COGNITIVE SCIENCE
A Multidisciplinary Journal

Cognitive Science 40 (2016) 2122–2136
Copyright © 2015 Cognitive Science Society, Inc. All rights reserved.
ISSN: 0364-0213 print / 1551-6709 online
DOI: 10.1111/cogs.12303

The Selective Laziness of Reasoning

Emmanuel Trouche,[a] Petter Johansson,[b,c] Lars Hall,[b] Hugo Mercier[d]

[a]*CNRS, Laboratory for Language, Brain and Cognition*
[b]*Cognitive Science, Lund University*
[c]*Swedish Collegium for Advanced Study, Uppsala University*
[d]*Center for Cognitive Sciences, University of Neuchâtel*

Received 7 October 2014; received in revised form 14 July 2015; accepted 16 July 2015

Abstract

Reasoning research suggests that people use more stringent criteria when they evaluate others' arguments than when they produce arguments themselves. To demonstrate this "selective laziness," we used a choice blindness manipulation. In two experiments, participants had to produce a series of arguments in response to reasoning problems, and they were then asked to evaluate other people's arguments about the same problems. Unknown to the participants, in one of the trials, they were presented with their own argument as if it was someone else's. Among those participants who accepted the manipulation and thus thought they were evaluating someone else's argument, more than half (56% and 58%) rejected the arguments that were in fact their own. Moreover, participants were more likely to reject their own arguments for invalid than for valid answers. This demonstrates that people are more critical of other people's arguments than of their own, without being overly critical: They are better able to tell valid from invalid arguments when the arguments are someone else's rather than their own.

Keywords: Reasoning; Argumentation; Choice blindness; Belief bias

1. Introduction

The way people produce arguments is doubly problematic. First, they mostly find arguments for their own side. Second, these arguments tend to be relatively weak. The first trait of argument production—the confirmation bias or myside bias—has been the topic of much attention (see, e.g., Nickerson, 1998). The later has been comparatively neglected, but is well supported by the existing evidence. When asked to justify their points of view, many participants can only generate arguments that make "superficial

Correspondence should be sent to Hugo Mercier, Centre de Sciences Cognitives, Université de Neuchâtel, Espace Louis-Agassiz 1, Neuchâtel 2000, Switzerland. E-mail: hugo.mercier@unine.ch

sense" (Perkins, 1985, p. 568), and they fail to offer genuine evidence (Kuhn, 1991). Similar results have been observed in social psychology (Nisbett & Ross, 1980) and in the study of formal reasoning (Evans, 2002). When people face simple problems ranging from the Wason selection task (Wason, 1966) to the Cognitive Reflection Test (Frederick, 2005), they typically start with a wrong intuition, which the subsequent reasoning fails to correct in most cases. This happens not only because people mostly look for arguments supporting their intuition (see Ball, Lucas, Miles, & Gale, 2003), but also because they are satisfied with the arguments they find—arguments that must be flawed given that they support a logically or mathematically invalid answer. Summarizing the perspective of dual process theories, Kahneman (2011) explains this poor performance of reasoning by the fact that "System 2 is sometimes busy, and often lazy" (p. 81): Reasoners do not make the effort that would be required to produce better arguments (see also, e.g., Evans, 2008).

This laziness, however, does not seem to apply to all arguments. When people evaluate other people's arguments—in particular, if they disagree with their conclusion—they appear to be more careful, and to mostly accept strong arguments. This result has been observed in research on persuasion and attitude change (for a review, see Petty & Wegener, 1998), and in Bayesian studies of argumentation (Hahn & Oaksford, 2007). Sound argument evaluation skills are also indicated by the fact that participants are convinced by arguments supporting the valid answer to reasoning problems such as those mentioned above (for the Wason selection task, see Moshman & Geil, 1998; for the CRT, see Trouche, Sander, & Mercier, 2014; and, more generally, Laughlin, 2011).

When it comes to evaluating others' arguments, the evaluation is most likely to be thorough when participants disagree with the argument's conclusion. When they agree with an argument's conclusion, not only are participants more likely to find the argument valid, but they also discriminate less between valid and invalid arguments, showing a relaxation of their evaluative criteria (Evans, Barston, & Pollard, 1983). Given that when participants produce arguments, they agree with the argument's conclusion, a more general way to frame the asymmetry between argument production and argument evaluation is as follows. When people agree with an argument's conclusion, they tend to evaluate it only superficially—this includes others' arguments whose conclusion one agrees with or arguments one produces. When people disagree with an argument's conclusion, they tend to evaluate it more thoroughly. Reasoning would thus only be *selectively* lazy.

The asymmetry that has the greatest ecological validity is that between the production of arguments and the evaluation of arguments whose conclusion one disagrees with—this is what happens in a standard exchange of arguments in which two or more people try to convince each other of their respective viewpoints. However, this asymmetry has only been indirectly demonstrated, from comparisons of disparate studies, and it is confounded by the fact that argument quality varies between different contexts and interlocutors. A convincing demonstration of this asymmetry would instead involve participants evaluating *their own arguments as if they were someone else's*. We would then expect that the participants would reject many of the arguments they deemed good enough to produce, if they thought the arguments came from someone else and they disagreed with their

conclusion. Moreover, they should be better at discriminating between their own good and bad arguments when they think they are someone else's and they disagree with their conclusion.

To test this prediction, we relied on the choice blindness paradigm, in which participants are led to believe that they have provided a given answer when in fact they answered something else. For example, in Hall, Johansson, and Strandberg (2012), the participants rated to what extent they agreed with moral issues, such as "If an action might harm the innocent, it is morally *reprehensible* to perform it." Using a sleight of hand, the participants' answers were at times reversed: If they had indicated that they agreed with the preceding statement, their answer now read that they agreed with an opposite statement (i.e. "... it is morally *permissible*..."). Participants were then asked to defend their positions, so that they would sometimes be asked to defend a moral position that was the opposite of their originally stated position. Not only did more than half of the participants often miss the switch, but they also gave coherent and detailed arguments supporting the opposite of their original opinion.

This general finding has been replicated in a number of different contexts and domains. Choice blindness has been demonstrated for attractiveness of faces (Johansson, Hall, Sikström, & Olsson, 2005; Johansson, Hall, Sikström, Tärning, & Lind, 2006; Johansson, Hall, Tärning, Sikström, & Chater, 2014), moral and political choices (Hall et al., 2012, 2013), and financial decision making (McLaughlin & Somerville, 2013). In addition, choice blindness has been demonstrated for taste and smell (Hall, Johansson, Tärning, Sikström, & Deutgen, 2010), for tactile stimuli (Steenfeldt-Kristensen & Thornton, 2013), and for auditory stimuli (Lind, Hall, Breidegard, Balkenius, & Johansson, 2014; Sauerland, Sagana, & Otgaar, 2013).

In the present case, we use a choice blindness manipulation in a reasoning task to make people believe that an answer and an argument they previously provided had been generated by another participant. The main prediction of the selective laziness account is that participants would reject many of the arguments they previously made, in particular bad arguments. By contrast, they should be more likely to accept their own good arguments.

2. Experiment 1

2.1. Method

2.1.1. Participants

We recruited 237 participants (100 females, M_{age} = 34.2, SD = 12.0) residing in the United States through the Amazon Mechanical Turk website. The total N was reached in two sessions: first an N of 160 and then an N of 77. We estimated the N for the first session based on the low detection rates (see below for an explanation of detection rates) obtained in previous choice blindness manipulations which would have allowed us to retain most of the participants for the comparison between manipulated and non-manipulated problems. However, the relatively higher rate of detection in the current experiment

allowed us to keep fewer participants than expected in the manipulated condition. The second session was conducted to approach the *N* initially aimed at for the manipulated condition. The experiments took about 10 min to complete, and we paid the participants standard rates for participation ($0.7).

2.1.2. Procedure and materials

The experiment consisted of two phases. In Phase 1, we presented the participants with five enthymematic syllogisms—syllogisms with an implicit premise—in succession (for an example syllogism see Fig. 1; all syllogisms can be found in the section "Materials" in the Supporting Information). For each syllogism, we asked the participants to choose which of five alternatives they thought was the valid answer and to explain why they gave their chosen answer (see Fig. 1).

At the start of Phase 2, which took place right after Phase 1, we told the participants that all five problems would be presented again, accompanied by the answer and explanation from another participant. The complete instruction read:

> We will now proceed to the second phase of the experiment. For each of the five problems, we will give you the answer given by another participant along with the explanation they gave. Each answer was provided by a different participant. You will be able to change your answer in light of this information if you wish.

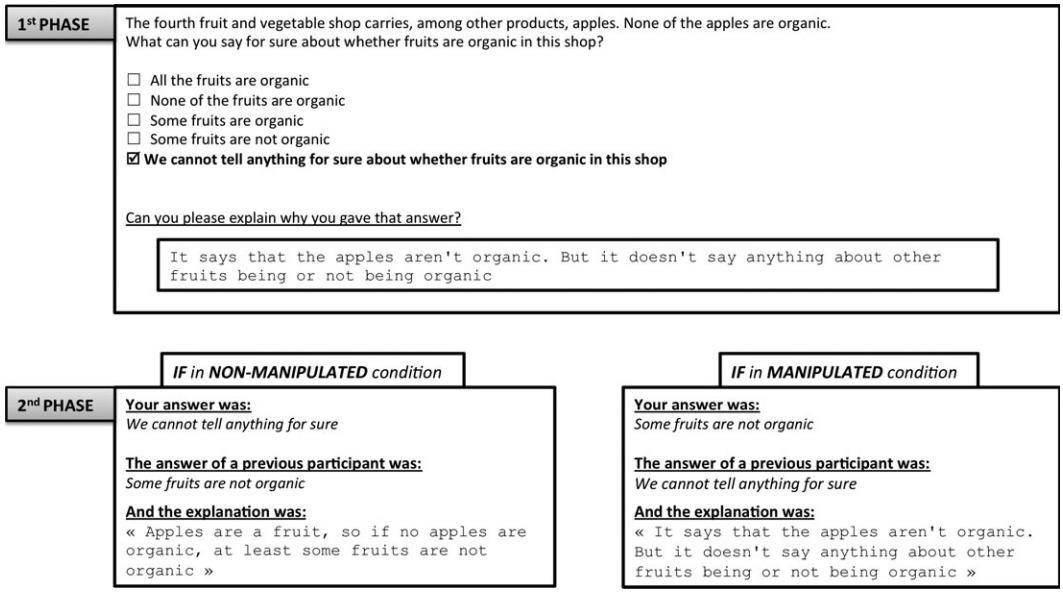

Fig. 1. Example of syllogism used in the experiment, shown both in the Manipulated and Non-Manipulated alternatives.

When the syllogisms were presented the second time, the participants were reminded of their own previous answer and provided with what was presented as someone else's answer and argument. They were told that they could change their answer in light of this information. The answer they had to evaluate was either the valid one (if the participants had previously given an invalid answer) or the most common invalid answer (if they had previously given the valid answer). The arguments presented for each answer were the same for all participants, constructed to be plausible explanations for making that particular (valid or invalid) choice (all the arguments used can be found in the Supporting Information).

However, for one of the syllogisms (the manipulated syllogism), instead of being truthfully reminded of their previous answer, participants were told that they had given an answer different from the one they had given: either the valid answer (if they had answered invalidly) or the most common invalid answer (if they had answered validly). Their own previous answer, and the argument that justified it, were presented as if they were those given by another participant. The external features of the presentation were strictly identical to those of the other four syllogisms (see Fig. 1 for an example of both conditions).

There were two different conditions, in which a different syllogism was manipulated. For half of the participants ($N = 119$), the EA3 syllogism was manipulated, and for the other half ($N = 118$) the EA4 was manipulated (see Supporting Information for the list of all syllogisms). This was to ensure that the outcome was not dependent on the properties of a particular syllogism. The Manipulated trials appeared on the fourth trial for all participants.

At the end of the experiment, we asked the participants a series of questions to determine if they had detected the manipulation of their answer. The first question simply asked what they thought of the experiment, the second if they thought anything was strange with the experiment, and in the third they were then presented with an actual description of the manipulation and asked if they had noticed this:

> During the second phase, in one or more problem your previous answers had been changed to another answer, and the "someone else's answer and argument" were actually the ones you had previously given. Did you notice this?

If they answered yes, we asked them how many times they had noticed this, followed by a question asking them to specify which of the five problems they thought had been altered. Finally, we asked the participants about their gender, age, and level of education.

2.1.3. Results

2.1.3.1. Detection rates: When the manipulation procedure was described in the debriefing, 53% of the participants indicated that they had noticed that their answers had been altered. Of these 53%, only 32% were able to retrospectively identify which of the syllogisms had actually been manipulated. This suggests that the 53% measure is conservative,

in that it includes participants who might not in fact have detected the manipulation. By contrast, it seems unlikely that the participants who said they had not detected the manipulation in fact had done so.

Using the conservative self-reported 53% as a criterion of detection (Detectors), we can ensure that the participants classified as Non-Detectors were in fact unaware of the manipulation, and thus treated their own argument as if it was someone else's (see Hall et al., 2012, 2013; Johansson et al., 2005 for further discussions of participant awareness and detection criteria in choice blindness experiments). The two target syllogisms did not differ in detection rates (EA3: 54%; EA4: 53%; Fisher's exact, $p = .90$, $OR = 0.95$), and they are therefore combined in the remainder of the analysis.

The participants who had given the right answer during the first phase were more likely to detect the manipulation compared to those that had given the wrong answer ($M = 64\%$ and 41%; Fisher's exact, $p < .001$, $OR = 2.5$). In Phase 2, three participants made a decision that was difficult to categorize: They neither stuck with the answer attributed to them nor accepted the argument, choosing instead a different invalid answer. Because their behavior was not easily interpretable in terms of sticking to one's answer versus accepting the arguments, which is the most relevant analysis for Phase 2, they were removed altogether from the analyses (including this statistic).

2.1.3.2. Phase 1 results: The mean score over five problems for all participants, counting 1 for a valid answer and 0 for an invalid answer, was 2.9 ($SD = 1.3$). Success rates for the syllogisms that were manipulated in Phase 2 were 60% for EA3 and 46% for EA4. It was, on average, 41% for Non-Detectors. For all the results that follow, we focus on the most relevant subset: the results from Non-Detectors on Manipulated problems. All the other results can be found in the Supporting Information.

2.1.3.3. Phase 2 results, rates of valid answers and comparison with Phase 1: Sixty-two percent of Non-Detectors gave a valid answer at Phase 2 on the Manipulated problem, a significant improvement over Phase 1 (Phase 1: 41% correct, exact McNemar's test, $\chi^2(1) = 7.93$, $p = .005$). The improvement in Phase 2 was also found for Non-Detectors as well as Detectors on the Non-Manipulated problems (see Supporting Information for details).

2.1.3.4. Phase 2 results, reaction to the arguments: Participants could react to the presentation of the arguments in three ways: they could keep the answer they had given in Phase 1 (or that had been attributed to them in the Manipulated trials), they could accept the answer supported by the argument (someone else's argument in the Non-Manipulated trials, their own argument in the Manipulated trials), or they could pick some other, new answer. As mentioned above, the three participants belonging to the last category have been removed from all analyses.

In Manipulated problems, Non-Detectors participants rejected what was in fact their own argument on 56% of the trials ([18 + 42]/[18 + 26 + 42 + 22] see Fig. 2). Participants who had given an invalid answer in Phase 1, and who were therefore presented

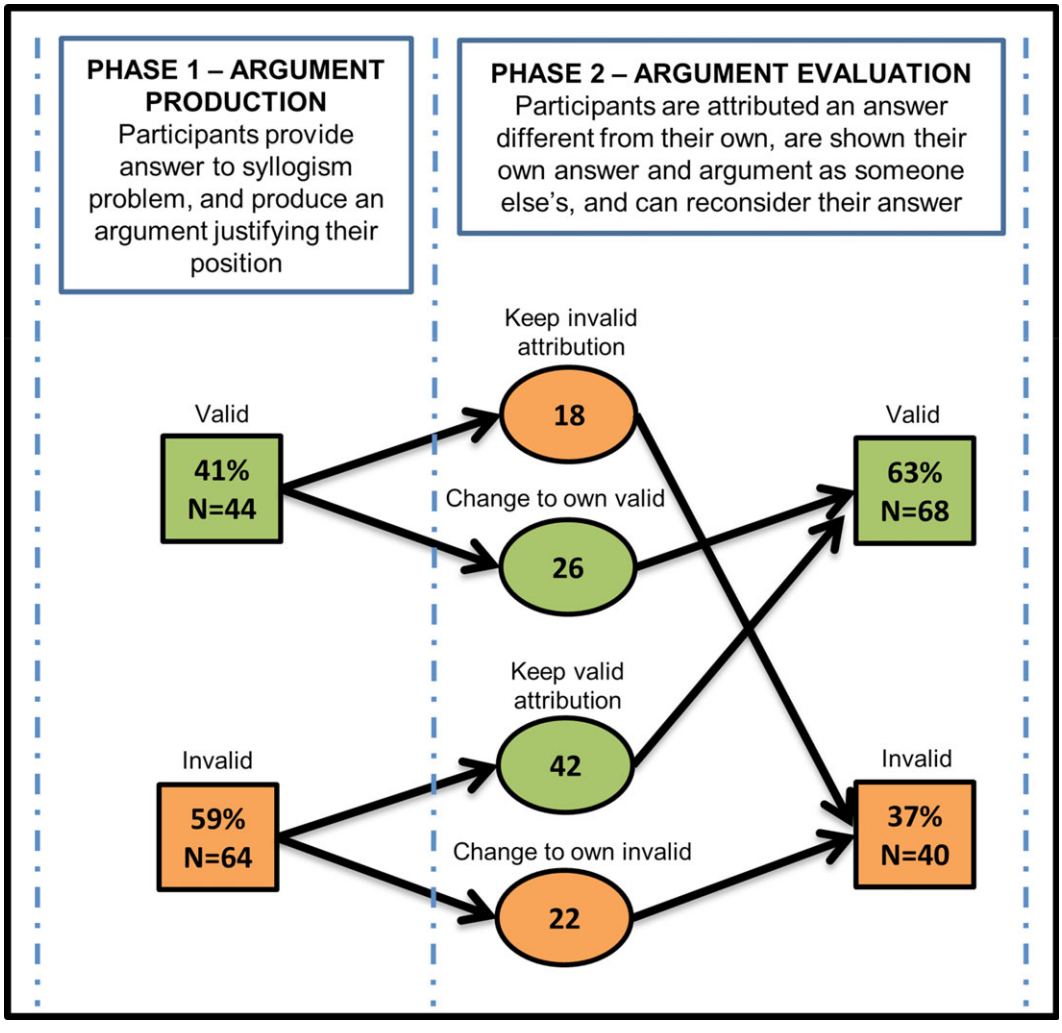

Fig. 2. Results for Non−Detectors on the Manipulated syllogism in Experiment 1. Boxes show percentages of valid and invalid answers at the end of each phase of the experiment. Arrows and ovals show the transition to valid and invalid answers as a consequence of the instructions in each phase. In Phase 1, participants provided the answer to a syllogism and provided an argument for their answer. In Phase 2, participants were attributed an answer different from their own and were confronted with their own Phase 1 answer and argument as if they were someone else's. They could then revise their attributed answer in light of this argument.

with their own argument for this invalid answer were more likely to reject the argument (42/[42 + 22] = 66%) than those who had given the valid answer in Phase 1, and who were therefore presented with their own argument for this valid answer (18/[18 + 26] = 41%) (Fisher's exact test, p = .02, OR = 2.7). For most of the Non-Manipulated syllogisms, both Non-Detectors and Detectors were also more likely to reject arguments for invalid than for valid answers (see Supporting Information).

2.1.4. Discussion

The goal of Experiment 1 was to compare the difference between how participants treat an argument when they produce it themselves and when they evaluate it as if it was someone else's. In Phase 1, participants were asked to solve reasoning problems and to provide arguments for their answer. If they evaluate their own arguments critically, they should realize that arguments for invalid answers are flawed and adopt valid answers. Thus, invalid answers reflect a poor evaluation of one's own arguments. In Phase 2, in which participants were asked to evaluate others' arguments, one problem was manipulated so that participants were in fact evaluating their own argument. Among the 47% who did not detect the manipulation, 56% rejected their own argument, choosing instead to stick to the answer that had been attributed to them. Moreover, these participants (Non-Detectors) were more likely to accept their own argument for the valid than for an invalid answer.

These results shows that people are more critical of their own arguments when they think they are someone else's, since they rejected over half of their own arguments when they thought that they were someone else's. It also shows that participants can discriminate strong from weak arguments when they think they are someone else's. However, a limitation of Experiment 1 is that it does not provide a good measure of how critical people are toward their own arguments when they produce them. Phase 1 performance is a mix of two factors. First, participants' initial intuitions (or initial models, see, e.g., Johnson-Laird & Bara, 1984), which might guide them toward the valid answer (see, e.g., Sperber, Cara, & Girotto, 1995). Second, participants' reasoning about this initial intuition. It is thus possible that reasoning played no positive role in Phase 1, even when participants provided the valid answer. Current "default-interventionist" models fit with this description of the participants' behavior in Phase 1, as they emphasize that reasoning often does not intervene to modify the intuitive, "default" answer (Evans, 2006; Kahneman, 2011; Stanovich, 2011).

To better examine the role of reasoning in the evaluation of one's own arguments, in Experiment 2 we introduced a Phase 0 in which participants had to provide a quick and intuitive answer to the same problems. In Phase 1, they were asked to justify this answer, and they could change their answer if they wished. This manipulation was similar to the "two responses" paradigm that has been previously used to study metacognitive monitoring (see, e.g., Thompson, Turner, & Pennycook, 2011; it should be noted that in the present experiment no metacognitive questions were asked).

Phase 2 was then identical to the Phase 2 of Experiment 1. This second experiment also addresses a potential concern with Experiment 1: The difference in performance between the production and evaluation of arguments might simply reflect the fact that people were confronted with the same problems for a second time when they were asked to evaluate arguments. As a result, they might simply have learned how to better solve this type of problems. In Experiment 2, the second presentation of the same problems occurs as people are asked to produce arguments, so it will be possible to tell if the mere repetition of the problems leads to improved performance.

Another possible concern is that the participants in the experiment might be afraid of not receiving their payment if they reported the manipulation. Experiment 2, tested for this possibility by introducing, for half the participants, a disclaimer prior to the debriefing reassuring the participants that they would get paid whatever they replied to the debriefing questions.

3. Experiment 2

3.1. Method

3.1.1. Participants

We recruited 174 participants (61 females, M_{age} = 35.5, SD = 12.0) residing in the United States through the Amazon Mechanical Turk website. The experiments took about 10 min to complete, and we paid the participants standard rates for participation ($0.7).

3.1.2. Procedure and materials

The experiment consisted of three phases. In Phase 0, the participants were presented with the five enthymematic syllogisms used in Experiment 1. For each syllogism, the participants were asked to choose which of five alternatives they thought was the valid answer. Participants were asked to provide a "fast, intuitive answer."

In Phase 1, participants were presented with the same problems, reminded of their initial answer, asked to provide an argument for this answer, and offered the possibility to give a new answer. No problem was manipulated in Phase 1.

Phase 2 was identical to the Phase 2 of Experiment 1, using the answers and arguments provided at Phase 1. Given that we had observed no interesting difference between the two syllogisms manipulated in Experiment 1, in this experiment we only manipulated one syllogism, an EA3.

The debriefing phase was identical to that of Experiment 1 with one exception: For half of the participants, it was preceded by a disclaimer reassuring them that they would get paid whatever they replied to the debriefing questions. Finally, we asked the participants about their gender, age, and level of education.

3.1.3. Results

Due to its design, Experiment 2 yielded a rich set of results, most of which do not speak to the point in hand. Here we focus on the results that pertain to the hypotheses laid out above. Other results can be found in the Supporting Information.

3.1.3.1. Detection rates: The debriefing manipulation (i.e., reassuring participants that they would get paid anyway) had no effect on detection rates (detection rate with special debriefing: 48%, normal debriefing: 44%, Fisher exact test p = .76, OR = 1.1); all results are thus presented together. When the manipulation procedure was described in the debriefing, 46% of the participants indicated that they had noticed that their answers had

been altered. Among the 46% of Detectors, only 40% were able to retrospectively identify which of the syllogisms had actually been manipulated. Using the inclusive self-reported 46% as a criterion of detection, we can ensure that the participants classified as Non-Detectors were in fact unaware of the manipulation, and thus treated their own argument as if it was someone else's. The participants who had given the right answer during Phase 1 were more likely to detect the manipulation compared to those that had given the wrong answer (M = 61% and 30%; Fisher's exact, $p < .001$, $OR = 3.6$). In Phase 2, eight participants made a decision that was difficult to categorize: they neither stuck with the answer attributed to them nor accepted the argument, choosing instead a different invalid answer. For the same reasons as in Experiment 1, they were removed altogether from the analyses (including this statistic).

3.1.3.2. Phase 0 results: The mean score over the five syllogisms for all participants, was 2.9 (SD = 1.19). It was 63% for the syllogism that was manipulated in Phase 2 (53% for Non-Detectors).

3.1.3.3. Phase 1 results: The mean score over the five syllogisms for all participants, counting 1 for a valid answer and 0 for an invalid answer, was 2.9 (SD = 1.30). It was 57% for the syllogism that was manipulated in Phase 2 (43% for Non-Detectors). The effects of participants' attempting to justify their intuitive answers can be broken down into the following categories. The participants who had initially provided a valid answer could either keep this valid answer, or change to adopt an invalid answer (see Fig. 3 for the results of Non-Detectors on the Manipulated problem). The participants who had initially provided an invalid answer could either keep this invalid answer, change to adopt the valid answer, or change to adopt another invalid answer. These results show that participants were more likely than not to keep their intuitive answer (23 changed and 141 did not, Binomial test, $p < .001$) and that they were not more likely to keep their initial answer when it was valid than when it was invalid (13% changed when valid, 17% changed when invalid, Fisher's exact, $p = .49$). The same pattern was observed among Non-Detectors for the Manipulated problem (note that this problem has not been manipulated yet) (19 changed and 67 did not, $p < .001$; 24% changed when valid, 20% changed when invalid, $p = .80$). For all the results that follow, we focus on the most relevant subset: the results from Non-Detectors on Manipulated problems. All the other results can be found in the Supporting Information.

3.1.3.4. Phase 2 results, rates of valid answers and comparison with Phase 1: Sixty-four percent of Non-Detectors gave a valid answer at Phase 2 on the Manipulated problem, a significant improvement over Phase 1 (Phase 1: 43% correct, exact McNemar's test, $\chi^2(1) = 5.78$, $p = .02$). The improvement in Phase 2 was also found for Non-Detectors as well as Detectors on the Non-Manipulated problems (see Supporting Information).

3.1.3.5. Phase 2 results, reaction to the arguments: Participants could react to the presentation of the arguments in three ways: they could keep the answer they had given (or

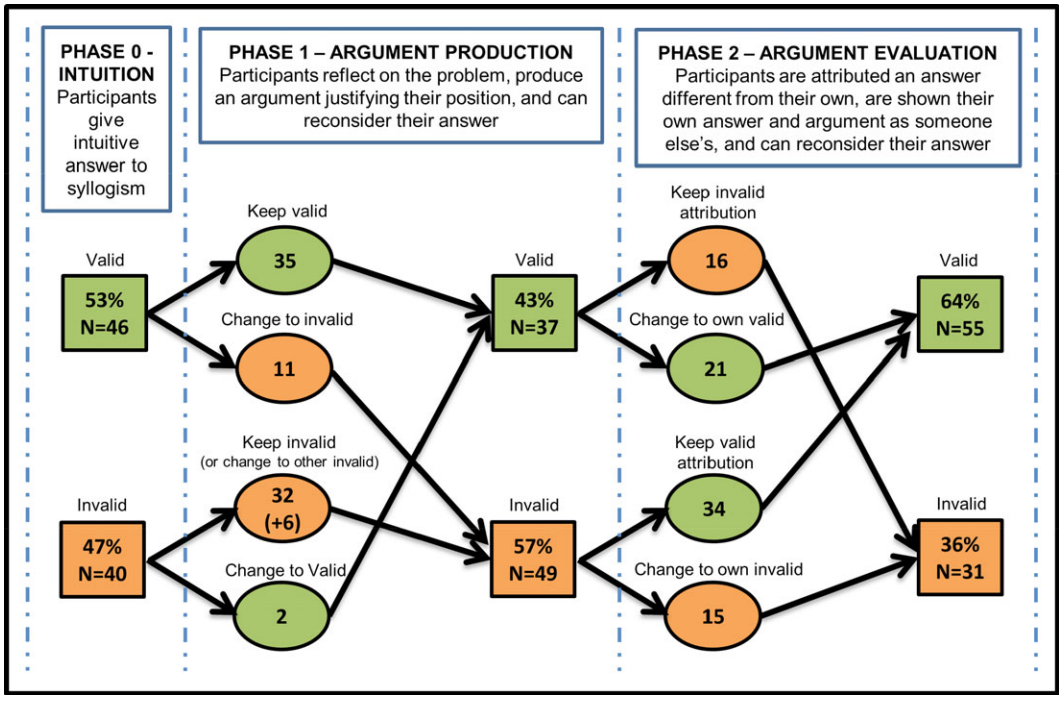

Fig. 3. Results for Non-Detectors on the Manipulated syllogism in Experiment 2. Boxes show percentages of valid and invalid answers at the end of each phase of the experiment. Arrows and ovals show the transition to valid and invalid answers as a consequence of the instructions in each phase. In Phase 0, participants provided an intuitive answer to a syllogism. In Phase 1, participants reflected on the same syllogism and provided an argument for their answer, which they could reconsider in the process. In Phase 2, still on the same syllogism, participants were attributed an answer different from their own, and were confronted with their own Phase 1 answer and argument as if they were someone else's. They could then revise their attributed answer in light of this argument.

that had been attributed to them in the Manipulated trials), they could accept the answer supported by the argument (someone else's in the Non-Manipulated trials, their own argument in the Manipulated trials), or they could pick some other, new answer (see Fig. 3). As mentioned above, the eight participants belonging to the last category have been removed from all analyses.

In Manipulated problems, Non-Detectors rejected what was in fact their own argument on 58% of the trials ([16 + 34]/[16 + 21 + 34 + 15]). Participants who had given an invalid answer in Phase 1, and who were therefore presented with their own argument for this wrong answer were more likely to reject the argument (34/[34 + 15] = 69%) than those who had given the valid answer in Phase 1, and who were therefore presented with their own argument for this valid answer (16/[16 + 21] = 43%) (Fisher's exact test, $p = .02$, $OR = 2.9$). We observed similar patterns for the Non-Manipulated syllogisms, both for Non-Detectors and Detectors (see Supporting Information).

3.1.4. Discussion

The goal of Experiment 2 was to further test the hypothesis tested in Experiment 1, namely, that participants are less critical of the same argument when they produce it than when they evaluate it as if it were someone else's. In particular, Experiment 2 aimed at a better understanding of the effects of argument production. Here, we focus on the Manipulated problem, for which we can compare the effects of argument production (in Phase 1), and the effects of the evaluation of the same argument when participants think it is someone else's (in Phase 2).

In Phase 0, participants provided an intuitive answer to the problem. In Phase 1 they were asked to give an argument to justify their answer, which they could then modify. If participants, during argument production, were critical of their own argument, in Phase 1 they would keep their intuitive answers given in Phase 0 when the answers are valid—which means that valid arguments can be found—and dismiss the intuitive answers when they are invalid—which means that no valid argument can be found.

The results from Phase 1 reveal that participants were not critical of their own arguments as they produced them. Not only did they only reject 5% of the invalid answers, but they did not reject valid answers at a different rate (13%). This result replicates previous results obtained with the "two responses" paradigm with similar problems (Thompson et al., 2011). Reasoning thus mostly provided post-hoc justifications for intuitive answers—a common outcome in this type of task (see, e.g., Evans & Wason, 1976)—and displayed no ability to discriminate between the participants' own valid and invalid answers.

By contrast, when participants thought the same arguments were someone else's, they prove more critical, rejecting 58% of the arguments. More important, reasoning also proved more discriminating, rejecting more arguments for invalid (69%) than for valid answers (43%).

4. Discussion

In two experiments, participants were asked to solve a series of simple reasoning problems, to produce arguments for their answers, and then to evaluate others' arguments about these answers. However, one of the problems was manipulated so that in fact participants were asked to evaluate their own argument as if it was someone else's. Across the two experiments, approximately half of the participants did not detect this manipulation. The participants who did not detect the manipulation thus evaluated an argument they had produced a few minutes before as if it was someone else's.

In the two experiments, participants proved critical of their own arguments when they thought that they were someone else's, rejecting more than half of the arguments. They also proved discriminating: They were more likely to reject their own arguments for invalid answers than their own arguments for valid answers.

Experiment 2 provides a contrast between the performance of reasoning when it evaluates others' arguments and when it produces arguments. In this experiment, participants were first asked to give intuitive answers to the problems, before being asked to produce arguments for these answers. The production of arguments had little effect on the answers, with the vast majority of participants keeping their intuitive answer. Moreover, participants were not more likely to change their invalid answers than their valid answers, so that reasoning did not exert any discrimination.

These experiments provide a very clear demonstration of the selective laziness of reasoning. When reasoning produces arguments, it mostly produces post-hoc justifications for intuitive answers, and it is not particularly critical of one's arguments for invalid answers. By contrast, when reasoning evaluates the very same arguments as if they were someone else's, it proves both critical and discriminating.

The present results are analogous to those observed in the belief bias literature (e.g., Evans et al., 1983). When participants evaluate an argument whose conclusion they agree with, they tend to be neither critical (they accept most arguments) nor discriminating (they are not much more likely to reject invalid than valid arguments). By contrast, when they evaluate argument whose conclusion they disagree with, they tend to be more critical (they reject more arguments) and more discriminating (they are much more likely to reject invalid than valid arguments). The similarity is easily explained by the fact that when reasoning produces arguments for one's position, it is automatically in a situation in which it agrees with the argument's conclusion.

Selective laziness can be interpreted in light of the argumentative theory of reasoning (Mercier & Sperber, 2011). This theory hypothesizes that reasoning is best employed in a dialogical context. In such contexts, opening a discussion with a relatively weak argument is often sensible: It saves the trouble of computing the best way to convince a specific audience, and if the argument proves unconvincing, its flaws can be addressed in the back and forth of argumentation. Indeed, the interlocutor typically provides counter-arguments that help the speaker refine her arguments in appropriate ways (for an extended argument, see Mercier, Bonnier, & Trouche, unpublished data). As a result, the laziness of argument production might not be a flaw but an adaptive feature of reasoning. By contrast, people should properly evaluate other people's arguments, so as not to accept misleading information—hence the selectivity of reasoning's laziness.

Acknowledgments

We thank the Swiss National Science Fund for its support through an Ambizione grant to H.M. P.J. thanks the Bank of Sweden Tercentenary Foundation and The Swedish Research Council (2014-1371). L.H. thanks the Bank of Sweden Tercentenary Foundation (P13-1059:1) and the Wallenberg Network Initiative (WNI). E.T. thanks the Direction Générale de l'Armement (DGA) for its support.

Data availability

All the data are available at this URL: https://sites.google.com/site/hugomercier/Data_ChoiceB_Reasoning.xls?attredirects=0

References

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

Evans, J. S. B. T. (2002). Logic and human reasoning: An assessment of the deduction paradigm. *Psychological Bulletin*, *128*(6), 978–996.

Evans, J. S. B. T. (2006). The heuristic-analytic theory of reasoning: Extension and evaluation. *Psychonomic Bulletin and Review*, *13*(3), 378–395.

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

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

Evans, J. S. B. T., & Wason, P. C. (1976). Rationalization in a reasoning task. *British Journal of Psychology*, *67*, 479–486.

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

Hahn, U., & Oaksford, M. (2007). The rationality of informal argumentation: A Bayesian approach to reasoning fallacies. *Psychological Review*, *114*(3), 704–732.

Hall, L., Johansson, P., & Strandberg, T. (2012). Lifting the veil of morality: Choice blindness and attitude reversals on a self-transforming survey. *PLoS ONE*, *7*(9), e45457.

Hall, L., Johansson, P., Tärning, B., Sikström, S., & Deutgen, T. (2010). Magic at the marketplace: Choice blindness for the taste of jam and the smell of tea. *Cognition*, *117*(1), 54–61.

Hall, L., Strandberg, T., Pärnamets, P., Lind, A., Tärning, B., & Johansson, P. (2013). How the polls can be both spot on and dead wrong: Using choice blindness to shift political attitudes and voter intentions. *PLoS ONE*, *8*(4), e60554.

Johansson, P., Hall, L., Sikström, S., & Olsson, A. (2005). Failure to detect mismatches between intention and outcome in a simple decision task. *Science*, *310*(5745), 116–119.

Johansson, P., Hall, L., Sikström, S., Tärning, B., & Lind, A. (2006). How something can be said about telling more than we can know: On choice blindness and introspection. *Consciousness and Cognition*, *15*(4), 673–692.

Johansson, P., Hall, L., Tärning, B., Sikström, S., & Chater, N. (2014). Choice blindness and preference change: You will like this paper better if you (believe you) chose to read it! *Journal of Behavioral Decision Making*, *27*(3), 281–289.

Johnson-Laird, P. N., & Bara, B. G. (1984). Syllogistic inference. *Cognition*, *16*(1), 1–61.

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

Kuhn, D. (1991). *The skills of arguments*. Cambridge, UK: Cambridge University Press.

Laughlin, P. R. (2011). *Group problem solving*. Princeton, NJ: Princeton University Press.

Lind, A., Hall, L., Breidegard, B., Balkenius, C., & Johansson, P. (2014). Speakers' acceptance of real-time speech exchange indicates that we use auditory feedback to specify the meaning of what we say. *Psychological Science*, *25*(6), 1198–1205.

McLaughlin, O., & Somerville, J. (2013). Choice blindness in financial decision making. *Judgment and Decision Making*, *8*(5), 561–572.

Mercier, H., & Sperber, D. (2011). Why do humans reason? Arguments for an argumentative theory. *Behavioral and Brain Sciences*, *34*(2), 57–74.

Moshman, D., & Geil, M. (1998). Collaborative reasoning: Evidence for collective rationality. *Thinking and Reasoning*, *4*(3), 231–248.

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomena in many guises. *Review of General Psychology*, *2*, 175–220.

Nisbett, R. E., & Ross, L. (1980). *Human inference: Strategies and shortcomings of social judgment*. Englewood Cliffs, NJ: Prentice-Hall.

Perkins, D. N. (1985). Postprimary education has little impact on informal reasoning. *Journal of Educational Psychology*, *77*, 562–571.

Petty, R. E., & Wegener, D. T. (1998). Attitude change: Multiple roles for persuasion variables. In D. Gilbert, S. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (pp. 323–390). Boston: McGraw-Hill.

Sauerland, M., Sagana, A., & Otgaar, H. (2013). Theoretical and legal issues related to choice blindness for voices. *Legal and Criminological Psychology*, *18*(2), 371–381.

Sperber, D., Cara, F., & Girotto, V. (1995). Relevance theory explains the selection task. *Cognition*, *57*, 31–95.

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

Steenfeldt-Kristensen, C., & Thornton, I. M. (2013). Haptic choice blindness. *I-Perception*, *4*(3), 207.

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

Trouche, E., Sander, E., & Mercier, H. (2014). Arguments, more than confidence, explain the good performance of reasoning groups. *Journal of Experimental Psychology: General*, *143*(5), 1958–1971.

Wason, P. C. (1966). Reasoning. In B. M. Foss (Ed.), *New horizons in psychology: I* (pp. 106–137). Harmandsworth, UK: Penguin.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Data S1. Materials and results from Experiments 1 and 2