they can feel highly confident in their solutions because they know that they did not fall for the wrong answer primed by the question. They might not base their confidence and their estimates of how well they performed on a *feeling of rightness* (Thompson, 2009; Thompson et al., 2011) but rather on a deeper understanding that there is a right answer and a wrong answer and that their solution is the right one (Slovic & Tversky, 1974; Stanovich, 1999). Second, the fact that deliberative responders know that there is an alternative solution that is compelling but incorrect allows them to infer that many others might have failed to reach the same correct solution that they came up with, which ultimately enables them to make more accurate estimates of how well they performed in comparison to others. Intuitive responders, on the other hand, are likely not only to believe mistakenly that their solution is correct but also that everybody else came up with the same obvious (but wrong) answer.

Competence and Metacognition

Research on people's ability to assess their competence suggests that the unskilled are often overconfident (for reviews, see Carter & Dunning, 2008; Dunning, 2011; Dunning, Johnson, Ehrlinger, & Kruger, 2003). Kruger and Dunning (1999, 2002) argued that the unskilled are unaware of their lack of skill precisely because they are unskilled. Because they do not know how to do better, they also do not realize that they could have done better. Therefore, they overestimate how good their performance was and how skillful they are (e.g., Ehrlinger, Johnson, Banner, Dunning, & Kruger, 2008; Moore & Healy, 2008), they overplace their performance or skills in comparison to others (e.g., Burson, Larrick, & Klayman, 2006; Ehrlinger et al., 2008; Kruger & Dunning, 1999; Larrick, Burson, & Soll, 2007), and they express levels of confidence in their performance or skills that are not warranted by reality (e.g., Klayman, Soll, Gonzalez-Vallejo, & Barlas, 1999; Larrick et al., 2007).

But it is not only the unskilled who have metacognitive shortcomings. The skilled have also been shown to be somewhat unaware of how well they performed in comparison to others (e.g., Burson et al., 2006; Kruger & Dunning, 1999). They lack the knowledge of how others performed, and, led by an illusion of false consensus (Ross, Greene, & House, 1977), they tend to believe that many others performed as well as they did. In a related finding, Nickerson, Baddely, and Freeman (1987) showed that people's estimates about others' ability to answer general knowledge questions are higher when they themselves know (or think they know) the answers than when they do not. Because skilled performers "fail to realize that their proficiency is not necessarily shared by their peers" (Kruger & Dunning, 1999, p. 1131), they underplace their performance level in relation to that of others. Some authors argue that good and bad performers are equally incompetent in metacognition and that people may be "skilled or unskilled, but still unaware of it" (Burson et al., 2006, p. 60).

This pessimistic view according to which people in general, even competent people, are bad at assessing their competence stands in contrast with our hypothesis of a metacognitive advantage for deliberative thinkers. A crucial difference between extant research on overconfidence and our research on confidence in deliberative versus intuitive thinking is that, in extant research, true skill and judgments of skill are assessed in domains where there is no clear conflict between intuitive solutions that come to mind immediately, and that feel right despite being wrong, and answers that are right but that require reasoning beyond one's initial intuition. For instance, in tests of grammar (Kruger & Dunning, 1999) or of knowledge about guns and safety (Ehrlinger et al., 2008), individuals either know or do not know the solution. And even if they know it, they have no clue as to whether others know it or not. But if people first have to overcome an initial intuition in order to come up with the correct answer, and if they feel how compelling the incorrect intuitive solution is and how easily people might be misled by it, they might feel warranted in ranking themselves above most others and be less likely to fall prey to false consensus.

In summary, whenever intuition and deliberative reasoning are in conflict, suggesting different responses, and the deliberative response is the correct one, we expect intuitive responders will be unskilled and unaware, whereas deliberative responders will be skilled and more aware. Intuitive responders know neither what the correct answer is nor that others might have given this other answer. Deliberative responders, on the other hand, not only have the knowledge of what the correct answer is but are also aware of the intuitive option and how compelling that intuition is. This knowledge enables them to guess that many others responded differently and incorrectly. Therefore, deliberative responders will be more accurate or realistic than intuitive responders in their assessment of how well they did, how they rank in comparison to others, how confident they are in their responses, and, most important, their assessment of how well others did. A dual-process (deliberative vs. intuitive) perspective of reasoning and JDM has been largely absent from research on overconfidence (but see Prowse Turner & Thompson, 2009; Shynkaruk & Thompson, 2006).

To test this hypothesis, we assessed three kinds of overconfidence (Moore & Healy, 2008)—overestimation (overestimating the quality of one's performance), overplacement (overestimating one's standing in comparison to others), and unwarranted confidence (expressing confidence levels that are not justified by one's actual performance)—as a function of the type of thinking that people used to solve problems and of the metacognitive knowledge about the different possible solutions to the problems. In Studies 1–4, type of thinking and metacognitive knowledge about alternative solutions were measured, and their relation with performance estimates was assessed. In Studies 5 and 6, each of these variables was manipulated, and their effect on performance estimates was tested.

Study 1

In the first study, participants took the CRT (see Appendix A) and then made both absolute estimates about how good their performance and others' performance was and comparative estimates about how well they performed in relation to others. Deliberative responders were expected to be more accurate than intuitive responders in these estimates. Participants were also asked to guess what others responded. Deliberative responders were expected to be aware of the alternative, intuitive solutions, whereas intuitive responders were expected to be aware only of the intuitive solutions.

Method

Participants. Eighty-six undergraduates from Indiana University participated and received course credit for their participation.

Procedure. Participants completed a paper and pencil questionnaire that started with the three CRT problems. Upon completing the test, participants were asked to make several absolute and comparative estimates.

With regard to absolute estimates, they were asked to indicate how many of the three problems they thought they had answered correctly (this estimate will be referred to in the Results section as *Number of correct answers by self*), as well as how many of the three problems they thought most participants had answered correctly (*Number of correct answers by others*).

With regard to comparative estimates, they were asked to indicate how good they thought their performance was in comparison to that of other students who participated in the study (*Better or worse than others;* scale from −3 = *much worse than most other students* through 0 = *same as most other students* to +3 = *much better than most other students*). They were also asked to estimate where they thought they ranked in terms of percentile of performance relative to the other participants (*percentile*; from 1 to 100, 1 being the worst performance and 100 the best). One item asked participants to rate how hard it had been for them to answer the CRT problems (*difficulty*; 1 = *very easy* to 9 = *very hard*).

There were two questions about confidence, one that asked participants to rate how confident they were in their answers (*confidence*; 1 = *not at all confident* to 9 = *very confident*) and one that asked participants whether they would be willing to gamble on their performance: "Imagine a hypothetical scenario where you would be given the opportunity to win or lose money depending on whether your answers to the three problems were correct or incorrect. Specifically, imagine that you would be offered the chance to win 5 dollars for every question you answered correctly and to lose 5 dollars for every question you answered incorrectly. Thus, you would win money if you solved most of the problems correctly but you would lose money if you solved most of them incorrectly. Would you be willing to play this game and bet on your answers— only hypothetically?" (The response options were "yes" and "no"; the results obtained with this measure will be referred to as the *Percentage of participants willing to bet on their answers.*).

Participants were also asked to guess what others had responded on each problem (*Others' responses*). They were informed that they would be seeing the same three problems again, but that this time, instead of providing their solutions, they were to provide the answers that they believed most participants gave to each problem. The instructions stated that these answers could be exactly the same as or different from their answers. For each of the three CRT problems, after indicating the answer that they believed most others gave, participants estimated the percentage of participants in this study who solved the problem correctly. They were also asked to estimate the percentage of other participants that they thought answered all three problems correctly (*Percentage of others who solved all problems correctly*).

Results and Discussion

Because our hypotheses concern the greater metacognitive accuracy of deliberative thinkers, we used planned contrasts pitting deliberative responders against all other responders or comparing them to intuitive responders in particular. For paired-samples t tests, Cohen's d was calculated correcting for dependence between means (Morris & DeShon, 2002). Degrees of freedom were adjusted whenever Levene's test revealed unequal variances.

Performance. In order to compare participants' metacognition as a function of whether they responded deliberatively or intuitively, we report data pertaining only to those participants who gave either the deliberative or the intuitive responses to each problem and not those who gave some other response or no response to at least one of the problems. However, results are generally the same regardless of the performance criterion that is adopted. That is, in all the studies, results are almost identical if we consider the data from all participants. For the few cases where performance criteria make a difference in the significance of the effects, we also report the results from analyses considering data from all participants.

Thirty-nine percent of the participants did not answer any problem correctly; 24.3% answered only one problem correctly; 17.6% answered two problems correctly; and 18.9% answered all three problems correctly. Throughout this paper, we refer to participants who gave the intuitive solution to every problem as *intuitive responders* and to those who gave a solution consistent with deliberative thinking (in this case, the correct responses) to every problem as *deliberative responders*. Table 1 shows the mean estimates and confidence across these performance levels.

Number of correct answers by self. Deliberative responders made higher estimates about the number of problems that they solved correctly than did either intuitive responders, $t(36) = 3.31$, $p = .002$, $d = 1.10$, or all other participants combined, $t(56) = 5.11$, $p < .001$, $d = 1.36$. The difference between the estimated and the actual number of correct answers reveals that intuitive responders overestimated their performance, one-sample $t(28) = 13.38$, $p < .001$, $d = 5.06$, whereas deliberative responders did not, $t(13) = -1.00$, $p = .336$, $d = 0.55$.

Table 1
Mean Estimates and Standard Deviations (SDs) by Number of Correct Answers (Study 1)

Estimate (range)	Actual number of correct answers			
	0	1	2	3
Number of correct answers by self (0–3)	2.31_a (0.93)	2.06_a (0.73)	2.62_{ab} (0.51)	2.93_b (0.27)
Number of correct answers by others (0–3)	2.43_a (0.69)	1.78_b (0.55)	1.77_b (0.60)	1.64_b (0.75)
Number of correct answers by self versus others (−3 to 3)	-0.14_a (0.71)	0.28_{ab} (0.75)	0.85_{bc} (0.55)	1.29_c (0.83)
Percentage of others who solved all problems correctly (0–100%)	74.50_a (28.06)	46.28_b (29.14)	46.69_b (28.13)	37.71_b (20.51)
Better or worse than others (−3 to 3)	0.14_a (0.92)	0.56_{ab} (1.10)	1.15_{bc} (1.21)	1.79_c (1.05)
Percentile (0–100%)	68.38_a (22.26)	66.72_a (16.51)	71.77_{ab} (15.05)	87.57_b (9.74)
Difficulty (1–9)	3.10_a (2.08)	3.72_a (2.11)	4.46_a (1.85)	3.76_a (2.05)
Confidence (1–9)	6.19_a (2.39)	6.50_a (1.54)	6.77_a (1.30)	7.57_a (2.14)
Percentage of participants willing to bet on their answers (0–100%)	65.5	83.3	92.3	92.9

Note. Means with different subscripts are different at $p \leq .05$ in a Tukey post hoc test.

Number of correct answers by others. When asked how many problems most participants answered correctly, deliberative responders estimated a lower number than did intuitive responders, $t(69) = 3.67$, $p < .001$, $d = 0.88$, and their estimates also tended to be lower than those of all other participants combined, $t(69) = 1.78$, $p = .080$, $d = 0.43$. Regardless of whether this question was interpreted as referring to the median (1) or the mode (0), all groups overestimated others' performance, but deliberative participants were better calibrated than other participants.

Number of correct answers by self versus others. The difference between the number of problems that participants thought they had answered correctly and the number of problems that they thought others had answered correctly was greater for deliberative responders than for either intuitive responders, $t(69) = 6.08$, $p < .001$, $d = 1.46$, or all other participants combined, $t(69) = 4.45$, $p < .001$, $d = 1.07$. Deliberative responders estimated a higher number of correct responses for themselves than for others, paired $t(13) = 5.83$, $p < .001$, $d = 1.69$, but there was no difference between the number of correct answers that intuitive responders thought they themselves and others had given, paired $t(27) = 1.07$, $p = .293$, $d = 0.21$.

Percentage of others who solved all problems correctly. Whereas intuitive responders estimated that most participants solved all three problems correctly, deliberative responders estimated that most participants failed to do so. The estimates of deliberative responders were closer to the true value (18.9%) than were those of intuitive responders, $t(67) = 4.10$, $p < .001$, $d = 1.00$, and those of all other groups combined, $t(67) = 2.23$, $p = .029$, $d = 0.54$.

Better or worse than others. Deliberative responders believed that they had performed better than others, whereas intuitive responders believed they had done as well as others. Deliberative responders were significantly more confident than intuitive responders, $t(23) = 5.02$, $p < .001$, $d = 2.09$, or all other groups combined, $t(21) = 3.66$, $p = .001$, $d = 1.60$. These comparative estimates were regressed on the estimated number of correct answers for the self and for most others, and both were significant predictors: $\beta_{self} = .65$, $p < .001$; $\beta_{others} = -.47$, $p < .001$; $R^2 = .45$.

Percentile. Deliberative responders made higher percentile estimates than did other participants in general, $t(37) = 5.31$, $p < .001$, $d = 1.75$, and intuitive responders in particular, $t(41) = 3.93$, $p < .001$, $d = 1.23$. To determine the accuracy of these estimates, we compared participants' estimated percentiles to the midpoint of the actual percentile range (for example, intuitive responders ranked in the bottom 39.2%, so the midpoint of their percentile range is 19.6). Intuitive responders overestimated their true percentile by 48.78%, one-sample $t(28) = 11.80$, $p < .001$, $d = 4.46$, whereas deliberative responders were accurate in their estimates ($M = -2.98\%$), one-sample $t(13) = 1.14$, $p = .273$, $d = 0.63$. Estimated percentiles were regressed on the estimated number of correct answers for the self and for most others: Estimates about one's own performance were a significant predictor, $\beta_{self} = .57$, $p < .001$; estimates of others' performance had a marginally significant effect, $\beta_{others} = -.18$, $p = .079$; $R^2 = .30$.

Others' responses. Participants made three guesses, one for each problem, about what most others had responded. Across the three problems, 96.4–100% of the intuitive responders guessed that most others had also given the intuitive response, not realizing that there was another solution. However, most (61.5–85.7%) deliberative responders guessed that most others had given the intuitive response, even though they themselves did not give that answer. First, deliberative responders were correct in estimating that most participants did not give the deliberative solutions. More important, they were aware that there were intuitive answers to these problems and that most others might have given them.

When asked to estimate the percentage of people who answered each of the three problems correctly, deliberative responders made lower estimates (*M*s ranging between 52 and 59%) than did either intuitive responders in particular (*M*s ranging between 77 and 88%) or all other participants combined (all $ps < .02$).

Difficulty. Experienced difficulty did not differ across performance levels ($ts < 1$).

Confidence. Deliberative responders tended to display greater confidence that their solutions were correct than did intuitive responders, $t(29) = 1.91$, $p = .066$, $d = 0.71$, and all other groups combined, $t(17) = 1.77$, $p = .095$, $d = 0.86$. (If all participants are considered and not just those who gave either the intuitive or the deliberative responses, both these differences are significant at $p \leq .05$.).

Percentage of participants willing to bet on their answers. Deliberative responders tended to be more willing than intuitive responders to bet on their answers, $\chi^2(1, N = 43) = 3.71$, $p = .054$, $\varphi = .29$. When deliberative responders are compared to the aggregate of other responders, this difference is not significant, $\chi^2(1, N = 74) = 1.84$, $p = .175$, $\varphi = .16$.

Summary. Intuitive responders clearly overestimated how many problems they solved correctly. Moreover, they thought that others gave the same intuitive solutions that they did. Consistently, they did not think that their performance was much better than that of others. Deliberative responders, on the other hand, estimated that they performed better than most others. Furthermore, in addition to knowing the correct responses, they proved to be aware of alternative intuitive solutions that others might have offered, which might explain why they estimated that their performance was better than that of others. Whereas deliberative thinkers were accurate, intuitive thinkers were clearly overconfident, as operationalized in the first two types of overconfidence identified by Moore and Healy (2008), overestimation and overplacement. There was also evidence that intuitive responders were highly confident in the quality of their performance, which was unrealistic, but deliberative responders were (correctly) even more confident in the quality of theirs.

Study 2

In the second study, instead of using the general measures used in the previous study (and in the following ones), which asked only about the confidence in one's overall performance, we asked for confidence ratings associated with each answer. This offered a better way to test our metacognitive advantage hypothesis with specific regard to confidence in performance. In this study, we wished to test (a) whether intuitive problem solvers are unrealistically confident and (b) whether deliberative problem solvers are more confident than intuitive problem solvers.

More important, by having confidence ratings for each problem, it was possible to test not only whether consistently deliberative responders and consistently intuitive responders express different levels of confidence but also whether the same responders express different levels of confidence depending on whether they respond in a deliberative or intuitive fashion to different problems. Consistent with our hypothesis, participants should express greater confidence when they respond deliberatively than when they respond intuitively. Such a result would provide stronger evidence that differences in confidence are indeed associated with differences in modes of thinking and not with some other individual-difference variable.

Method

Participants. Seventy-five undergraduates from the University of Lisbon participated and received course credit for their participation.

Procedure. Participants completed a paper and pencil questionnaire with the CRT (see Appendix A), and, after answering each problem, they rated their confidence in their response on a scale from 1 (*not at all confident*) to 9 (*very confident*). After this, they were asked whether they would be willing to bet on their answers (the gamble measure of Study 1). In the end, they were asked what they thought others had responded, just as in the first study.

Results and Discussion

Performance. Thirty-nine percent of the participants did not answer any problem correctly (the intuitive responders); 26.9% answered only one problem correctly; 17.3% answered two problems correctly; and 17.3% answered all three problems correctly (the deliberative responders).

Confidence for individual problems. In all three problems, participants who provided the correct, deliberative response were more confident than those who provided the incorrect, intuitive response ($ts > 2.57$, $ps < .013$, $ds > 0.67$).

The previous analysis involved between-subjects comparisons. In order to test whether the same participants would feel different levels of confidence depending on whether they responded in a deliberative or intuitive fashion to different problems, we examined the confidence ratings of those participants who responded both ways; that is, those who gave one or two deliberative responses to the three problems (and therefore also gave two or one intuitive responses, respectively). Those participants expressed higher confidence when they responded deliberatively than when they responded intuitively ($M = 8.30$, $SD = 1.17$ vs. $M = 7.07$, $SD = 1.98$), paired $t(22) = 2.63$, $p = .015$, $d = 0.56$.

Average confidence across problems. The confidence scores associated with the answers to the three CRT problems were aggregated into an average confidence score (see Table 2). Deliberative responders were more confident overall than both intuitive responders, $t(48) = 2.63$, $p = .011$, $d = 0.76$, and all other participants combined, $t(48) = 2.30$, $p = .026$, $d = 0.66$.

Percentage of participants willing to bet on their answers. All deliberative responders indicated that they were willing to bet on their answers. This rate was higher than that for other participants: comparing deliberative versus intuitive responders, $\chi^2(1, N = 29) = 4.15$, $p = .042$, $\varphi = .38$; comparing deliberative versus all other responders, $\chi^2(1, N = 52) = 3.63$, $p = .057$, $\varphi = .26$.

Others' responses. For all three problems, 100% of the intuitive responders guessed that most others had also given the intuitive response, not realizing that there was another solution. However, most (77.8–100%) deliberative responders guessed that most others had given the intuitive response.

Table 2
*Mean Confidence and Standard Deviations (*SD*s) by Number of Correct Answers (Study 2)*

	Actual number of correct answers			
Estimate (range)	0	1	2	3
Average confidence across problems (1–9)	7.30_{a} (1.59)	7.38_{ab} (1.45)	8.04_{ab} (0.98)	8.70_{b} (0.48)
Percentage of participants willing to bet on their answers (0–100%)	64.3	65.0	89.9	100

Note. Means with different subscripts are different at $p \leq .05$ in a Tukey post hoc test.

Summary. First, both intuitive and deliberative responders were confident; their confidence ratings were consistently above the middle point of the scale. Whereas intuitive respondents were overconfident, with their performance not warranting such high confidence, deliberative respondents were justly confident, given their good performance. Second, deliberative responders expressed very high confidence (their average confidence was 8.70 on a scale from 1 to 9), significantly higher than the confidence levels of intuitive respondents. This confidence boost presumably came from their awareness not only of the deliberative solutions but also of the highly compelling intuitive solutions (most deliberative responders were indeed aware of the intuitive solutions), although we did not test this causal process directly. Third, a within-subjects analysis revealed that the same responders expressed greater confidence when they responded deliberatively than when they responded intuitively, which further suggests that the observed differences in metacognition follow from differences in modes of thinking.

Study 3

In this study, participants solved several base rate (BR) problems, some in which the intuitive response (based on the stereotypical description) and the deliberative response (based on the BR) differed (conflict condition), and others in which the two ways of thinking concurred (no conflict). Then, they made estimates about their performance as well as that of others. By using no-conflict problems with the exact same structure as conflict problems except for the way BR information is presented (supportive of vs. opposed to stereotypical information), we expected to remove deliberative responders' metacognitive advantage while keeping everything else constant. Thus, our primary interest in this study was not how accurate deliberative and intuitive responders are in assessing their performance but rather whether they make different assessments for the different types of problems.

Deliberative responders were expected to think that they performed better than others in the conflict condition, because for those problems they would realize that there was an intuitive solution that other participants might have chosen. However, they were not expected to make such high comparative assessments for no-conflict problems, because for these problems the intuitive solution was the same as the deliberative one. Intuitive responders, on the other hand, were not expected to make different estimates for the different types of problems, because they should not be aware of the existence of the deliberative solutions or whether these were in conflict with the intuitive solutions.

Method

Participants. Forty-six undergraduates from the University of Lisbon participated and received course credit for their participation.

Procedure. Participants completed a paper and pencil questionnaire with two blocks of four BR problems each (conflict and no-conflict conditions), adapted from De Neys and Franssens (2009; see Appendix B), presented in counterbalanced order. After presentation of each of the two blocks of problems, participants were asked to assess their performance and that of others. They did so by completing the same measures as in Study 1, except for (a) the estimated percentages of other participants who solved each problem correctly, because this measure did not add much to the other measures, and (b) willingness to bet on accuracy, which was not considered well suited for the kind of problems used in this study.

Results and Discussion

Performance. In the no-conflict problems, participants gave the expected BR- and stereotype-consistent answers to all four problems (87% of the participants) or to three of the four problems (13%). Because our interest is in how deliberative and intuitive problem solvers differ in their estimates of their performance in comparison to others, we used performance in the conflict problems to identify deliberative versus intuitive responders. In the conflict problems, 56.5% of the participants did not answer any problem according to the BRs (the intuitive responders); 13.0% gave one BR-consistent response; 15.2% gave two BR-consistent responses; 2.2% (one participant) gave three BR-consistent responses; and 13.0% (the deliberative responders) answered all four problems according to the BRs. (Because there were few consistently deliberative responders, in the following data analyses, we grouped the one participant who answered three problems according to the BRs with those who answered all four problems according to the BRs. Doing so did not change the pattern of results but ensured greater statistical power.)

The mean estimates across all performance levels are displayed in Table 3. To test our hypothesis, we compared only intuitive versus deliberative responders (the pattern of results is the same if deliberative responders are compared not only to intuitive responders but to all other responders).

Number of correct answers by self and others. A mixed model analysis of variance (ANOVA) with the target of estimates (one's own performance vs. others' performance) and the type of problem (conflict vs. no-conflict) as within-subjects variables and performance (deliberative vs. intuitive) as a between-subjects variable revealed a main effect of target, $F(1, 31) = 28.02$, $p < .001$, $\eta_p^2 = .48$, such that estimates for the self were higher than estimates for others; a target × performance interaction, $F(1, 31) = 27.19$, $p < .001$, $\eta_p^2 = .47$, such that only deliberative responders made higher estimates for self than others; a target × type of problem interaction, $F(1, 31) = 9.14$, $p = .005$, $\eta_p^2 = .23$, such that higher estimates for self versus others were made only for conflict problems; and a target × performance × type of target interaction, $F(1, 31) = 11.19$, $p = .002$, $\eta_p^2 = .27$, such that the higher estimates for self versus others were made only by deliberative responders and more for conflict problems, paired $t(6) = 3.12$, $p = .021$, $d = 1.30$, than for no-conflict problems, paired $t(6) = 1.99$, $p = .094$, $d = 1.36$. Intuitive responders did not make different estimates for self versus others in either conflict or no-conflict problems (ts < 1). Deliberative responders did not make different estimates about their own performance for the different types of problem ($t < 1$), but their estimates of others' performance were higher for no-conflict problems than for conflict problems, paired $t(6) = 2.47$, $p = .049$, $d = 0.94$.

Percentage of others who solved all problems correctly. When asked to estimate the percentage of participants who solved all four problems correctly, the estimates of intuitive responders for conflict and no-conflict problems were very similar, but the

Table 3
Mean Estimates and Standard Deviations (SDs) by Number of Base-Rate-Consistent Answers and Type of Problem (Conflict vs. No Conflict; Study 3)

	Number of base-rate-consistent answers							
	0		1		2		4	
Estimate (range)	Conflict	No conflict	Conflict	No conflict	Conflict	No conflict	Conflict	No conflict
Number of correct answers by self (0–4)	3.08 (1.13)	3.00 (0.69)	2.83 (1.94)	3.67 (0.52)	2.43 (0.79)	3.00 (0.82)	3.71 (0.49)	3.86 (0.38)
Number of correct answers by others (0–4)	3.10 (1.22)	2.96 (0.77)	3.20 (0.84)	3.20 (0.84)	2.29 (0.49)	2.86 (0.69)	1.86 (1.46)	3.14 (1.21)
Percentage of others who solved all problems correctly (0–100%)	60.36 (35.22)	63.68 (32.90)	53.33 (37.24)	66.67 (35.59)	35.14 (20.63)	62.14 (29.98)	37.43 (27.89)	81.00 (30.09)
Better or worse than others (−3 to 3)	0.04 (0.34)	0.04 (0.20)	0.17 (0.41)	0.17 (0.41)	0.14 (0.90)	0.14 (0.38)	1.43 (1.27)	0.43 (0.79)
Percentile (0–100%)	65.38 (19.80)	65.31 (19.14)	63.33 (19.66)	67.50 (24.03)	57.86 (10.75)	71.43 (18.64)	70.71 (18.13)	67.14 (20.59)
Difficulty (1–9)	2.77 (2.05)	3.12 (2.08)	5.67 (2.66)	4.00 (2.19)	4.00 (2.00)	2.29 (1.50)	2.86 (1.35)	2.71 (2.06)
Confidence (1–9)	6.54 (2.50)	6.73 (1.95)	5.33 (2.66)	6.83 (2.04)	5.00 (1.73)	7.00 (1.63)	7.43 (1.62)	7.71 (1.11)

percentage estimated by deliberative responders more than doubled from conflict to no-conflict problems (see Table 3). There was a significant main effect of the type of problem, $F(1, 29) = 9.81$, $p = .004$, $\eta_p^2 = .25$, and type of problem interacted with performance, $F(1, 29) = 8.01$, $p = .008$, $\eta_p^2 = .22$, such that deliberative responders made higher estimates for no-conflict versus conflict problems, paired $t(6) = 3.56$, $p = .012$, $d = 1.38$, whereas intuitive responders did not make different estimates for the different types of problems ($t < 1$). The main effect of performance was not significant ($F < 1$).

Better or worse than others. There was a significant main effect of performance, with deliberative responders making higher comparative assessments than intuitive responders, $F(1, 31) = 20.29$, $p < .001$, $\eta_p^2 = .40$, and a significant main effect of type of problem, with higher values for conflict versus no-conflict problems, $F(1, 31) = 21.37$, $p < .001$, $\eta_p^2 = .41$. There was a significant interaction of performance and type of problem, $F(1, 31) = 21.37$, $p < .001$, $\eta_p^2 = .41$, such that deliberative responders made higher comparative assessments for conflict versus no-conflict problems, $t(6) = 2.65$, $p = .038$, $d = 1.11$, whereas intuitive responders made the exact same assessments for both kinds of problems.

Percentile. Estimated percentiles showed the same pattern as the better-or-worse question, with intuitive responders making the same estimates for conflict and no-conflict problems and deliberative responders making higher estimates for conflict problems than for no-conflict problems; but this difference was not significant. There was no significant main effect of either performance or type of problem, and the interaction between these two was also not significant (Fs < 1).

Others' responses. Across the four conflict problems, 84.6–100% of the intuitive responders guessed that most others had also given the intuitive response, not considering the alternative solution. However, most (57.1–71.4%) deliberative responders guessed that most others had given the intuitive response. These guesses were related to how deliberative responders compared their performance to that of others: When they were asked whether they had performed better or worse than others, deliberative responders who guessed that most other participants had also responded deliberatively to all four problems estimated that their performance was the same as that of most other participants ($M = 0.00$, $SD = 0.00$), but deliberative responders who guessed that most other participants had not responded deliberatively to any of the four problems estimated that their performance was better than that of most others ($M = 2.00$, $SD = 1.00$), $t(3) = 2.68$, $p = .075$, $d = 3.09$. The same pattern emerged for percentile estimates: Deliberative responders who thought that most others had also responded deliberatively to all four problems ranked themselves at the median ($M = 50.00$, $SD = 0.00$), whereas deliberative responders who thought that most others had not responded deliberatively to any of the four problems ranked themselves higher ($M = 81.67$, $SD = 11.55$), $t(3) = 3.68$, $p = .035$, $d = 4.25$.

Difficulty. There were no significant differences in the overall difficulty experienced by intuitive versus deliberative responders or in the difficulty experienced in conflict versus no-conflict problems, and there was no interaction between these factors (all Fs < 1).

Confidence. Confidence did not vary across the different types of problem, and there was no interaction between performance and type of problem (Fs < 1). Deliberative responders were not significantly more confident than intuitive responders, $F(1, 31) = 1.62$, $p = .212$.

Summary. As predicted, deliberative thinkers made different comparative assessments depending on whether there was a conflict between deliberation and intuition or not, presumably because they were aware of both the deliberative and the intuitive solutions and whether these were in agreement or not. Intuitive participants, on the other hand, did not make different comparative assessments for conflict versus no-conflict problems, presumably because they relied only on intuition and were not aware of the deliberative solutions. For them, there was no difference between conflict and no-conflict problems.

Study 4

This study provided a conceptual replication of Study 1 with another reasoning task: Participants solved several syllogistic problems for which the intuitive response (based on the believability of the conclusion) and the deliberative response (based on logic) differed, and then they assessed their performance and that of others.

More important, this study tested whether guesses about what others responded are related to how deliberative responders compare their performance to that of others. According to the metacognitive advantage hypothesis, the reason for deliberative thinkers' greater awareness and self-enhancing estimates is not just because deliberative thinkers are deliberative but also because they know that many others think only of the intuitive solution. Therefore, if deliberative thinkers are not aware of the intuitive solution or if they do not believe that most others respond intuitively, they are expected to make lower comparative estimates than deliberative responders who think that most others give the intuitive responses.

In Studies 1 and 2, it was not possible to test this hypothesis, because there were no deliberative responders who believed that most others had been deliberative like them. In Study 3, results were consistent with this prediction, but still there were few deliberative responders who believed that most others had been deliberative as they had. Previous research shows that accuracy levels are typically higher for syllogisms than for the CRT or BR problems (e.g., De Neys & Franssens, 2009; Frederick, 2005). In studies of syllogistic reasoning, participants are usually explicitly instructed to assume the truthfulness of the premises and then to evaluate if the conclusions follow logically from these premises. Such instructions may well call their attention to the fact that they have to inhibit their previous beliefs in order to get their answers right. This kind of hint is not given in the CRT or BR problems, and it may lead to a higher number of deliberative responders, as well as more deliberative responders believing that others also answered deliberatively. Indeed, in Study 4, there were more deliberative responders who thought that most others had responded deliberatively to all problems, just as many as those who thought that most others did not respond deliberatively to any of the problems. Therefore, it was possible to compare their estimates.

Method

Participants. Seventy-four undergraduates from the University of Lisbon participated and received course credit for their participation.

Procedure. Participants completed a paper and pencil questionnaire containing four syllogistic problems (adapted from De Neys & Franssens, 2009) that either have conclusions that are believable but invalid or have conclusions that are valid but unbelievable (see Appendix C). For each syllogism, participants were asked or not whether the conclusion follows logically from the premises. The instructions said that a conclusion should be accepted only if it follows logically from the premises and that the syllogisms should be analyzed assuming that the premises are true. After they had solved these problems, participants were asked to assess their performance and that of others by completing the same measures as in Studies 1 and 3.

Results

Performance. Fifteen percent of the participants did not answer any problem correctly (the intuitive responders); 28.4% answered two problems correctly; 13.5% answered three problems correctly; and 43.2% answered all four problems correctly (the deliberative responders). Table 4 shows the mean estimates and confidence across these performance levels.

Number of correct answers by self. Participants from all performance levels estimated that they solved three or more problems correctly on average, with no significant differences between groups (ts < 1). Intuitive participants overestimated their actual number of correct answers, one-sample $t(10) = 9.02$, $p < .001$,

Table 4
Mean Estimates and Standard Deviations (SDs) by Number of Correct Answers (Study 4)

	Actual number of correct answers			
Estimate (range)	0	2	3	4
Number of correct answers by self (0–4)	3.09_a (1.14)	3.14_a (1.10)	3.00_a (1.33)	3.44_a (1.11)
Number of correct answers by others (0–4)	3.45_a (0.69)	3.00_{ab} (1.00)	2.70_{ab} (0.48)	2.63_b (0.98)
Number of correct answers by self versus others (−4 to 4)	-0.36_a (0.67)	0.14_{ab} (1.20)	0.30_{ab} (1.06)	0.81_b (1.33)
Percentage of others who solved all problems correctly (0–100%)	75.00_a (18.44)	63.67_a (27.03)	59.50_a (20.06)	55.94_a (25.82)
Better or worse than others (−3 to 3)	-0.09_{ab} (1.14)	-0.33_a (1.07)	0.30_{ab} (0.68)	0.84_b (1.10)
Percentile (0–100%)	56.36_a (15.67)	57.14_a (17.79)	58.50_a (11.56)	63.28_a (17.94)
Difficulty (1–9)	4.09_a (2.74)	4.14_a (2.18)	4.50_a (1.58)	3.75_a (2.21)
Confidence (1–9)	5.82_a (2.93)	5.91_a (2.19)	5.10_a (1.97)	6.81_a (2.16)

Note. Means with different subscripts are different at $p < .05$ in a Tukey post hoc test.

d = 5.70, whereas deliberative participants underestimated theirs, one-sample $t(31) = -2.88$, $p = .007$, $d = 1.03$. To compare the degree of misestimation across performance levels, we examined absolute differences (independent of sign) between estimated and actual number of correct answers. Deliberative responders were more accurate than all other participants in general, $t(70) = 5.13$, $p < .001$, $d = 1.23$, and intuitive responders in particular, $t(70) = 7.16$, $p < .001$, $d = 1.71$.

Number of correct answers by others. When asked how many problems most participants answered correctly, deliberative responders made lower estimates than either intuitive responders, $t(25) = 3.08$, $p = .005$, $d = 1.23$, or all other groups combined, $t(55) = 2.07$, $p = .043$, $d = 0.56$.

Number of correct answers by self–others. Deliberative responders estimated a higher number of correct responses for themselves than for others, paired $t(31) = 3.46$, $p = .002$, $d = 0.61$, whereas intuitive responders did not make different estimates for themselves and for others, paired $t(10) = -1.79$, $p = .104$, $d = 0.70$. The difference between the numbers of correct answers estimated for self and others was larger for deliberative responders than for either intuitive responders, $t(70) = 2.84$, $p = .006$, $d = 0.68$, or all other groups combined, $t(70) = 2.76$, $p = .007$, $d = 0.66$.

Percentage of others who solved all problems correctly. Deliberative responders' estimates about the percentage of participants who solved all four problems correctly were lower and closer to the actual percentage (43.2%) than were the estimates of intuitive responders, $t(70) = 2.22$, $p = .030$, $d = 0.53$, or of all other participants combined, $t(70) = 1.71$, $p = .091$, $d = 0.41$.

Better or worse than others. When asked whether they had performed better or worse than others, deliberative responders believed that they had performed better, whereas intuitive responders believed that they were average. Deliberative responders were more confident than intuitive responders, $t(69) = 2.53$, $p = .014$, $d = 0.61$, and all other groups combined, $t(69) = 3.46$, $p = .001$, $d = 0.83$. These comparative estimates were regressed on the estimated number of correct answers for the self and for most others, and both were significant predictors: $\beta_{self} = .56$, $p < .001$; $\beta_{others} = -.46$, $p < .001$; $R^2 = .37$.

Percentile. Percentile estimates did not differ significantly across performance levels (ts < 1.47, ps $> .147$). The mean estimated percentiles fell above the actual percentile range of each group for all but deliberative responders, whose estimates fell in the correct range. Comparison of these estimates to the midpoint of the actual percentile range revealed that intuitive responders overestimated their ranking by 48.91%, one-sample $t(10) = 10.36$, $p < .001$, $d = 6.55$, whereas deliberative responders underestimated theirs by 15.12%, one-sample $t(31) = -4.77$, $p < .001$, $d = 1.71$. To compare the degree of misplacement (over- or underplacement) across performance levels, we considered the absolute differences (positive or negative) between estimated and actual percentile. Deliberative responders were more accurate than both intuitive responders, $t(70) = 6.27$, $p < .001$, $d = 1.50$, and the aggregate of all other participants, $t(70) = 2.83$, $p = .006$, $d = 0.68$. Percentile estimates were regressed on the estimated number of correct answers for the self and for most others, and both were significant predictors: $\beta_{self} = .49$, $p < .001$; $\beta_{others} = -.38$, $p = .001$; $R^2 = .27$.

Others' responses. Across the four syllogisms, 90.9–100% of the intuitive responders guessed that most others had also given the intuitive response, not considering alternative solutions. As for deliberative responders, some believed that others had given the correct answers but other deliberative responders did not. Importantly, these guesses were related to their comparative assessments. When they were asked whether they had performed better or worse than others, deliberative responders who guessed that most others had also responded deliberatively to all four problems estimated that their performance was similar to that of others, whereas deliberative responders who guessed that most others had not responded deliberatively to any of the four problems estimated that they had performed better than others ($M = 0.13$, $SD = 0.35$ vs. $M = 1.13$, $SD = 0.64$), $t(14) = 3.86$, $p = .002$, $d = 2.06$. Likewise, when making percentile estimates, deliberative responders who thought that most others had also responded deliberatively to all four problems ranked themselves close to 50% ($M = 53.13$, $SD = 7.04$), below other deliberative responders who thought that most others had not responded deliberatively to any of the four problems ($M = 63.13$, $SD = 12.23$), $t(14) = 2.00$, $p = .065$, $d = 1.07$.

Difficulty. Experienced difficulty did not differ across performance levels (ts < 1).

Confidence. When asked how confident they were in their answers, deliberative responders displayed greater confidence than other participants in general, $t(70) = 2.21$, $p = .031$, $d = 0.53$. The contrast comparing them specifically with intuitive responders did not reveal a significant difference, $t(70) = 1.25$, $p = .215$, $d = 0.30$.

Discussion

The results of Study 4 largely replicate those of Studies 1 and 3 (in the case of conflict problems) with yet another kind of reasoning task: syllogistic reasoning. Once more, deliberative thinkers were better than intuitive thinkers at most absolute and comparative judgments of performance. This study offered a chance to test whether guesses about what others responded are related to how deliberative responders compare their performance to that of others. According to our hypothesis, deliberative thinkers make comparative assessments that are flattering to themselves not just because they are deliberative but also because they know that many others think only of the intuitive solution. Indeed, deliberative responders who thought that most others had responded deliberatively made lower comparative estimates than deliberative responders who thought that others had not respond deliberatively. Note that comparative estimates were made before the guesses about others' responses. This result suggests that, when participants made those estimates, their guesses about what others might have responded were already in their mind.

The fact that some deliberative responders in this study and in Study 3 did not indicate the intuitive solutions when they were asked to guess what others had responded may appear to be incompatible with a default-interventionist model. This is only the case, however, if we interpret these guesses strictly as a measure of awareness of the intuitive solution, because according to a default-interventionist model we should not expect any deliberative thinker to lack intuition. But the measure asked not what the intuitive solution was but rather what participants thought that most others had responded. It is possible that deliberative responders were aware of the intuitive solution but thought that others were able to produce the deliberative solution. This would mean that these deliberative responders were not unaware of the intuitive alternative but rather were just confident about others' reasoning skills. To assess in a more direct way whether deliberative responders are aware of the intuitive solutions and whether they indeed consider them before giving the deliberative responses, we asked participants in Studies 5 and 6 whether they considered alternative solutions before coming to the response that they gave.

In both this study and Study 3, percentile estimates did not always agree with other comparative measures, such as asking directly whether participants believed that they performed better or worse than others or calculating the difference between their estimates about their own performance and their estimates about others' performance. It is possible that translating an impression of one's performance into an exact percentile is more difficult than other tasks (see also Ehrlinger et al., 2008). Still, the percentile estimates made by deliberative responders were overall more accurate than those made by others, particularly intuitive responders, as predicted.

Study 5

Studies 5 and 6 manipulated each of the main variables in this research—thinking mode (intuitive vs. deliberative) and metacognitive knowledge about the alternative solutions to the problems—and assessed the effect of these manipulations on performance estimates.

Study 5 used a manipulation that was intended to change the thinking mode of participants and then tested whether this change was accompanied by changes in performance estimates. We used an *attention capturer* (Sanford, Sanford, Molle, & Emmott, 2006) to draw attention to the information in the premises of a problem. Attention capturers are stylistic devices that serve to highlight certain parts of an utterance, so that they are attended to in greater detail and processed in greater depth. Examples for written text include underlining, italics, and boldface. In this study, in order to improve performance, we underlined the crucial part of a conflict problem, the one that makes the deliberative solution correct and the intuitive solution incorrect.

Method

Participants. Sixty-five participants were recruited through Amazon's Mechanical Turk (www.mturk.com), which has proven to be a valid resource for research in psychology (see Buhrmester, Kwang, & Gosling, 2011). Participants were located in the United States, and they were required to have an approval rate in previous assignments of at least 95%. Participants received 0.25 U.S. dollars.

Procedure. Participants were given an adapted version of the bat-and-ball problem to solve: *A TV and a DVD together cost 110 dollars. The TV costs 100 dollars more than the DVD. How much does the DVD cost?* (the contents of the bat-and-ball problem of the CRT were changed so that participants would not be able to look up the answers on the Internet; see Goodman, Cryder, & Cheema, in press). After they had answered this problem, participants were asked (a) what percentage of people in general answer this problem correctly (this measure will be referred to as the *percentage of others who solved the problem correctly*) and (b) how well people in general perform on this problem (1 = *very poorly* to 9 = *very well*; *performance rating*).

Then participants were informed that they would be seeing the same problem again, so that, if they wanted to, they could change their initial answer. They were told that, if they decided not to change their initial answer, they could give the same answer as before. At this time, the problem was presented again, but this time with the conflict part in the second premise underlined: *A TV and a DVD together cost 110 dollars. The TV costs 100 dollars more than the DVD*. How much does the DVD cost? After they solved the problem for the second time, participants were again asked to estimate how well other people in general do on this problem, using the same two items as before.

Finally, the survey assessed participants' *awareness of alternative responses*: "Did the solution 10 dollars come to your mind at any time while you were trying to solve the problem?" (yes or no), and "Did the solution 5 dollars come to your mind at any time while you were trying to solve the problem?" (yes or no).

Results and Discussion

Performance. A comparison of the rate of participants who gave the correct answer in Times 1 and 2 revealed that performance improved from 41.7% to 60% (McNemar's $p = .002$).

Estimates. Before we tested whether changes in performance were accompanied by the expected changes in estimates, we examined the estimates made at Time 1 to test if, just as in Study 2, the confidence of deliberative responders and intuitive responders differs at the level of specific problems. Indeed, for both measures, the estimates made at Time 1 about the ability of people in general to solve the bat-and-ball problem (here adapted to TV-and-DVD) were higher for those participants who responded intuitively than for those who responded correctly (ts > 4.27, ps $< .001$, ds $>$ 1.25).

Using mixed model ANOVA, we analyzed estimates and ratings at Times 1 and 2 as a function of how participants responded at Times 1 and 2 (see Table 5).

For the estimated *percentage of others who solved the problem correctly*, there was a main effect of time, $F(1, 55) = 10.97$, $p = .002$, $\eta_p^2 = .17$, with higher estimates at Time 1 than at Time 2 ($M = 71.64$ vs. $M = 64.94$). There was also a main effect of type of responding at Times 1 and 2, $F(2, 55) = 10.64$, $p < .001$, $\eta_p^2 = .28$, with the highest estimates made by those who responded intuitively both times ($M = 85.02$), followed by those who responded intuitively at first but then changed to the correct response ($M = 62.25$), and those who responded correctly both times ($M = 57.60$). Pairwise comparisons revealed that those who responded intuitively both times estimated higher percentages than each of the other two groups did (ps $< .006$); the other two groups did not differ from each other ($p = .561$); no participant changed from the correct to the intuitive solution. More important, there was a time × type of responding interaction, $F(2, 55) = 25.24$, $p < .001$, $\eta_p^2 = .48$. Planned comparisons revealed that the decrease in estimates from Time 1 to Time 2 was observed only for participants who changed from the intuitive to the correct response ($M_1 = 77.00$, $SD = 28.50$ vs. $M_2 = 47.50$, $SD = 23.00$), paired $t(9) = 3.68$, $p = .005$, $d = 1.19$. Those who responded correctly

Table 5
Mean Estimates and Standard Deviations (SDs) at Times 1 and 2 by Type of Responding (Study 5)

	Type of responding at Times 1 and 2					
	Intuitive–intuitive		Intuitive–correct		Correct–correct	
Estimate (range)	Time 1	Time 2	Time 1	Time 2	Time 1	Time 2
Percentage of others who solved the problem correctly	84.13 (15.67)	85.91 (14.35)	77.00 (28.50)	47.50 (23.00)	53.80 (27.36)	61.40 (25.15)
Performance rating (1–9)	7.74 (1.39)	7.78 (1.20)	6.80 (2.15)	4.30 (1.83)	5.12 (2.15)	5.84 (2.08)

both times made higher estimates at Time 2 than at Time 1 (M_1 = 53.80, *SD* = 27.36 vs. M_2 = 61.40, *SD* = 25.15), paired $t(24) = -2.63$, $p = .015$, $d = 0.53$. Those who responded intuitively both times also made higher estimates at Time 2 than at Time 1 (M_1 = 84.13, *SD* = 15.67 vs. M_2 = 85.91, *SD* = 14.35), paired $t(22) = -2.41$, $p = .025$, $d = 0.54$ (if we consider all incorrect responses and not just intuitive responses, this difference is not significant, $t < 1$).

For *performance ratings*, there was a main effect of time, $F(1, 55) = 10.96$, $p = .002$, $\eta_p^2 = .17$, with higher ratings at Time 1 than at Time 2 ($M = 6.55$ vs. $M = 5.97$). There was a main effect of type of responding at Times 1 and 2, $F(2, 55) = 12.25$, $p < .001$, $\eta_p^2 = .31$, with the highest ratings made by those who responded intuitively both times ($M = 7.76$), followed by those who changed from the intuitive to the correct response ($M = 5.55$), and those who were consistently correct ($M = 5.48$). Pairwise comparisons revealed that consistently intuitive responders made higher ratings than the other two groups did (*p*s < .001); the other two groups did not differ from each other ($p = .913$). Again, there was a trial × type of responding interaction, $F(2, 55) = 24.96$, $p < .001$, $\eta_p^2 = .48$. Planned comparisons revealed that the decrease in the ratings from Time 1 to Time 2 was observed only for participants who changed from the intuitive to the correct response (M_1 = 6.80, *SD* = 2.15 vs. M_2 = 4.30, *SD* = 1.83), paired $t(9) = 3.34$, $p = .009$, $d = 1.06$. Those who responded correctly both times made higher ratings at Time 2 than at Time 1 (M_1 = 5.12, *SD* = 2.15 vs. M_2 = 5.84, *SD* = 2.08), paired $t(24) = -3.27$, $p = .003$, $d = -0.66$. Those who responded intuitively both times did not change their ratings ($t < 1$).

For each measure, the absolute differences in ratings/estimates between the first and the second trials were examined in order to assess the amount of change in estimates/ratings regardless of whether that change was a decrease or an increase. For both measures, planned contrasts comparing participants who changed from the intuitive to the correct response to those who did not change their answer confirmed that participants who changed answers revised their estimates to a greater extent (*t*s > 2.98, *p*s < .015, *d*s > 1.90).

At Time 1, those who responded intuitively at first but who would then change to the correct solution expressed more confidence in their estimates and ratings than did those who responded correctly from the beginning (*t*s > 2.09, *p*s < .044, *d*s > 0.73). But at Time 2, the estimates and ratings made by those who changed from intuitive to correct were not different from the ratings that had been made at Time 1 by those who responded correctly from the beginning (*t*s < −1.06, *p*s > .296). This suggests that the improvement in performance led participants who changed from intuitive to correct to make estimates similar to those made by participants who responded correctly from the beginning.

Awareness of alternative responses. When asked whether the solutions 10 dollars and 5 dollars had come to their mind while they were trying to solve the problem, most participants (72.2%) who gave the deliberative response at Time 2 indicated that they had thought of the intuitive solution; this included both 64% of consistently correct responders and 90% of those who changed from intuitive to correct. In the case of those who changed from intuitive to correct, this high percentage was expected, as these participants had just given the alternative intuitive answer in the previous trial, and therefore they had to have been aware of it. This result is mostly informative for deliberative responders who did not give the intuitive response at any point and yet proved to be aware of it. Only one intuitive responder (4.2%) indicated having thought of the deliberative solution.

Being aware of the alternative solution or not predicted the estimates at Time 2, with participants who indicated having thought of the alternative making lower ratings and estimates than those who did not show awareness of the alternative (*t*s > 2.44, *p*s < .018, *d*s > 0.62), and it also predicted the absolute differences in estimates from Time 1 to Time 2, with larger differences in the estimates and ratings made by those who proved to be aware of the alternative (*t*s > 3.67, *p*s < .001, *d*s > 0.94).

Summary. Participants who at first responded intuitively to the TV-and-DVD problem were confident about the ability of other people in general to solve this problem. Of those participants, some changed to the correct, deliberative solution. The participants who improved their performance across trials became less confident about how well people in general perform on this problem. Their changed estimates were similar to the ones made by participants who recognized the correct, deliberative solution from the beginning.

Participants who changed from the intuitive to the correct solution also became more aware of alternative responses. Even though their awareness of possible alternative responses was measured only after the second trial (we did not measure it after the first trial because this might have influenced their performance in the second trial; i.e., asking "Did the solution 5 dollars come to your mind?" might have made participants realize that this is the correct solution), the results of the previous studies suggest that intuitive responders are generally unaware of the correct, deliberative alternative. Therefore, when these participants changed their answer to the correct solution, they necessarily gained awareness that there are alternative solutions to the problem: a compelling but incorrect solution and a correct, deliberative solution.

Even though participants who changed from the intuitive to the deliberative solution were the ones who changed their estimates more drastically, participants who did not change solutions sometimes also changed their estimates. They did so in the opposite direction to those who changed solutions; that is, they became more favorable in their estimates of others' performance. This result was not consistent for participants who responded intuitively on both trials (they increased their percentage estimates but not performance ratings), but it was consistent for participants who responded deliberatively on both trials. It is possible that these participants made the second estimates not about the TV-and-DVD problem per se but rather about the specific version of the problem where the conflict part was underlined; therefore, they might expect people to do better (which they did).

Finally, one might argue that participants who changed from the intuitive to the deliberative solution lowered their estimates not because they became aware of a different solution but because they were made aware that they had made a mistake in the first trial. Therefore, their lowered estimates might have been driven by a motivation to make their initial error seem more prevalent and normal (see Sherman, Presson, & Chassin, 1984). Apart from its main objective—to manipulate the metacognitive awareness of alternative solutions—the following study also tests this alternative explanation by controlling for whether people are asked to respond to the problems or not.

Study 6A

This study used a manipulation that was intended to change the metacognitive awareness of alternative solutions. Participants first completed the CRT and made performance estimates. After that, participants were informed either about the intuitive solutions or about the deliberative solutions. Then, participants were again asked to make performance estimates. If intuitive responders are aware only of the intuitive solutions, whereas deliberative responders are aware of both the intuitive and the deliberative solutions, then showing solutions different than the ones they gave should affect the estimates of intuitive responders more than the estimates of deliberative responders.

Method

Participants. One hundred forty-two participants were recruited at the University of Lisbon. Participants received a gift card in the value of 5 euros.

Procedure. Participants were given the three CRT problems to solve, and they were then asked to estimate how well people in general perform on these problems. in particular, they were asked (a) How many of those 3 problems do you think most people answer correctly? (this measure will be referred to as *Number of correct answers by others*); (b) What percentage of people answer correctly to all 3 problems? (*Percentage of others who solved all problems correctly*); (c) Indicate how good you think is the performance of people in general on those 3 problems (*Performance rating*; 1 = *very poor* to 9 = *very good*).

Then, participants were assigned either to the intuitive-solutions condition, where they were told that they would be seeing possible answers to the problems that they had just solved, or to the deliberative-solutions condition, where they were told that they would be seeing the solutions to the problems that they had just solved. In the intuitive-solutions condition, participants were shown the intuitive solutions to every problem. In the deliberative-solutions condition, they were shown the deliberative solutions (see Appendix D). After that, participants were again asked to estimate the performance of people in general on the CRT problems using the same measures as before.

In the intuitive-solutions condition, there were additional measures pertaining to participants' estimates and ratings of their own performance, and not just others' performance (these measures were not presented in the deliberative-solutions condition because in that condition participants were told that the presented solutions were the solutions to the problems, and therefore it would be only natural that intuitive responders would become less confident about their performance): (d) How many of these 3 problems do you think you answered correctly? (*Number of correct answers by self*); (e) Indicate how good you think your performance was on these 3 problems in comparison to other people in general (*Better or worse than others*; 1 = *much worse than most other people* to 9 = *much better than most other people*); and (f) How confident are you that the answers that you gave to the 3 problems are correct? (*Confidence*; 1 = *not at all confident* to 9 = *very confident*). These measures were presented before and after participants were shown the intuitive solutions. The question of interest here was whether deliberative responders would change their estimates and ratings after seeing responses different than the ones they gave.

Any changes in estimates that intuitive responders make in the deliberative-solutions condition might be a simple consequence of the instructions in that condition directly proving them wrong. The instructions might lead them to make lower estimates for motivational purposes, so that their error would seem more prevalent and therefore more normal and acceptable (Sherman et al., 1984). To control for this alternative explanation, we assigned 30 participants to the same condition where they were shown the deliberative responses, but they were not asked to respond to the problems. They were simply asked to read the problems and to think about what they would respond if they were asked to do so. Just as the other participants did, these participants made estimates both before and after they were shown the solutions.

Finally, in both conditions there was a question that assessed participants' *awareness of alternative solutions*: "As you started to solve the problems, were there any other answers that came to your mind before you decided on the answers that you gave? Now you are going to see the same three problems again, but this time, instead of repeating the responses that you gave before, we want you to provide any answers that came to your mind before you decided on the answers that you gave. If you did not think of other answers, you do not have to say anything."

Results and Discussion

Performance. Forty-three percent of the participants did not answer any problem correctly (the intuitive responders); 18.5% answered only one problem correctly; 17.3% answered two problems correctly; and 21% answered all three problems correctly (the deliberative responders).

Estimates. Table 6 shows estimates and ratings as a function of estimation time (before vs. after participants were shown the solutions), performance (number of correct vs. incorrect/intuitive responses), and condition (whether they were shown the intuitive or the deliberative solutions). To test our hypothesis, we will

Table 6
Mean Estimates and Standard Deviations (SDs) Before and After the Solutions Were Shown by Number of Correct Responses and Type of Solutions Shown (Study 6A)

		Actual number of correct answers							
		0		1		2		3	
Type of solution	Estimate (range)	Before	After	Before	After	Before	After	Before	After
Deliberative	Number of correct answers by others (1–3)	2.25 (0.97)	1.20 (0.77)	1.60 (0.55)	1.00 (0.71)	1.75 (0.71)	1.75 (0.71)	1.57 (0.98)	1.71 (0.95)
	Percentage of others who solved all problems correctly	70.40 (29.98)	34.50 (26.64)	38.20 (28.66)	28.20 (32.69)	39.63 (25.29)	34.13 (24.42)	30.00 (23.98)	32.86 (22.33)
	Performance rating (1–9)	7.05 (1.76)	4.15 (2.23)	5.60 (1.52)	4.40 (2.51)	5.13 (1.73)	5.38 (1.77)	5.00 (1.41)	5.00 (1.00)
Intuitive	Number of correct answers by others (1–3)	2.20 (0.68)	2.53 (0.64)	2.00 (0.47)	1.90 (0.57)	1.50 (0.55)	1.33 (0.52)	1.40 (0.70)	1.20 (0.79)
	Percentage of others who solved all problems correctly	65.87 (27.03)	73.87 (24.80)	38.80 (30.25)	42.30 (31.01)	21.17 (27.54)	22.83 (29.50)	25.10 (15.35)	22.20 (12.58)
	Performance rating (1–9)	6.40 (1.72)	6.87 (1.92)	5.80 (1.69)	5.30 (2.00)	4.67 (2.34)	4.67 (2.25)	5.00 (1.25)	4.40 (1.08)
	Number of correct answers by self (1–3)	2.53 (0.64)	2.67 (0.82)	2.40 (0.52)	2.10 (0.74)	2.67 (0.52)	2.83 (0.41)	2.80 (0.42)	2.80 (0.42)
	Better or worse than others (1–9)	6.00 (2.00)	6.13 (2.17)	6.10 (1.52)	5.90 (1.60)	7.00 (1.67)	7.17 (1.33)	7.50 (0.97)	7.70 (0.82)
	Confidence (1–9)	7.20 (1.82)	8.00 (2.04)	6.50 (2.27)	6.50 (1.84)	7.67 (0.82)	8.00 (0.63)	8.00 (1.05)	8.00 (1.05)

compare only intuitive versus deliberative responders, but the pattern of results is the same if deliberative responders are compared to all other responders.

A consistent pattern of results emerged such that only intuitive responders changed their estimates, and only when they saw the deliberative solutions, yielding significant time × performance × condition interactions for all three measures that were common to both conditions (Fs > 7.07, ps $< .011$, $\eta_p^2 > .13$). As expected, deliberative responders did not change their estimates when they were shown the deliberative solutions (ts ≤ 1.00, ps $\geq .356$). More important, they also did not change their estimates when they were shown the alternative intuitive solutions: Five out of six of their estimates did not change (ts ≤ 1.72, ps $\geq .120$); there was only a difference for performance ratings, with lower ratings after versus before the solutions were shown, paired $t(9) = 2.25$, $p = .051$, $d = 0.72$. Intuitive responders did not change their estimates when they were shown the intuitive solutions (ts ≤ 1.44, ps $\geq .173$). (When all incorrect responses and not just intuitive responses are considered, there is one difference in one of the measures: The estimated number of correct answers by others was higher at Time 2 than at Time 1 ($M_1 = 2.05$, $SD = 0.65$ vs. $M_2 = 2.41$, $SD = 0.66$), paired $t(21) = -2.16$, $p = .042$, $d = 0.94$.) But they lowered all their estimates and ratings after they were shown the alternative deliberative solutions (ts > 3.80, ps $< .001$, ds > 1.21). The absolute differences from Time 1 to Time 2 were larger for intuitive responders than for deliberative responders for all estimates and ratings (ts > 4.52, ps $< .001$, ds > 1.37).

When we compared the estimates made by participants in the deliberative-solutions condition who solved the problems to those made by the participants in the same condition who were not asked to respond to the problems, both before and after they were shown the solutions, the pattern of results is the same: higher estimates or ratings before versus after seeing the deliberative solutions (Fs > 25.15, ps $< .001$, $\eta_p^2 > .23$); responding versus not responding to the problems did not have any significant effect (Fs < 1.12, ps $> .294$). Thus, the difference in estimates before versus after seeing the deliberative solutions seems to be attributable to intuitive participants' lack of awareness of the deliberative solutions and not to the fact that participants were shown to have made a mistake and tried to make their mistake seem more prevalent and normal.

In the deliberative-solutions condition, the estimates that intuitive and deliberative responders made at Time 1 were always different, with intuitive responders making higher estimates (ts $>$ 1.74, ps $< .089$, ds > 0.58; this difference was marginally significant for only one of the measures). But none of the estimates made by intuitive responders at Time 2 were different from the ones that deliberative responders had made at Time 1 (ts < 1.03, ps $> .314$). This result suggests that the enhanced metacognition that came from seeing the deliberative solutions led intuitive responders to make estimates similar to the ones that deliberative responders made at first.

Awareness of alternative responses. When participants were asked to indicate any answers that came to their mind before they decided on the answers, many did not respond for any of the problems (61.7–76.5%). When they did respond, 75–90.9% of deliberative responders indicated the intuitive solutions across the three problems, whereas only 0–37.5% of intuitive responders indicated the deliberative solutions.

Summary. When deliberative responders were shown the intuitive solutions, they did not change their estimates, whereas when intuitive responders were shown the deliberative solutions they did revise their estimates.

Study 6B

This study replicated Study 6A with the following changes. First, greater care was taken than in the previous study to make the two conditions equivalent. In that study, the deliberative solutions were presented as "the solutions" whereas the intuitive solutions were presented as "possible answers." This difference could contribute to the greater impact of showing the deliberative solutions. In this study, both the deliberative and the intuitive solutions were presented as "possible solutions." Also, in Study 6A, the explanations for the deliberative solutions were longer (see Appendix D) and thus perhaps more convincing than the explanations for the intuitive solutions. In Study 6B, the explanations for the two types of solutions had the same number of words. Finally, in the previous study, participants in the deliberative condition made estimates and ratings only about other people in general, whereas participants in the intuitive condition made estimates and ratings both about themselves and about other people in general. In this study, participants in both conditions rated their performance and that of other people in general.

Method

Participants. Eighty participants were recruited through Amazon's Mechanical Turk. Participants were located in the United States, and they were required to have an approval rate in previous assignments of at least 95%. Participants received 0.25 U.S. dollars.

Procedure. Participants were given the three CRT problems to solve (as in Study 5, the contents of the CRT problems were altered so that participants would not be able to look up the answers on the Internet; see Appendix E). Then they were asked to answer the following questions: (a) How many of those 3 problems do you think you answered correctly? (this measure is referred to as *Number of correct answers by self*); (b) And how many of those 3 problems do you think other people in general answer correctly? (*Number of correct answers by others*); (c) What percentage of people in general answer correctly to all 3 problems? (*Percentage of others who solved all problems correctly*); (d) Indicate how good you think your performance was on those 3 problems in comparison to other people in general (*Better or worse than others*; 1 = *much worse than most other people* to 9 = *much better than most other people*); (e) How confident are you that your initial answers to the three problems are correct? (*Confidence*; 1 = *not at all confident* to 9 = *very confident*).

Then, participants were told, "Now you will see possible solutions to the problems that you have just solved." In one condition, they were shown the intuitive solutions. In another condition, they were shown the deliberative solutions (see Appendix E). Afterward, participants were again asked to make performance estimates and confidence ratings with the measures used earlier in this study.

Results and Discussion

Performance. Twenty-four percent of the participants did not answer any problem correctly (the intuitive responders); 23.8% answered only one problem correctly; 9.5% answered two problems correctly; and 42.9% answered all three problems correctly (the deliberative responders).

Estimates. Table 7 shows the mean estimates and ratings as a function of estimation time (before vs. after participants were shown the solutions), performance (number of deliberative vs. intuitive responses), and condition (whether they were shown the intuitive or the deliberative solutions). As in Study 6A, we compared only intuitive versus deliberative responders, but, again, the pattern of results is the same if consistently deliberative responders are compared to all other responders.

A consistent pattern of results emerged such that only intuitive responders became less confident after they saw alternatives to the solutions they had presented, yielding significant time × performance × condition interactions for all five measures (Fs > 5.62, ps $< .023$, $\eta_p^2 > .13$). As expected, deliberative responders did not change their estimates when they were shown the deliberative solutions, and there was only a marginally significant difference for confidence, with higher ratings after versus before the solutions were shown, paired $t(10) = -2.21$, $p = .052$, $d = 0.76$; for the other four measures, ts ≤ -1.00, ps $\geq .341$. More important, they also did not change four out of five of their estimates when they were shown the alternative intuitive solutions (ts ≤ 1.46, ps $\geq .164$); only their estimates about the percentage of correct responders decreased, paired $t(15) = 2.53$, $p = .023$, $d = 0.72$. When intuitive responders were shown the intuitive solutions, they did not change either their estimates of the number of correct responses by self or others, or their estimates of whether they performed better or worse than others (ts < -1.87, ps $> .104$). But they estimated a higher percentage of correct responders, paired $t(7) = -3.67$, $p = .008$, $d = 1.32$, and they became more confident, paired $t(7) = -3.64$, $p = .008$, $d = 2.41$; the last difference does not hold if all consistently incorrect responders and not just intuitive responders are considered, $t(9) = -1.67$, $p = .129$. More important, intuitive responders lowered four out of five estimates after they were shown the alternative deliberative solutions (ts > 2.25, ps $< .066$, ds > 0.88; if all incorrect responses are considered, ts > 3.50, ps $< .017$); only their estimates of the percentage of people who answer correctly did not change significantly ($p = .118$), even though these too were lower after the participants saw the solutions.

The absolute differences from Time 1 to Time 2 were larger for intuitive responders than for deliberative responders for all estimates and ratings (ts > 1.90, ps $< .077$, ds > 0.98), except for the better-or-worse estimate, $t(40) = 1.11$, $p = .274$. When all incorrect responses are considered, these differences were larger for consistently incorrect responders than for deliberative responders for all the measures (ts > 1.94, ps $< .056$).

In the deliberative-solutions condition, at Time 1, intuitive responders estimated a higher number of correct responses for others and a higher percentage of correct responders than deliberative responders did (ts > 2.50, ps $< .017$, ds > 0.79); considering all incorrect responses, the difference in estimated percentages is not significant, $t(15) = 1.57$, $p = .131$. But at Time 2 the estimates made by intuitive responders were not different from those made

Table 7
Mean Estimates and Standard Deviations (SDs) Before and After the Solutions Were Shown by Number of Correct Answers and Type of Solutions Shown (Study 6B)

		Actual number of correct answers							
		0		1		2		3	
Type of solution	Measure (range)	Before	After	Before	After	Before	After	Before	After
Deliberative	Number of correct answers by self (1–3)	2.57 (0.53)	1.00 (1.29)	2.33 (0.52)	1.33 (0.52)	2.67 (0.58)	2.33 (0.58)	2.91 (0.30)	3.00 (0.00)
	Number of correct answers by others (1–3)	2.29 (0.76)	1.14 (1.22)	2.00 (0.63)	1.50 (0.55)	2.00 (0.00)	1.67 (0.58)	1.55 (0.52)	1.64 (0.51)
	Percentage of others who solved all problems correctly	51.86 (40.69)	30.00 (32.15)	55.00 (32.71)	40.83 (29.23)	35.00 (21.79)	40.00 (20.00)	34.82 (19.11)	36.18 (21.13)
	Better or worse than others (1–9)	4.86 (1.22)	3.86 (1.57)	5.33 (0.52)	4.17 (0.98)	5.33 (1.53)	5.00 (1.73)	6.45 (0.52)	6.55 (0.52)
	Confidence (1–9)	7.43 (1.81)	3.71 (3.40)	7.50 (0.55)	4.17 (1.84)	7.00 (2.00)	5.33 (4.04)	8.36 (0.81)	8.91 (0.30)
Intuitive	Number of correct answers by self (1–3)	2.88 (0.35)	3.00 (0.00)	2.22 (0.97)	1.89 (0.60)	2.33 (0.58)	1.33 (0.58)	2.88 (0.34)	2.75 (0.78)
	Number of correct answers by others (1–3)	2.50 (0.54)	2.50 (0.54)	2.00 (0.50)	1.78 (0.44)	2.00 (1.00)	1.67 (0.58)	1.94 (0.77)	1.81 (0.91)
	Percentage of others who solved all problems correctly	74.88 (11.59)	83.50 (10.18)	52.78 (27.51)	48.33 (19.20)	50.67 (39.00)	39.33 (31.01)	49.00 (24.80)	45.06 (27.64)
	Better or worse than others (1–9)	5.00 (1.20)	5.50 (0.76)	4.89 (1.83)	4.56 (1.67)	5.33 (1.16)	3.33 (1.16)	5.88 (0.96)	5.81 (1.22)
	Confidence (1–9)	6.63 (2.20)	8.50 (0.93)	6.22 (2.73)	5.67 (2.83)	7.00 (0.00)	3.00 (0.00)	8.38 (0.96)	8.13 (1.59)

by deliberative responders at Time 1 (ts < 1.50, ps $> .148$). This suggests that the double awareness that came from seeing the deliberative solutions led intuitive responders to make estimates similar to the ones that deliberative responders (who presumably had that double awareness all along) made at first.

Summary. Intuitive responders changed most of their estimates when they were shown alternative solutions, whereas deliberative responders did not.

General Discussion

Across seven studies using different reasoning and judgment tasks, a consistent pattern suggesting that deliberative thinkers have a metacognitive advantage over intuitive thinkers emerged. In Studies 1–4, deliberative thinkers were more accurate than intuitive thinkers in the absolute assessments that they made about both their own performance and that of others; they were also more accurate in the comparative assessments that they made about their performance versus that of others; and their confidence levels were more realistic than those expressed by intuitive thinkers. Confirming the role of type of thinking in metacognitive awareness, in Study 5, those participants who changed from an intuitive to a deliberative mode of thinking also changed their impressions of how competent people are in thinking about reasoning problems. Finally, Studies 6A and 6B showed that deliberative and intuitive responders indeed do not have the same metacognitive awareness. Consistent with the hypothesis that deliberative responders are aware of both their deliberative solutions and the alternative intuitive solutions, whereas intuitive responders are aware only of the intuitive solutions that they give, when deliberative responders were shown the intuitive solutions, they did not change their estimates about how well they and others perform. However, intuitive responders did change their estimates when they were shown the deliberative solutions; their estimates became similar to the ones that deliberative responders made at first. These results have several implications, which we now discuss.

Judgments of Performance

Deliberative problem solvers were more accurate than their intuitive peers in their estimates of their own performance, and they also made more accurate estimates of others' performance. The former finding is not new. It is the Dunning–Kruger effect, whereby the skilled are better at assessing their performance than the unskilled (e.g., Ehrlinger et al., 2008; Moore & Healy, 2008). However, in previous research, skilled performers are not as good at estimating others' performance. Often, they demonstrate a false consensus bias (Ross et al., 1977) and believe that many others performed as well as they did (e.g., Kruger & Dunning, 1999). In our studies, deliberative responders (the skilled) were less likely to show false consensus. When asked what others answered, deliberative respondents were aware of solutions that were different from their own, whereas intuitive respondents simply repeated their solutions. To the best of our knowledge, previous research has not shown this metacognitive advantage such that the skilled have a precise notion of alternative responses that others might have given (e.g., saying that others answered 10 instead of 5 in the bat and ball problem).

Because deliberative responders had a much better sense of how good their performance was, as well as how good others' performance was, they also had a clear advantage over their intuitive peers when it came to making comparative estimates. In problems where the deliberative and the intuitive solutions differed, deliberative responders accurately estimated that they had done better than others, presumably because they were aware of the intuitive response and aware that most others might have given it, although we did not test this causal process directly. However, in problems where the deliberative and the intuitive solutions were the same (Study 3), deliberative responders lowered their comparative assessments, most likely because they knew that it was not necessary to think deliberatively to succeed in these problems. Intuitive thinkers, on the other hand, oblivious to the deliberative solutions, did not change their comparative assessments depending on whether the deliberative and the intuitive solutions agreed or not. Presumably, they could think only of the intuitive solutions, and they believed that most others also provided these solutions. Consequently, they believed that they had performed as well as others, which was overly confident.

Also consistent with the hypothesis that deliberative responders have a greater metacognitive awareness than their intuitive peers, in Studies 6A and 6B, deliberative responders did not change their estimates when they were shown the intuitive solutions, but intuitive responders did change their estimates when they were shown the deliberative solutions. These results stand in contrast to those found by Kruger and Dunning (1999). In their Study 3, when poor performers were shown alternative responses that other people gave, they did not change their performance estimates, because they were not able to recognize that some of those alternative responses were better than the ones they had given. But seeing the alternative responses that their peers had given made the better performers realize that they had excelled even more than what they thought at first. On the other hand, in our Studies 6A and 6B, intuitive responders were able to recognize the correctness of the deliberative solutions, which made them less confident, whereas deliberative responders were aware all along that there were compelling but incorrect alternative responses and therefore did not change their estimates when they were shown these alternatives.

Unskilled and Unaware, Skilled and Aware

Our results argue against the skilled-or-unskilled-but-still-unaware perspective, at least where deliberative versus intuitive judgment and problem solving are concerned. Proponents of this perspective argue that "the skilled and the unskilled are similarly limited in judging how their performance compares with that of others" (Burson et al., 2006, p. 75; see also Krueger & Mueller, 2002). According to this perspective, people are generally confident and some happen to be accurate in their confident estimates whereas others are not, but the accuracy of the skilled is fortuitous, just as a broken clock shows the right time twice a day. In our studies, we found several results that argue against this perspective. First, deliberative (the skilled) and intuitive responders (the unskilled) were not confident to the same degree. Deliberative responders were overall more confident than intuitive responders. In fact, both in direct measures, such as estimating whether they performed better or worse than others, and in indirect measures, such as calculating the difference between the number of correct responses estimated for the self versus others, intuitive responders did not place themselves above others. Therefore, it was not the case that everyone made similar estimates, merely influenced by the desire to rank above average, and some happened to be accurate whereas others were not. Second, the fact that, in Study 3, deliberative responders made different comparative estimates when deliberative reasoning and intuition were in conflict versus when they were not also suggests that their metacognitive accuracy reflected real awareness of the alternative solutions rather than simply being fortuitous.

Just as the unskilled are said to suffer from a "double curse" (Dunning et al., 2003), not knowing how to perform well and not knowing that they do not know how to perform well, it can also be said that, when deliberation and intuition are at odds, deliberative responders benefit from a double blessing: They know how to avoid compelling but incorrect responses, and, because they realize what these potential errors are and that many others might not be able to avoid them, they are also in a privileged position to compare themselves to others.

Comparative Judgments Need Not Be Egocentric

Our results also differ from previous research with regard to the egocentric nature of comparative judgments and the effect that difficulty has on such judgments. Several studies have found that people tend to overweigh their estimated performance when judging their relative standing or skills and that they fail to take into account how well others are likely to perform (e.g., Chambers & Windschitl, 2004; Klar & Giladi, 1999; Kruger, 1999; Kruger & Burrus, 2004; Windschitl, Kruger, & Simms, 2003). This makes sense in domains where there is no good direct cue to how others performed (see Dunning, 2011), and people can infer this only from how easy it was for themselves to perform. In our studies, however, estimates about others, and not just about one's own performance, were found to predict comparative assessments, and there were no differences in experienced difficulty across performance levels.

A possible reason for this difference is that, in comparison to previous research, ours used problems where it was easier for people to make informed guesses about how most others performed, even when, in the case of intuitive participants, these guesses were wrong. Intuitive thinkers thought of responses that were so intuitive and compelling that they believed that most others must have thought of them too. Deliberative thinkers were able to go beyond those intuitive initial responses, but presumably because they realized how compelling these were, they imagined that most others must have thought only of these responses. In Study 3, it is the difference in the estimates about others that explains why deliberative responders changed their comparative assessments from conflict to no-conflict problems.

Also, unlike previous research, difficulty does not seem to account for our results. There were no differences in experienced difficulty between poor and better performers. One might expect deliberative responders to have felt greater difficulty than intuitive responders because their solutions were more demanding. However, the reasoning that is necessary to come up with the correct answers to these problems is not particularly hard; what is hard is overcoming the highly compelling intuitive answers. Our measure probably failed to capture this difficulty.

Correct Versus Incorrect Metacognition

According to our hypothesis, it is not simply engaging in deliberative reasoning that grants deliberative thinkers a metacognitive advantage over intuitive thinkers; it is being both aware that there is a deliberative solution and a different, intuitive solution. Thus, the metacognitive advantage of deliberative responders is also dependent on having intuition. If deliberative thinkers are not aware of the intuitive solutions, or if the intuitive response was not compelling enough for them to believe that others provided them, they should not place themselves above others. Indeed, in Studies 3 and 4, deliberative responders made higher comparative estimates when they estimated that most others did not respond deliberatively to any problem than when they estimated that most others responded deliberatively to all problems as they did. In this regard, the findings in this research do not favor either the idea that deliberation is better, because intuition leads to judgment errors (e.g., Tversky & Kahneman, 1974), or that intuition is better and can outperform deliberation (e.g., Dijksterhuis, Bos, Nordgren, & van Baaren, 2006). The metacognitive advantage comes from having both intuition and deliberation.

Our results also suggest that the metacognitive advantage of deliberative thinkers over their intuitive peers is not about having versus not having metacognition but rather about having correct versus incorrect metacognition. Indeed, both deliberative and intuitive responders were fairly consistent in their absolute and comparative estimates, comparing themselves to others according to how they believed that others performed. Intuitive responders believed that most others gave the same answers that they did and, thus, that others had a good performance, similar to theirs. Their answers were miscalibrated because the basis for their metacognition was wrong, but their metacognitive judgments were internally consistent. Thus, it is wrong metacognition and not lack of metacognition that distinguishes intuitive problem solvers from their deliberative peers.

Confidence

Deliberative participants' greater metacognitive skills also emerged in their confidence ratings, which were both more realistic and higher than those expressed by intuitive responders. We believe that the great confidence shown by deliberative participants comes from their awareness that there is a right answer and a wrong (albeit highly compelling) intuitive answer, and that their solution is the right one. If we are correct, deliberative participants' confidence is based on a knowledge of what is right and wrong, rather than a feeling of rightness (FOR; Thompson, 2009). According to Thompson and colleagues (Thompson, 2009; Thompson et al., 2011), intuitive responses come with an associated FOR that signals whether additional deliberative thinking is necessary or whether responders can trust their initial intuitions. One way to integrate their findings with ours is to think of them as dealing with different forms of metacognition that come into play at different stages. FOR is a metacognitive experience that comes with the initial intuitive response and can influence the processes that determine the final response. The metacognitive advantage of deliberative responders that was explored in our research is post response. It is not decisive to choose which response to give but rather to assess the quality of that response. However, responders can come to feel this deliberative confidence only if they have reasoned beyond their initial intuition, and that is likely to be determined by the initial FOR. Future research should explore the interplay between the different modes of thinking and these different metacognitive experiences.

Models of Conflict in Dual Process Theories of Thinking and Reasoning

Our results also have implications for how the different modes of thinking are believed to operate. We based our hypothesis on default-interventionist models (e.g., Evans, 2006; Kahneman, 2011; Kahneman & Frederick, 2002; Stanovich & West, 2000; Thompson, 2009; Thompson et al., 2011), according to which intuition is the default and deliberative reasoning is an option that can come into play or not, but only after the intuitive response has emerged. However, other authors (e.g., Epstein, 1994; Sloman, 1996) endorse *parallel-competitive* models according to which intuitive and deliberative thought processes are simultaneously activated and compete for the response. Because the two processes are activated at all times, when there is a conflict between them, people realize it. If people come up with the intuitive solution rather than the deliberative solution, it is not because they failed to detect the conflict but rather because they failed to inhibit their compelling intuitions. And there are also *preemptive conflict resolution* models (see Evans, 2007), according to which problems are analyzed or judgments made either through intuitive processing or through deliberative deliberation but not both at the same time.

Our results are more consistent with a default-interventionist model than with any of the other two kinds of model. According to a preemptive conflict resolution model, people should not be cognizant of both the deliberative and the intuitive alternatives, whereas according to a parallel-competitive model, everyone should be cognizant of both alternatives; therefore, deliberative responders should not have privileged information of which intuitive responders would be unaware. However, whereas most of the deliberative responders in our studies were aware of the intuitive solutions (contrary to a preemptive conflict resolution model), very few intuitive responders proved to be aware of the deliberative solutions (contrary to a parallel-competitive model). Also contrary to a parallel-competitive model is the fact that, in Study 3, deliberative responders made different estimates depending on whether there was a conflict or not, whereas intuitive responders did not. This seems inconsistent with the idea that intuitive responders are just as likely as deliberative responders to detect a conflict between reason and intuition.

Conclusion

When there is a conflict between reason and intuition and the deliberative solution is better than the intuitive one, intuitive responders are unskilled and unaware (they do not know and they do not know that they do not know), whereas deliberative responders are skilled and aware (they know, they know that they know, and they know that others might not know). We believe that this metacognitive advantage is not restricted to reasoning and JDM tasks that are cleverly designed to tease apart intuitive and deliberative thought processes. It should manifest itself in any domain where people have learned, through training or experience, to think or proceed in a different, better way but still remain aware of the old and sometimes less effective way. The awareness that there are alternative solutions and that others might know only the most basic or intuitive of them puts them in a better vantage point both to judge their performance and to estimate their standing in comparison to others.

References

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science, 6,* 3–5. doi:10.1177/1745691610393980

Burson, K. A., Larrick, R. P., & Klayman, J. (2006). Skilled or unskilled, but still unaware of it: How perceptions of difficulty drive miscalibration in relative comparisons. *Journal of Personality and Social Psychology, 90,* 60–77. doi:10.1037/0022-3514.90.1.60

Carter, T. J., & Dunning, D. (2008). Faulty self-assessment: Why evaluating one's own competence is an intrinsically difficult task. *Social and Personality Psychology Compass, 2,* 346–360. doi:10.1111/j.1751-9004.2007.00031.x

Chaiken, S., & Trope, Y. (1999). *Dual-process theories in social psychology*. New York, NY: Guilford Press.

Chambers, J. R., & Windschitl, P. D. (2004). Biases in social comparative judgments: The role of nonmotivational factors in above-average and comparative-optimism effects. *Psychological Bulletin, 130,* 813–838. doi:10.1037/0033-2909.130.5.813

De Neys, W. (2006). Automatic-heuristic and executive-analytic processing during reasoning: Chronometric and dual-task considerations. *Quarterly Journal of Experimental Psychology, 59,* 1070–1100. doi:10.1080/02724980543000123

De Neys, W. (2012). Bias and conflict: A case for logical intuitions. *Perspectives on Psychological Science, 7,* 28–38. doi:10.1177/1745691611429354

De Neys, W., & Franssens, S. (2009). Belief inhibition during thinking: Not always winning but at least taking part. *Cognition, 113,* 45–61. doi:10.1016/j.cognition.2009.07.009

Dijksterhuis, A., Bos, M. W., Nordgren, L. F., & van Baaren, R. B. (2006, February 17). On making the right choice: The deliberation-without-attention effect. *Science, 311,* 1005–1007. doi:10.1126/science.1121629

Dunning, D. (2011). The Dunning–Kruger effect: On being ignorant of one's own ignorance. In J. Olson & M. P. Zanna (Eds.), *Advances in experimental social psychology* (Vol. 44, pp. 247–296). New York, NY: Elsevier.

Dunning, D., Johnson, K., Ehrlinger, J., & Kruger, J. (2003). Why people fail to recognize their own competence. *Current Directions in Psychological Science, 12,* 83–87. doi:10.1111/1467-8721.01235

Ehrlinger, J., Johnson, K., Banner, M., Dunning, D., & Kruger, J. (2008). Why the unskilled are unaware: Further explorations of (absent) self-insight among the incompetent. *Organizational Behavior and Human Decision Processes, 105,* 98–121. doi:10.1016/j.obhdp.2007.05.002

Epstein, S. (1994). Integration of the cognitive and psychodynamic unconscious. *American Psychologist, 49,* 709–724. doi:10.1037/0003-066X.49.8.709

Evans, J. S. B. T. (2006). The heuristic-analytic theory of reasoning: Extension and evaluation. *Psychonomic Bulletin & Review, 13,* 378–395. doi:10.3758/BF03193858

Evans, J. S. B. T. (2007). On the resolution of conflict in dual-process theories of reasoning. *Thinking & Reasoning, 13,* 321–339. doi:10.1080/13546780601008825

Evans, J. S. B. T., & Curtis-Holmes, J. (2005). Rapid responding increases belief bias: Evidence for the dual-process theory of reasoning. *Thinking & Reasoning, 11,* 382–389. doi:10.1080/13546780542000005

Ferreira, M. B., Garcia-Marques, L., Sherman, S. J., & Sherman, J. W. (2006). Automatic and controlled components of judgment and decision making. *Journal of Personality and Social Psychology, 91,* 797–813. doi:10.1037/0022-3514.91.5.797

Finucane, M. L., Alhakami, A., Slovic, P., & Johnson, S. M. (2000). The affect heuristic in judgments of risks and benefits. *Journal of Behavioral Decision Making, 13,* 1–17. doi:10.1002/(SICI)1099-0771(200001/03)13:1<1::AID-BDM333>3.0.CO;2-S

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives, 19,* 25–42. doi:10.1257/089533005775196732

Goodman, J. K., Cryder, C. E., & Cheema, A. (in press). Data collection in a flat world: The strengths and weaknesses of Mechanical Turk samples. *Journal of Behavioral Decision Making*.

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

Kahneman, D., & Frederick, S. (2002). Representativeness revisited: Attribute substitution in intuitive judgement. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 49–81). Cambridge, England: Cambridge University Press. doi:10.1017/CBO9780511808098.004

Klar, Y., & Giladi, E. E. (1999). Are most people happier than their peers, or are they just happy? *Personality and Social Psychology Bulletin, 25,* 586–595. doi:10.1177/0146167299025005004

Klayman, J., Soll, J. B., Gonzalez-Vallejo, C., & Barlas, S. (1999). Overconfidence: It depends on how, what, and whom you ask. *Organizational Behavior and Human Decision Processes, 79,* 216–247. doi:10.1006/obhd.1999.2847

Krueger, J., & Mueller, R. A. (2002). Unskilled, unaware, or both? The contribution of social-perceptual skills and statistical regression to self-enhancement biases. *Journal of Personality and Social Psychology, 82,* 180–188. doi:10.1037/0022-3514.82.2.180

Kruger, J. (1999). Lake Wobegon be gone! The "below average effect" and the egocentric nature of comparative ability judgments. *Journal of Personality and Social Psychology, 77,* 221–232. doi:10.1037/0022-3514.77.2.221

Kruger, J., & Burrus, J. (2004). Egocentrism and focalism in unrealistic optimism (and pessimism). *Journal of Experimental Social Psychology, 40,* 332–340. doi:10.1016/j.jesp.2003.06.002

Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology, 77,* 1121–1134. doi:10.1037/0022-3514.77.6.1121

Kruger, J., & Dunning, D. (2002). Unskilled and unaware—But why? A reply to Krueger and Mueller. *Journal of Personality and Social Psychology, 82,* 189–192. doi:10.1037/0022-3514.82.2.189

Larrick, R. P., Burson, K. A., & Soll, J. B. (2007). Social comparison and confidence: When thinking you're better than average predicts overconfidence (and when it does not). *Organizational Behavior and Human Decision Processes, 102,* 76–94. doi:10.1016/j.obhdp.2006.10.002

Moore, D. A., & Healy, P. J. (2008). The trouble with overconfidence. *Psychological Review, 115,* 502–517. doi:10.1037/0033-295X.115.2.502

Morris, S. B., & DeShon, R. P. (2002). Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. *Psychological Methods, 7,* 105–125. doi:10.1037/1082-989X.7.1.105

Nickerson, R. S., Baddeley, A., & Freeman, B. (1987). Are people's estimates of what other people know influenced by what they themselves know? *Acta Psychologica, 64,* 245–259. doi:10.1016/0001-6918(87)90010-2

Prowse Turner, J. A., & Thompson, V. A. (2009). The role of training, alternative models, and logical necessity in determining confidence in syllogistic reasoning. *Thinking & Reasoning, 15,* 69–100. doi:10.1080/13546780802619248

Roberts, M. J., & Newton, E. J. (2001). Inspection times, the change task, and the rapid-response selection task. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 54*(A), 1031–1048.

Ross, L., Greene, D., & House, P. (1977). The false consensus phenomenon: An attributional bias in self-perception and social perception processes. *Journal of Experimental Social Psychology, 13,* 279–301. doi:10.1016/0022-1031(77)90049-X

Sanford, A. J. S., Sanford, A. J., Molle, J., & Emmott, C. (2006). Shallow processing and attention capture in written and spoken discourse. *Discourse Processes, 42,* 109–130. doi:10.1207/s15326950dp4202_2

Schroyens, W., Schaeken, W., & Handley, S. J. (2003). In search of counterexamples: Deductive rationality in human reasoning. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 56*(A), 1129–1145.

Sherman, S. J., Presson, C. C., & Chassin, L. (1984). Mechanisms underlying the false consensus effect: The special role of threats to the self. *Personality and Social Psychology Bulletin, 10,* 127–138. doi:10.1177/0146167284101015

Shynkaruk, J. M., & Thompson, V. A. (2006). Confidence and accuracy in deductive reasoning. *Memory & Cognition, 34,* 619–632. doi:10.3758/BF03193584

Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin, 119,* 3–22. doi:10.1037/0033-2909.119.1.3

Slovic, P., & Tversky, A. (1974). Who accepts Savage's axiom? *Behavioral Science, 19,* 368–373. doi:10.1002/bs.3830190603

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

Stanovich, K. E., & West, R. F. (2000). Individual differences in reasoning: Implications for the rationality debate. *Behavioral and Brain Sciences, 23,* 645–665. doi:10.1017/S0140525X00003435

Thompson, V. A. (2009). Dual process theories: A metacognitive perspective. In J. S. B. T. Evans & K. Frankish (Eds.), *In two minds: Dual processes and beyond* (pp. 171–195). New York, NY: Oxford University Press. doi:10.1093/acprof:oso/9780199230167.003.0008

Thompson, V. A., Prowse Turner, J. A., & Pennycook, G. (2011). Intuition, reason, and metacognition. *Cognitive Psychology, 63,* 107–140. doi:10.1016/j.cogpsych.2011.06.001

Tsujii, T., & Watanabe, S. (2010). Neural correlates of belief-bias reasoning under time-pressure: A near-infrared spectroscopy study. *NeuroImage, 50,* 1320–1326. doi:10.1016/j.neuroimage.2010.01.026

Tversky, A., & Kahneman, D. (1974, September 27). Judgment under uncertainty: Heuristics and biases. *Science, 185,* 1124–1131. doi:10.1126/science.185.4157.1124

Windschitl, P. D., Kruger, J., & Simms, E. (2003). The influence of egocentrism and focalism on people's optimism in competitions: When what affects us equally affects me more. *Journal of Personality and Social Psychology, 85,* 389–408. doi:10.1037/0022-3514.85.3.389

Appendix A

Problems Used in Studies 1 and 2

1. A bat and a ball together cost 110 cents. The bat costs 100 cents more than the ball. How much does the ball cost?
2. If it takes 5 machines 5 minutes to make 5 widgets, how long would it take 100 machines to make 100 widgets?
3. 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?

Appendix B

Problems Used in Study 3

In a big research project a number of studies were carried out in which short personality descriptions of the participants were made. In every study there were participants from two population groups (e.g., carpenters and policemen). In each study one participant was drawn at random from the sample. You'll get to see the personality description of this randomly chosen participant. You'll also get information about the composition of the population groups tested in the study in question. You'll be asked to indicate to which population group the participant most likely belongs.

Conflict Problems

1. In a study 100 people were tested. Among the participants there were 5 people who drive a used Nissan and 95 people who drive a BMW. José is a randomly chosen participant of the study. José is 38 years old. He works in a steel plant. He lives in a small apartment in the outskirts of Barreiro*. His wife has left him.

What is most likely?
José drives a BMW.
José drives a used Nissan.
*Note: Barreiro is a town associated with the lower class.

2. In a study 100 people were tested. Among the participants there were 5 sixteen-year-olds and 95 forty-year-olds. Joana is a randomly chosen participant of the study. Joana likes to listen to techno and electro music. She often wears tight sweaters and jeans. She loves to dance and has a small nose piercing.

What is most likely?
Joana is 16 years old.
Joana is 40 years old.

(*Appendices continue*)

3. In a study 100 people were tested. Among the participants there were 95 Swedes and 5 Italians. Mario is a randomly chosen participant of the study. Mario is 25 years old. He is a charming young man and is a real womanizer. His favorite dish is the spaghetti his mother makes.
What is most likely?
Mario is a Swede.
Mario is an Italian.

4. In a study 100 people were tested. Among the participants there were 95 Muslims and 5 Buddhists. Sara is a randomly chosen participant of the study. Sara is 19 years old. She likes to philosophize and she hates materialism. She wears secondhand clothes and would love to go to India one day.
What is most likely?
Sara is a Buddhist.
Sara is a Muslim.

No-Conflict Problems

1. In a study 100 people were tested. Among the participants there were 95 people who like to watch independent French movies and 5 people who like to watch Hollywood action movies. Mariana is a randomly chosen participant of the study. Mariana is 35 years old. She writes reviews for a cultural magazine. Her husband is a university professor. She loves painting and photography.
What is most likely?
Mariana likes to watch independent French movies.
Mariana likes to watch Hollywood action movies.

2. In a study 100 people were tested. Among the participants there were 95 people who live in the country and 5 people who live in the city. Débora is a randomly chosen participant of the study. Débora is 22 years old. She rides a horse. After school she takes care of the animals at home. In the weekends she rises early and visits her grandparents.
What is most likely?
Débora lives in the country.
Débora lives in the city.

3. In a study 100 people were tested. Among the participants there were 5 people who vote for a left-wing ecologist party and 95 people who vote for a right-wing nationalist party. Lúcia is a randomly chosen participant of the study. Lúcia is 67 years old. She believes that traditional values are important and lives in an area where there's a lot of crime.
What is most likely?
Lúcia votes for a left-wing ecologist party.
Lúcia votes for a right-wing nationalist party.

4. In a study 100 people were tested. Among the participants there were 5 women and 95 men. Zé is a randomly chosen participant of the study. Zé is 32 years old and is a self-confident and competitive person. Zé's goal is to build a career. Zé does a lot of sport and is well-muscled.
What is most likely?
Zé is a woman.
Zé is a man.

Appendix C

Problems Used in Study 4

1. All flowers need water
Roses need water
Therefore, roses are flowers

2. All things with an engine need oil
Cars need oil
Therefore, cars have engines

3. All mammals can walk
Whales are mammals
Therefore, whales can walk

4. All vehicles have wheels
A boat is a vehicle
Therefore, a boat has wheels

(*Appendices continue*)

Appendix D

Solutions Shown in Study 6A

1. Bat-and-ball problem

Intuitive solution: *The ball costs 10 euros. 110 euros minus 100 euros equals 10 euros.*

Deliberative solution: *The ball costs 5 euros. The bat costs 100 euros more than the ball, that is, 105 euros. Therefore, the bat and the ball together cost 110 euros.*

2. Machines-and-widgets problem

Intuitive solution: *100 minutes. If 5 machines take 5 minutes to make 5 widgets, then 100 machines take 100 minutes to make 100 widgets.*

Deliberative solution: *5 minutes. If you increase the number of machines, you automatically increase the number of widgets that they will make without increasing the time that is necessary to make them. The time it takes to make the machines is always the same. Therefore, if 5 machines take 5 minutes to make 5 widgets, 100 machines take 5 minutes to make 100 widgets, just as 1 machine takes 5 minutes to make 1 widget.*

3. Lake-and-lily pads problem

Intuitive solution: *24 days. Half of 48 is 24.*

Deliberative solution: *47 days. If every day the patch doubles in size, then on the 48th day the patch of lily pads is twice the size that is was on the 47th day. Therefore, on Day 47 the patch is half the size that it will be on Day 48. Thus, if it takes 48 days for the patch of lily pads to cover the entire lake, then it takes 47 days for the patch to cover half the lake.*

Appendix E

Solutions Shown in Study 6B

1. Bat-and-ball problem (adapted): A TV and a DVD together cost 110 dollars. The TV costs 100 dollars more than the DVD. How much does the DVD cost?

Intuitive solution: *One possible solution is 10 dollars. The TV and the DVD together cost 110 dollars. 110 dollars minus 100 dollars is 10 dollars. Therefore, the DVD costs 10 dollars.*

Deliberative solution: *One possible solution is 5 dollars. The DVD costs 5. The TV costs 100 more than the DVD: 105. Therefore, the TV and the DVD together cost 110 dollars.*

2. Machines-and-widgets problem (adapted): *If it takes 5 hens 5 days to lay 5 eggs, how long would it take 100 hens to lay 100 eggs?*

Intuitive solution: *One possible solution is 100 days. If 5 hens take 5 days to lay 5 eggs, then 100 hens take 100 days to lay 100 eggs, just as 20 hens would take 20 days to lay 20 eggs.*

Deliberative solution: *One possible solution is 5 days. If you increase the number of hens, you automatically increase the number of eggs that they will lay. The time it takes for them to lay the eggs is always the same.*

3. Lake-and-lily pads problem (adapted): *A computer virus is spreading through the system of a computer. Every minute, the number of infected files doubles. If it takes 48 minutes for the virus to infect all of the system, how long would it take for the virus to infect half of the system?*

Intuitive solution: *One possible solution is 24 minutes. If it takes 48 minutes for the virus to infect all of the system, then it takes half that time for it to infect half of the system.*

Deliberative solution: *One possible solution is 47 minutes. If the number of infected files doubles every minute, then the number of infected files on Minute 47 is half the number of infected files on Minute 48.*

Received February 6, 2012
Revision received May 30, 2013
Accepted May 30, 2013 ■