Judgment and Decision Making, Vol. 17, No. 2, March 2022, pp. 331–361

Cognitive miserliness in argument literacy? Effects of intuitive and analytic thinking on recognizing fallacies

Annika M. Svedholm-Häkkinen* Mika Kiikeri†

Abstract

Fallacies are a particular type of informal argument that are psychologically compelling and often used for rhetorical purposes. Fallacies are unreasonable because the reasons they provide for their claims are irrelevant or insufficient. Ability to recognize the weakness of fallacies is part of what we call argument literacy and imporatant in rational thinking. Here we examine classic fallacies of types found in textbooks. In an experiment, participants evaluated the quality of fallacies and reasonable arguments. We instructed participants to think either intuitively, using their first impressions, or analytically, using rational deliberation. We analyzed responses, response times, and cursor trajectories (captured using mouse tracking). The results indicate that instructions to think analytically made people spend more time on the task but did not make them change their minds more often. When participants made errors, they were drawn towards the correct response, while responding correctly was more straightforward. The results are compatible with "smart intuition" accounts of dual-process theories of reasoning, rather than with corrective default-interventionist accounts. The findings are discussed in relation to whether theories developed to account for formal reasoning can help to explain the processing of everyday arguments.

Keywords: argumentation; informal reasoning; fallacies; mouse tracking; intuitive; analytic

*Tampere University, Tampere Institute for Advanced Study and Faculty of Education and Culture; University of Helsinki, Department of Psychology and Logopedics. Faculty of Education and Culture, P.O. Box 700, 33014 Tampere University, Finland. Email: annika.svedholm-hakkinen@tuni.fi. https://orcid.org/0000-0002-6462-6630.

†University of Helsinki, Department of Psychology and Logopedics. https://orcid.org/0000-0002-9325-053X.

Thanks to Riinu Elonheimo, Emma Kuokkanen, Solina Laevuo, Emma Lehtola, Merit Morikawa, Emma Oksanen, Saara Silakoski, Tanja Teräväinen, Maria Ukkonen, Suvi Öhman and Marjaana Lindeman, for creating stimuli and collecting data. Thanks to Eero Häkkinen for programming assistance, to Jon Baron for helpful comments on an earlier draft and for sharing R code, and to Tarja Peromaa for advice on SDT analyses.

1 Introduction

It can be argued that good argumentation and critical thinking are central aspects of our everyday reasoning, and crucial abilities for democracy and society. Yet, the ability to distinguish well-supported informal arguments from poor arguments, i.e., argument literacy, and the cognitive underpinnings of this ability, have not received recognition as central topics for mainstream cognitive psychology. Calls to bring the attention of psychologists to informal everyday argumentation have been made for years (e.g., Hornikx & Hahn, 2012; Kuhn, 1991; Shaw, 1996; Voss & Means, 1991; Voss et al., 1986), but so far, the research has tended to focus on precisely defined topics and particular types of arguments, rather than more broadly on argumentation as it appears in everyday discourse (Bonnefon, 2012). Thus, the cognitive processes that explain why people often trust poorly justified arguments are still largely unknown.

In everyday discourse, arguments can take many shapes, but any set of statements in which sets of reasons are given in support of a claim is an argument. Argument are strong when the reasons given to support a claim are relevant and sufficient. In weak arguments, the reasons given may be irrelevant, insufficient, or misleading.

The classic fallacies make up a particularly salient type of weak argument. These are arguments that turn out to be weak if inspected closely, but that are psychologically compelling on the surface. Thus, they are often used strategically and rhetorically. Textbooks on argumentation and critical thinking list many of these typical errors of argumentation that even educated and intelligent people typically make and fall prey to (Hamblin, 1970; Tindale, 2007). For example, the fallacy of a false dilemma is common in political debates, which present topics as if only two options were possible: immigrants are presented either as innocent victims or as opportunists taking advantage of their hosts. Another is the slippery slope fallacy, which argues against actions by claiming that they would lead to a chain of events culminating in terrible outcomes: "You should not let the authorities impose lockdowns even during a pandemic, because doing so will lead to fascism."

There is largely consensus on what is erroneous in each famous type of fallacy (e.g., Hamblin, 1970; Tindale, 2007), although debates continue (e.g., Fumagalli, 2020). More detailed analyses of particular types of fallacies have been presented at least in a Bayesian framework (Hahn, 2020; Hahn & Oaksford, 2007) and in the framework of the pragma-dialectical rules of fair discussions (van Eemeren et al., 2009). These discussions are, however, outside the scope of the present study, as our focus is on the cognitive underpinnings of the ability to distinguish the weakness of fallacies in general. Therefore, for the present purposes, we use classic textbook examples of fallacies, whose weakness we will for now take as given.

1.1 Studying informal arguments as a form of reasoning

Currently, there is little knowledge of the thinking processes that explain the appeal of fallacies. Previous psychological research on argumentation has examined, for example, how contextual factors affect people's ability to identify fallacies. This research shows that people are better at recognizing that fallacies are problematic when they are instructed to consider the opposite perspective or engaging in dialogue, and that familiarity with fallacies helps recognize them in real examples (Mercier, 2016; Neuman et al., 2006; Weinstock et al., 2004). People also show an awareness of context: they consider fallacies to be an acceptable tactic if one is quarreling, but not if the aim of a discussion is to convince the other part (Neuman et al., 2006).

Other research examined the processing steps involved in thinking about arguments. Philosophers argue that to evaluate the soundness of an argument, one should consider first, whether the reasons given are acceptable, true and relevant in themselves. Next, one should evaluate the relevance of the reasons for the claim (Angell, 1964). Existing research indicates that people spontaneously engage in the first step, in particular evaluating the believability of the premises (Thompson & Evans, 2012; Voss & Means, 1991). However, evaluating the relationship between claims and reasons seems to be more cognitively demanding and people often avoid doing it. For example, when asked to formulate counterarguments, many people elaborate on their own preferred explanations rather than providing evidence against the original argument (Kuhn & Modrek, 2017). In another study, even when explicitly instructed to evaluate the relationship between claims and reasons, people evaluated plausibility instead (Neuman et al., 2004). Similarly in a third study, the counterarguments that people produced tended to question the truth of the premises, rather than to question how well the premises support the claims (Shaw, 1996). Performance improved when participants were asked to rate the believability of the premises and the strength of the arguments separately, indicating that people can evaluate argument strength better than they do spontaneously, if they put their mind to it.

1.2 Argumentation and dual-process theories of reasoning

In the present study we are interested in relating argument evaluation to theories of how reasoning unfolds. According to popular dual-process theories of reasoning (DPTs), decisions arise from two main types of processing: fast and autonomous intuitive thinking, and slow and deliberate reasoning (Evans & Stanovich, 2013; Kahneman, 2011; Morewedge & Kahneman, 2010). In this framework, humans are depicted as cognitive misers, who use analytic processing sparingly, "operating most often under a least-effort principle, with intellectual values too weak to support the effort that thinking deeply requires" (Kuhn & Modrek, 2017, p. 97). Many findings in research on formal reasoning have been explained as instances of cognitive miserliness. For example, on base-rate tasks, people tend to respond in line with stereotypes rather than statistical frequency information (De Neys & Franssens,

2009; Tversky & Kahneman, 1974). Similarly, people treat ratios expressed using larger numbers, such as 18/100, as if they were higher than ratios expressed using small numbers, such as 2/10, because of the intuitive salience of the numerator (Bonner & Newell, 2010). On the popular Cognitive Reflection Test, which consists of items such as "A bat and a ball cost $1.10. The bat costs $1 more than the ball. How much does the ball cost?" the initial response for many people is 10 cents, rather than 5 cents, which requires some calculation (Frederick, 2005).

In the DPT framework, falling prey to fallacies could be explained as an instance of cognitive miserliness: it might be that fallacies appeal to intuitive thinking, in which people evaluate arguments quickly and without putting much effort into evaluation. Similar sentiments have been expressed by Schellens et al. (2017), whose discussion connects argument evaluation to Petty and Cacioppo's (1986) Elaboration Likelihood Model of persuasion, which is a form of DPT. These authors suggested that argument quality may be processed either superficially or in more detail, but they did not explicitly test predictions derived from the theory. Similarly, Thompson and Evans (2012) suggested that argument evaluation is influenced by intuition, because people's response choices in their study were different from their verbal explanations for these same choices. However, no studies to date have directly tested whether our tendency to fall for fallacies can be explained in terms of dual-process theories.

If fallacy acceptance can be explained as an instance of cognitive miserliness, the relationship might nevertheless not be straightforward. Dual-process theories of reasoning have largely been developed based on findings that concern people's understanding of formal logic and statistical reasoning (Kahneman, 2003; Stanovich, 2011). Meanwhile, evaluating informal arguments differs from formal reasoning in several ways. Everyday reasoning involves dealing with problems that are ill-structured and debatable, and that may have no definitive solutions (Galotti, 1989; Kuhn, 1991). By definition, informal arguments are not categorically valid or invalid, the way that formal arguments are. Rather, they may be relatively sound or unsound (Angell, 1964). Based on the ways in which everyday argumentation differs from formal reasoning, it is not clear whether the same cognitive processes that are at play in formal reasoning apply in the same ways to everyday argumentation.

An important topic of debate in DPT concerns whether correct responses require deliberative analytic processing or whether they can arise intuitively. According to long-standing default-interventionist, or corrective, accounts of DPT, initial notions that are based on heuristics, and which are therefore often wrong, are corrected by later analytic processing when necessary (Evans, 2008). That is, correct responses tend to require analytic intervention, such as using mental arithmetic on the CRT (Cognitive Reflection Test) tasks. As regards fallacies, for people to correctly recognize fallacies as being weak, their initial attraction towards these alluring arguments would typically have to be corrected through careful deliberation.

In contrast, more recent "logical intuition" models ("DPT 2.0"; De Neys, 2018) have been formulated to accommodate evidence showing that even normatively correct responses on formal reasoning tasks may in fact arise immediately and effortlessly (e.g., De Neys, 2012; Newman et al., 2017). These accounts suggest that decision-making may not always require deliberation, but may instead involve a choice between competing processes that all begin immediately. Some of these intuitions may be logical if the individual has internalized basic principles of logic to the point where these can be activated and applied effortlessly. For example, on base-rate tasks people may, through practice, have learned to immediately pay attention to the base-rates rather than to stereotypes (Bago & De Neys, 2017; De Neys & Pennycook, 2019; Newman et al., 2017).

To date, logical intuitions have been discussed only in the context of formal reasoning. For informal arguments, we can imagine that thinking about them could involve comparable well-justified intuitions. For fallacies, such "smart intuitions" would mean that people would have learned, to the point of automaticity, to recognize fallacies as being weak. For example, on arguments that appeal to authority, people would intuitively know that they should pay attention to whether the authority is relevant, because that is the relevant question to consider when evaluating an authority argument (van Eemeren et al., 2009). On arguments that appeal to consequences, people would intuitively pay attention to the plausibility of the proffered consequences, thereby intuitively recognizing most slippery slope arguments as being exaggerated, implausible and weak.

DPTs lead to several predictions regarding how people evaluate fallacies. First, a corrective account predicts that rapid intuitions first lead thinkers to accept fallacies, and slower deliberation is needed to reject fallacies. By this account, we could expect that if people are encouraged to rely on intuitive thinking, they would accept an increased number of fallacies, and to do it faster, as they would straightforwardly allow themselves to accept their first intuitions. Conversely, being encouraged to think more analytically or deliberately could be expected to have the reverse effects, as the decision would now involve an enhanced conflict between the first impression and careful, deliberative scrutiny of arguments that often results in the reasoner detecting that fallacies are weak. Thus, we could expect responses to become slower and to involve more decisional conflict and weighings between options.

If, on the other hand, as logical intuitions models would suggest, judgments about the strength of arguments are based on a competition between different types of processes that all begin immediately, we could expect that manipulations to think in either way should have less effect on how people respond, as their responses would to a larger extent have been conflicted from the outset.

To gain insight into the types of processing underlying responses, one method is the analysis of response times (RT). This approach assumes that longer latencies indicate the activation of more competing elements. Previously, Voss and colleagues have used RT in studies of argumentation, finding for example that evaluating arguments that one disagrees

with is slower than evaluating arguments that are in line with one's own opinions (e.g., Voss et al., 1993). Response times are nevertheless a "black-box" method, as they do not reveal how the decision-making process unfolds (Schneider et al., 2015). Moreover, response times may be affected by other factors besides decisional conflict, such as other sources of task difficulty. Finally, the reliability of RT may suffer as a result of speed-accuracy tradeoffs and confounding variables such as differences in how certain people want to be before they give their answers. Because RT does not tell us what it is that slows down responses, using RTs as the sole indicator of the processes involved in decision making is insufficient.

1.3 Mouse tracking and changes of mind

Another method that offers to shed insight on how decisions unfold over time is mouse tracking. Mouse tracking involves following the trajectory of the computer cursor as people make decisions. This method rests on the finding that cognitive processes such as decisional conflict and hesitation between response options continually influence motoric activity, which can be seen as deviations in the path on which people move the mouse (Freeman & Ambady, 2010; Stillman et al., 2018). Indices of decisional conflict derived from mouse tracking tend to be partly dissociable from RT, typically showing correlations around $r = .40$ (Stillman et al., 2018). Measures of mouse trajectories have previously been used to examine the competing influences on decisions on topics ranging from self-control and food preferences to racial attitudes (Gill & Ungson, 2018; Stillman & Ferguson, 2018).

Mouse tracking methodology has also been applied to study formal reasoning. A notable finding from these studies is that the types of results that inspired the logical intuition models have not been found in mouse tracking studies. Rather, the results from mouse tracking have conformed to corrective DPTs. In one study, Szaszi et al. (2018) studied the denominator neglect task. There were no indications that correct responses could arise intuitively. Instead, better reasoners exhibited trajectories that indicated changes of mind towards the correct response. However, the denominator neglect task is very fast-paced and may thus be a poor point of comparison for the present argument evaluation task.

For making predictions about how argument evaluation will unfold in a mouse tracking experiment, it may be more relevant to turn to findings obtained using more verbose tasks that involve more elaborate reasoning. One such study is by Travers et al. (2016), who used mouse tracking to operationalize decisional conflicts on the CRT. Even these results were most compatible with corrective DPTs, because corrective mouse movements were mostly found on trials ending in correct responses, while heuristic responses did not show attraction towards the correct option.

Another task type that also involves more verbose material and more elaborate reasoning are sacrificial moral dilemmas, which ask people whether it is acceptable to kill one to save the lives of many (utilitarian thinking) or not (deontological thinking). These tasks have been studied using mouse tracking by both Gürçay and Baron (2017) and Koop (2013).

These studies also looked for mind changes of the sort that corrective DPT would predict. The overall rate of switches from one side of the screen to the other was low (around 20 % of trials), and switches were not more common on trials ending in either type of response (deontological or utilitarian). Thus, in these types of tasks it seems that people usually made their choices early, with little interference on mouse trajectories from the competing response option. These findings illustrate that DPTs may not be directly applicable to reasoning on tasks outside the formal reasoning domain.

1.4 The present study

The present study tests the applicability of DPTs to everyday argumentation and asks whether humans are cognitive misers even when evaluating informal arguments. In line with this assumption and with DPTs, we predict that prompting participants to respond in line with their first impressions will make responses quicker, while prompting people to think analytically and carefully will make responses slower. In terms of performance and mouse trajectories, different DPTs lead to different predictions.

By the corrective account, correct responses typically require that analytic processing intervenes with the outputs of intuitive processes. Thus, by this account we should expect trials that end in correct responses (e.g., rating a fallacy as being weak) to exhibit attraction towards the incorrect response (e.g., rating a fallacy as being strong)[1], followed by corrective mouse movements. Incorrect responses, in turn, should be more straightforward as they would involve no analytic intervention. Encouraging people to think intuitively should decrease corrective movements and decrease performance, while encouraging analytic thinking should increase corrective movements and increase performance.

In contrast, by an account that postulates logical (or "smart" or "sound") intuitions, no corrective movements are necessary to respond correctly, as correct responses stem from correct intuitions. An integral part of the logical intuition account is the suggestion that people are intuitively sensitive to norms of thinking even if they end up giving the incorrect response. For example, skin conductance increases when people give incorrect responses to categorial syllogisms (De Neys et al., 2010). So far, evidence for this view has been found using multiple behavioral and physiological methods, but only on formal reasoning tasks (reviewed in De Neys, 2012). So far, no studies using mouse tracking have captured evidence for this phenomenon. If it did occur, we could expect it to mean that it is the trials that end in errors that would exhibit most attraction towards the opposite response option. Instructions to think in either way (intuitively or analytically) should have little effect on performance or on mouse trajectories.

[1]Note that we use the term "correct" for short to describe responses on which the participant rates a fallacy as being weak or a reasonable argument as being strong. Similarly, we use "incorrect" and "error" to refer to the reverse responses. These terms are used as convenient shorthands and should not be taken to imply that informal arguments are categorically valid or invalid in the same way as formal logic is.

To sum up, how people evaluate and react to fallacies and other informal everyday arguments has not previously been examined in the context of DPT or from the perspective of decisional conflict, leaving the cognitive processes involved largely unknown. The present study addresses the underlying processes by experimentally manipulating thinking to be either intuitive or analytic, all the while recording our participants' response times and mouse movements to reveal how their decisions about arguments' quality unfold in real time.

2 Method

2.1 Participants

Fifty-eight Finnish volunteers participated in the experiment (35 female, 21 male, 1 other, 1 NA). The mean age was 30.4 years, range 17–62, SD = 13.5 years, median = 24.0 years. Thirty-six were students, 21 working, and one's occupation was unknown. Fifteen had completed a tertiary degree and two a doctoral degree. All were native speakers of Finnish and had normal vision. The sample was a convenience sample recruited from the wider community. To control for the possible effects of familiarity with analyzing arguments, we asked whether the participants were familiar with the topic to the extent that they for example knew what is meant by "straw-man" and "slippery slope" arguments. The majority (65 %) reported no familiarity with these concepts, while 28 % reported having 'some' and 7 % reported 'much' knowledge about argumentation analysis.

2.2 Materials

2.2.1 Arguments

The stimuli originally consisted of 63 fallacious arguments and 57 reasonable arguments. The stimuli were formulated to sound like natural conversation in terms of both content and form. The fallacies were formulated with the help of textbooks and online resources (Downes, 1995; Sagan, 1997; Tenhunen, 1998) and included the following types: ad hominem, straw man, reference to irrelevant authorities, ad ignorantiam, ad baculum, ad consequentiam, false dilemma, circular argument, slippery slope, non sequitur, statistical fallacies, mixing correlation with causation, post hoc ergo propter hoc, and unfalsifiability. The reasonable arguments were formulated to be similar to the fallacies in terms of length and grammatical form. An example of a fallacy is "Petting cats relieves stress. Hence, cat owners are more easygoing than other people". An example of a reasonable argument is "You should not drive while intoxicated because alcohol lowers reaction speed, and driving requires a good ability to react quickly".

To make the comparison between fallacies and reasonable arguments as clear as possible, three scholars in argumentation analysis and philosophy screened the original materials to

identify ambiguous items. Independently of each other, the experts rated each argument on a five-point scale, indicating whether they found the argument to be clearly a fallacy (-2), a milder fallacy (-1), difficult to assess (0), a fairly reasonable argument (+1), or a strong argument (+2). Items that were deemed difficult to assess (0) by at least one expert, or on which the experts disagreed to the extent that one of them rated it on the opposite side of the scale midpoint from the others, were excluded from further analyses. The experts agreed that the 53 remaining fallacies were fallacies and that the 29 remaining reasonable arguments were reasonable. Appendix A presents all the arguments.

2.2.2 Manipulation check

Fourteen items from the Situation-Specific Thinking Style scale (SSTS; Novak & Hoffman, 2009), translated into Finnish, were used as a manipulation check. The participants were asked to indicate on a five-point scale (1 = *very poor description*, 5 = *very much on point*) to what extent the items described the manner in which the participant had responded after the experimental manipulation. Seven items described an intuitive style of thinking (e.g., "I trusted my hunches", Cronbach's α = .93) and seven items an analytic style (e.g., "I reasoned things out carefully", α = .86).

2.3 Procedure

Upon arrival at the laboratory, participants gave written informed consent. On the computer screen, they received instructions to evaluate arguments by whether the justifications given in them were strong or weak. They were given examples of strong and weak arguments and two practice trials. See Appendix B for the complete instructions.

On each trial, an argument was displayed in the middle bottom of the computer screen. In the top left and right corners of the screen were buttons with the text "Strong" and "Weak". The participants indicated their responses by moving the computer mouse to either button and clicking it. The argument was displayed until the participant responded. For half of the participants, Strong was on the left and Weak on the right, and for the other half, the order was the reverse. To display the next argument, participants had to click "Continue" in the bottom middle of the screen. The first 60 trials were presented in this baseline condition. Then, participants received one of two experimental manipulations. Thirty participants were instructed to think intuitively "based on their first impressions" on the rest of the trials, and 28 were instructed to think analytically "as reasonably and carefully as you can" on the rest of the trials. After these manipulations, the last 60 trials were presented as above, with an additional reminder to think intuitively or analytically after 20 trials. All instructions are found in Appendix B.

The order of presentation of the items was randomized across conditions individually for each participant. Thus, the number of fallacies and reasonable arguments presented to

individuals in the baseline and experimental conditions varied randomly. The number of fallacies in the baseline condition ranged from 27 to 37.

There were four combinations of manipulation and layout (intuitive with "strong" on the left; intuitive with "strong" on the right; analytic with "strong" on the left; analytic with "strong" on the right). Participants were assigned to one of these in rotating order as they arrived at the laboratory. The participants were not informed before the experiment that mouse trajectories would be recorded.

At the end of the experiment, the participants filled in a questionnaire containing the SSTS manipulation check and background questions. The participants were debriefed and they received a voucher worth 5 euros to compensate for lost time.

2.4 Mouse tracking measures

The experiment was implemented in the MouseTracker software which presents the stimuli and records the x and y coordinates of the mouse on the computer screen 60–75 times per second as participants are responding (Freeman & Ambady, 2010). To ease data processing, we used data that was time-normalized by MouseTracker to 101 steps using linear interpolation. To measure changes of mind after the participant has given consideration to both options, we calculated what proportion of each trajectory was spent on the side of the screen opposite of the final response. This was calculated as follows: X_ratio = X_opposite/(X_same + X_opposite), where X_opposite is the sum of the X-coordinates of all the time steps during which the cursor was on the opposite side of the screen from the final response, and X_same is the sum of all the time steps during which the cursor was on the same side as where the final response was made. R code for calculations is provided in the Supplementary material. The resulting measure (X ratio) is sensitive to how far to the opposite side the trajectory reaches as well as to the proportion of time spent on the opposite side.

In addition, to capture smaller deviations from a straight line, we analyzed the Area Under Curve (AUC) between the realized mouse trajectory and an idealized straight line from the start button to the response button, which was computed by MouseTracker and is a standard measure in mouse tracking research (Freeman & Ambady, 2010). Unlike the X_ratio, a high AUC does not necessarily indicate that the cursor was ever on the opposite side.

2.5 Data processing

Because the RT distribution was highly skewed to the right (skewness value = 3.47), we analyzed the natural logarithm of response times. After logarithm transformation, RT was approximately normally distributed. The data were analyzed using R (R core team, 2021), and the psych (Revelle, 2021) and psycho (Makowski, 2018) packages for R.

3 Results

3.1 Manipulation check

The results obtained using the SSTS scales indicate that the manipulations worked as intended. Participants who were encouraged to think intuitively reported higher scores on the SSTS Intuition scale (M = 3.78, SD = .69) than participants who were encouraged to think analytically (M = 2.15, SD = .75, $t(56)$ = 8.625, p < .001). Cohen's d was 2.26, denoting a large effect. Conversely, scores on the SSTS Analytic scale were notably higher among those receiving the analytic manipulation (M = 4.11, SD = .52) than those in the intuitive group (M = 3.12, SD = .68, $t(56)$ = 6.209, p <.001, Cohen's d = 1.63). As noted by a reviewer, these results may partly reflect demand effects, as the SSTS used many of the same words as the manipulations (e.g., analytic, logical, careful vs. intuition, first impression, instinct).

3.2 Effects of intuitive and analytic manipulations

Figure 1 presents means of the studied measures before and after the manipulations. Table 1 presents descriptive statistics of the change in each of these variables from baseline to after the manipulation. Table 1 also presents the results of t-tests on the size of the change depending on the type of manipulation received. All dependent variables were positively associated with each other (all p's < .001). As the X_ratio and AUC both describe the mouse trajectories, they overlapped substantially across trials (r = .48). Their associations with RT were smaller (X_ratio: r = .24, AUC: r = .17).

Table 1: Change from baseline in the studied variables by type of manipulation.

	Manipulation		Test of group difference		
	Intuitive (n=30) M (SD)	Analytic (n=28) M (SD)	t	df	p
d-prime	–0.13 (0.83)	0.03 (0.86)	–0.75029	55.296	0.456
c (criterion)	–0.23 (0.43)	–0.14 (0.41)	–0.88562	55.916	0.380
Error rate	0.03 (0.09)	0.00 (0.08)	0.95679	55.904	0.343
RT (log)	–0.38 (0.18)	0.06 (0.19)	–9.1187	54.674	<.001
X_ratio	0.01 (0.07)	0.01 (0.06)	–0.13786	55.545	0.891
AUC	0.10 (0.21)	–0.01 (0.22)	2.0129	55.247	0.049

3.2.1 Ability to distinguish fallacies from reasonable arguments

To describe how participants distinguished fallacious from reasonable arguments, we first used the error rates to calculate the d' (d-prime, discriminability) and c (criterion) measures

(Green & Swets, 1966). Based on signal detection theory (SDT), the discriminability measure indicates the ability to distinguish between response options, and the criterion measure indicates the participants' bias to prefer one response option over the other. In the present study, discriminability describes participants' individual abilities at distinguishing fallacies as being weak, and well-justified arguments as being strong. The c measure reflects the participants' overall tendency to rate arguments as being either strong or weak, regardless of actual strength. Larger criterion values indicate a higher threshold for accepting an argument as being strong.

Examination of the SDT measures (Table 1) and the error rates on fallacies and reasonable arguments separately (Figure 1A) indicates that participants who were encouraged to think intuitively relaxed their criterion such that they started rating both fallacies and reasonable arguments as more often being strong. Those who were encouraged to think analytically did not seem to change the way they assessed arguments. However, these trends were not statistically significant. Overall, the error rates were very low throughout the experiment.

3.2.2 Response times

As predicted, the manipulations affected RT differently. As Figure 1B shows, responses became quicker after the intuitive manipulation. After the analytic manipulation, responses seemed to become slower. As noted by a reviewer, these results offer additional support to the validity of the manipulations.

3.2.3 Analyses of trajectories

Figure 1C shows the proportions of the trajectories spent on the opposite side (the X_ratio). As the figure shows, there was no significant difference between the effects of the two manipulations on this variable.

However, as Figure 1D shows, the two manipulations affected AUC differently. Among those encouraged to respond intuitively, trajectories became more curved. Figure 2 illustrates the curvature of the trajectories averaged over trials. Note that the actual trajectories were highly variable. Figure 3 shows the actual trajectories of two participants.

3.3 Correct versus error trials

To test differing predictions derived from corrective and logical intuition accounts of DPT, we compared all dependent variables across trials with correct and erroneous responses. Table 2 shows descriptive statistics for these variables. T-tests showed that RT, the X_ratio, and AUC were larger for error trials than for correct trials. This indicates that when making errors, participants' mouse trajectories exhibited more attraction towards the correct response and strayed longer into the opposite side, than they did towards the incorrect response when responding correctly. This is clearly visible in Figure 2, which shows that

the average trajectories of error trials initially leered towards the correct response button, before moving to the incorrect one.

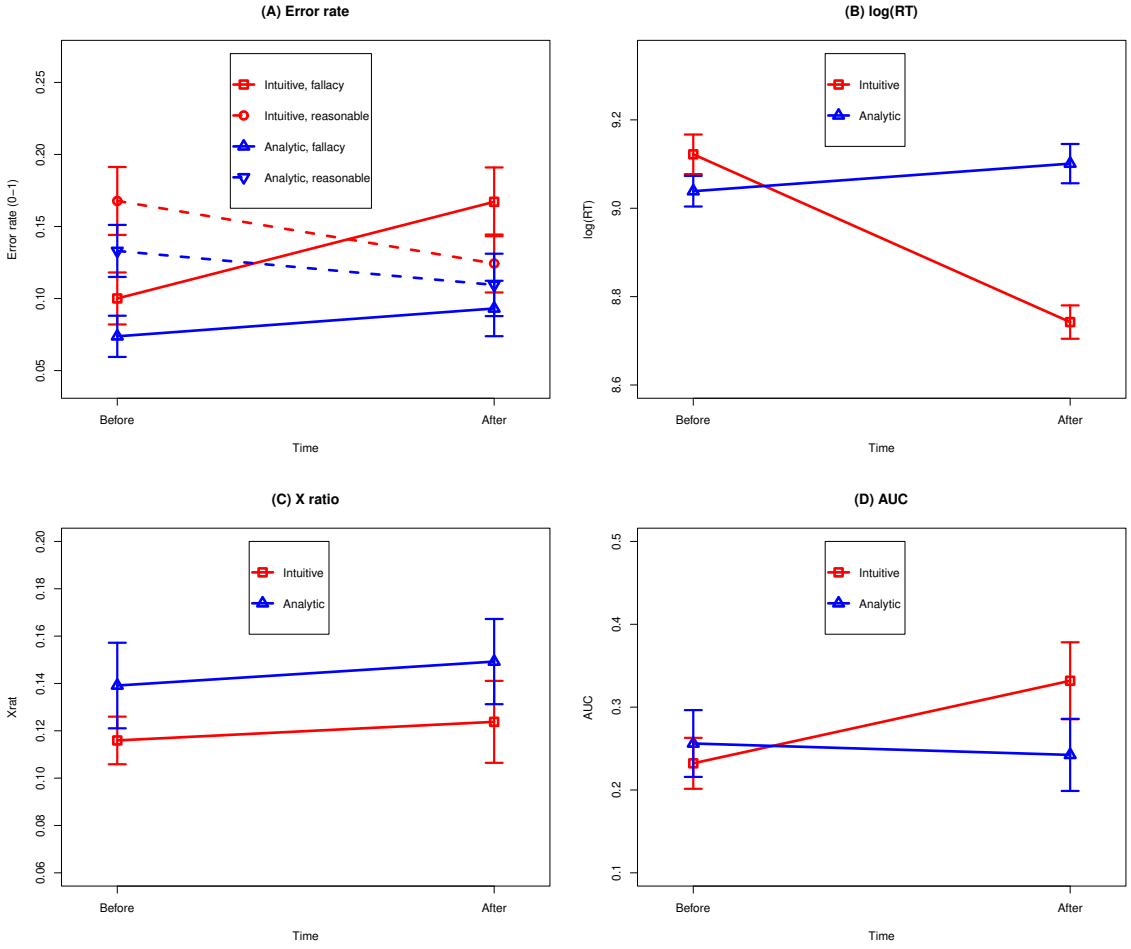

FIGURE 1: Means of the studied measures before and after the different manipulations. Results concerning (A) error rate, (B) response times, (C) X_ratio, (D) area under curve. Error bars are 1 standard error of the mean.

TABLE 2: Descriptive statistics for measures on trials ending in correct responses or errors.

	Correct trials M (SD)	Error trials M (SD)	t (df=57)	p
RT	8.96 (0.22)	9.38 (0.32)	–11.234	<.001
X_ratio	0.12 (0.08)	0.22 (0.14)	–5.7054	<.001
AUC	0.23 (0.19)	0.54 (0.38)	–6.5733	<.001

The error rates were too low to reliably investigate whether the intuitive and analytic manipulations differently affected the RT, X_ratio, or AUC of correct trials and error trials.

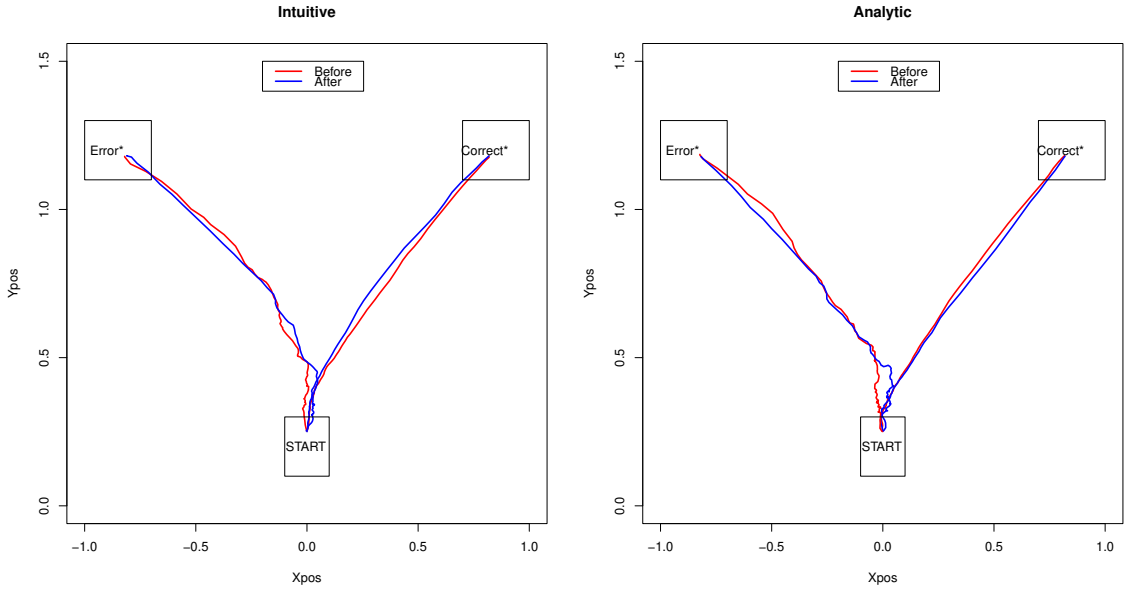

FIGURE 2: Average trajectories before and after manipulations. In the experiment, the response buttons were labeled “Strong” and “Weak”. In the figure, all trajectories have been turned so that the correct responses (classifying fallacies as weak and reasonable arguments as strong) end on the right, all incorrect responses on the left.

3.4 Final checks

To rule out that the results were affected by how familiar the participants were with argumentation analysis, we ran the analyses again, excluding those who reported having ‘much’ knowledge about argumentation analysis. We also inspected all dependent variables for systematic differences between the participants reporting different levels of knowledge. No systematic differences were found.

We also ran all analyses separately for fallacies and reasonable arguments, but as the pattern of results was highly similar and separate analyses indicated no further conclusions, we report the results for both types of arguments pooled (except the error rates for fallacies and reasonable arguments, which Figure 1A shows separately for clarity).

Finally, we ran all analyses with the ambiguous arguments included. The pattern of results and the conclusions were unchanged.

4 Discussion

How people distinguish fallacies as being weak arguments has large practical meaning for their ability to make reasoned choices, with implications for many areas of life. In the present study, argument literacy was examined in light of a prominent notion in the literature on formal reasoning, namely that people are cognitive misers. A cognitive miser saves effort by forgoing careful scrutiny of all relevant aspects of a task and settles for an easy answer

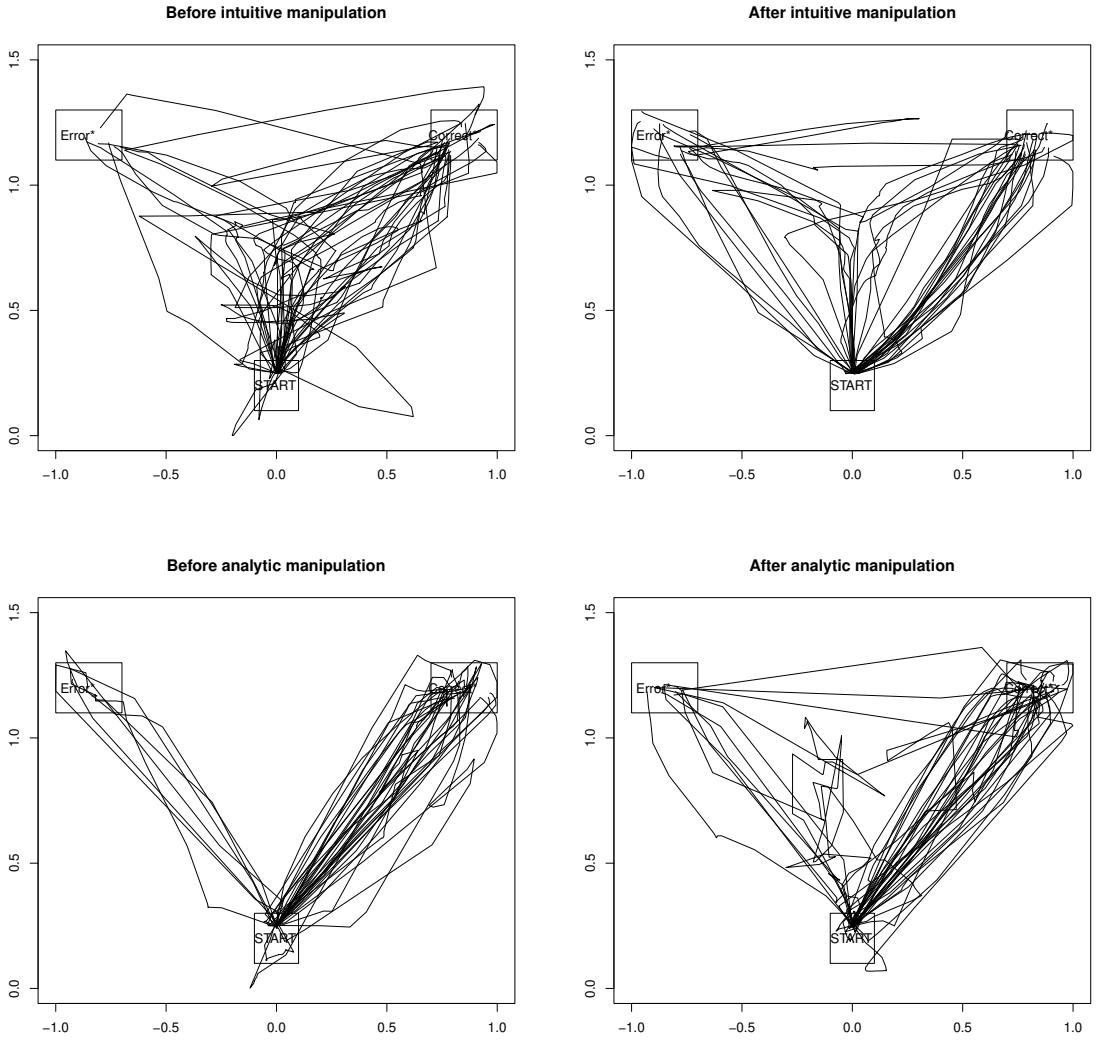

FIGURE 3: Illustration of the mouse trajectories of one participant before and after the intuitive manipulation (top row) and one participant before and after the analytic manipulation (bottom row).

(Toplak et al., 2011, 2014). The present paper asked whether humans are cognitive misers even when evaluating informal arguments. Studies indicate that evaluating the relationship between reasons and claims is the most demanding stage of argument analysis and one that is rarely engaged in spontaneously (Shaw, 1996). Thus, miserliness in evaluating arguments could manifest as reasoners avoiding evaluating whether the reasons support the claims, basing their responses instead on more superficial features or general believability of the presented information (Kuhn & Modrek, 2017; Shaw, 1996).

To test for miserliness in argument evaluation, we subjected our study participants to the simplest of manipulations: instructions to use either a more intuitive way of thinking, or a more analytic way of thinking. Under instructions to think in one of these ways, participants

evaluated whether arguments were weak or strong. To track argument evaluation as it unfolded, we analyzed error rates, response times, and indices of decisional conflict as measured through mouse tracking. In addition, we compared the cursor trajectories on trials that ended in correctly rating fallacies as weak and reasonable arguments as strong, to trials that ended in incorrectly accepting fallacies or rejecting reasonable arguments. The research results were only partly in line with the idea of a cognitive miser.

First, response times changed as a function of thinking style manipulations in the way predicted by dual-process theories, decreasing in the intuitive condition and increasing after the analytic probe. Response times are often interpreted as indicative of the amount of conflict, with faster responses taken to be less conflicted (Bonner & Newell, 2010; Voss et al., 1993). The present RT results are compatible with this view. They could suggest that in the intuitive condition, our participants gave carefree intuitive responses, while in the analytic condition they experienced more conflict and used more reasoning. However, such a conclusion is not plausible because the increased time spent in the analytic condition had no bearing on either the responses that participants gave or on the cursor trajectories leading to these responses, as we will discuss next.

Examining the responses that participants gave makes it clear that correct responses did not require analytic processing. Participants correctly tended to see that fallacies are weak and reasonable arguments are strong even when they were under instructions to think intuitively and answering quickly. By the corrective account, encouraging people to think intuitively should have increased errors, and encouraging analytic thinking should have decreased errors (Evans, 2008). In contrast, by an account that postulates smart intuitions (De Neys, 2018), instructions to think in either way should have had little effect on performance. The results showed no significant difference in how the manipulations affected the error rates. However, the overall low error rate makes it difficult to interpret the results as supporting either account, as the lack of effects might be due to a ceiling effect.

Fortunately, the trajectory analyses offer clearer conclusions. We may note that the present findings are in line with previous findings documenting that mouse tracking can reveal processes that are not easily available to introspection and also, that mouse tracking indices are at least partly dissociable from RT (Stillman et al., 2018). While RTs may be influenced by multiple factors, research indicates that mouse paths more reliably reflect decisional difficulty arising from ambivalence towards the target and from decision changes (Maldonado et al., 2019; Schneider & Schwarz, 2017; Schneider et al., 2015).

The two measures describing trajectories that we used, AUC and X_ratio, responded differently to the thinking style manipulations. AUC grew larger after the intuitive instructions, and smaller after the analytic instructions. If we assume that AUC is an indication of the amount of decisional conflict that participants experience, this result is the opposite of what DPTs would predict. However, since these responses also became quicker, the most likely explanation seems to be that the increased AUC is an artefact brought on by decreased precision as people were moving the mouse quickly. It should be noted that AUC values

increase even if trajectories stay on the side of the final response button throughout the trial. Thus, they may not be the ideal measure to capture the types of mind changes that DPTs are concerned with.

In contrast to AUC, the X_ratio may be a more appropriate measure for describing changes of mind, as the X_ratio describes the extent to which the cursor trajectory strays over to the side of the screen opposite from the final response. In the present experiment, the X_ratio was equally unaffected by both manipulations.

The most interesting differences were revealed by the comparisons between trials with correct responses and trials with incorrect responses. The corrective, default-interventionist account predicts that trials that end in correct responses should exhibit attraction towards the incorrect response, reflecting incorrect heuristics, followed by corrective mouse movements, reflecting the activation of analytic thinking processes (Evans, 2008). Incorrect responses, in turn, should be more straightforward as they would involve no analytic intervention. The present results did not conform to such a model.

The results revealed the opposite: corrective movements were more common on trials that ended in errors, as shown by both AUC and the X_ratio. This type of finding is predicted by an account that postulates smart intuitions (De Neys, 2018). By such an account, no corrective movements are necessary to respond correctly. When making errors, however, the existence of these smart intuitions, or some type of implicit monitoring for conflicts between heuristics and normative principles (Johnson et al., 2016), is assumed to influence the decision making process. The present results, in which the trajectories on error trials showed attraction towards the correct response option, could be predicted by this model. Thus, when responding incorrectly, participants seemed to have some implicit awareness that fallacies are weak, and that reasonable arguments are strong. Yet they gave in to temptation and responded that the fallacies were strong or that the reasonable arguments were weak.

Previous evidence for "logical intuition" accounts has stemmed from studies focused on formal reasoning tasks. These studies have presented evidence that for many people, some principles of good reasoning have been automatized to the point of having become intuitive (e.g., Brisson et al., 2014; Handley et al., 2011; Newman et al., 2017; reviewed in De Neys, 2012, 2018). The present study extends this existing research by showing that a similar effect also seems to apply to the evaluation of everyday arguments. Thus, even though there are no definitive correct or incorrect responses on informal arguments in the way that there are on formal reasoning tasks (Angell, 1964), comparable cognitive processes may be involved in how people solve both types of thinking tasks. This finding is theoretically important because until now, research investigating whether the same cognitive processes apply to both formal and informal reasoning has been scarce and the findings inconclusive (Thompson & Evans, 2012).

The present study is also the first to report support for smart intuitions using mouse tracking methodology. Previous mouse tracking studies with reasoning tasks have been

compatible with corrective DPTs, as they have found corrective cursor movements to be related to correct responses and better reasoning ability (Szaszi et al., 2018; Travers et al., 2016). The difference between our results and those of Szaszi et al. and Travers et al. may reflect the type of task used. Travers et al. (2016) discuss the possibility that on the CRT, which they studied, the logical and other principles needed to correctly solve the tasks may not be easy to internalize. The logical intuition account is largely based on studies using logical syllogisms and base-rate tasks (De Neys, 2012). It may be easier to internalize the logical principles needed on those tasks because they are relatively simple. In this comparison, the processes needed to evaluate fallacies and reasonable everyday arguments thus seem to be more similar to the processes needed for logical reasoning than to the perhaps more complex types of processes that are needed on the CRT. This is somewhat surprising, given that informal arguments, by definition, concern problems that are debatable, and that may have no definitive solutions (Angell, 1964; Galotti, 1989; Kuhn, 1991). The present task format likely required less complex processing than is needed to evaluate everyday arguments in more naturalistic settings.

4.1 Caveats

A few methodological issues regarding this study are important to be discussed further. One is that the present was a high-achieving population, as indicated by the overall low error rate. To establish whether argument evaluation involves smart intuitions more generally, it is important to replicate the present analyses in populations who more often accept fallacies at baseline. It is possible that for people who are generally good at telling fallacies from reasonable arguments, allowing themselves to respond incorrectly is particularly difficult.

A related observation is that the context of participating in a reasoning study on university premises may have acted as an unintended manipulation to think more analytically from the outset, and unintended demand effects may have been at play (Phillips et al., 2016). Studies indicate that the potential to increase analytic thinking experimentally tends to be small in any case (Lawson et al., 2020), and the present context may have restricted it further. Therefore, more research is needed to find whether the present findings replicate in other settings, or whether more intuitive thinking may decrease accuracy, and more analytic thinking may increase accuracy on argument evaluation, if the baseline is less high. For example, evaluating arguments in more naturalistic settings may better reveal how people typically think about arguments.

Unfortunately, the small sample size and resulting low statistical power precluded more detailed analyses of the effects of background variables, such as previous exposure to argumentation terminology and concepts. Thus, studies with larger samples are needed to investigate possible moderators of the ability and threshold for recognizing argument quality. Argument evaluation likely interacts with individual characteristics of the reasoner, such as reasoners' more stable preferences for intuitive or analytic thinking. Influences of reasoners' characteristics are likely, because thinking styles tend to predict performance on

most reasoning tasks and to affect thinking and behavior widely (Pennycook et al., 2015; Petty et al., 2009; Toplak et al., 2011). Particularly relevant for evaluating arguments, a recent study showed that understanding scientific argumentation is positively predicted by general cognitive abilities and epistemological sophistication (Münchov et al., 2019). More research is needed to determine whether this also applies to fallacies and to everyday informal argumentation more generally.

By the present results it is not clear whether the findings apply to all poorly justified arguments or only to fallacies. Fallacies are defined as being unreasonable but psychologically compelling (e.g., Copi & Burgess-Jackson, 1996). That is, fallacies are not just any type of weak argument but a very specific class of weak argument that are often used for rhetoric effect precisely because there is something in them that makes people accept them. In future research, an interesting question is how people evaluate other kinds of weak arguments. There can also be arguments that are weak in a mundane, more obvious way, without conforming to any of the famous formats known as fallacies. For example, an argument such as "The postal service in Finland is excellent since it is available to anyone" is not among the famous types of fallacies recognized in textbooks, but nevertheless cannot be considered particularly strong because the reasons given are not sufficient in relation to the strength of the claim being made.

Finally, bringing research on argumentation to the cognitive lab comes with some challenges. Argumentation is an inherently social activity, and integral to argumentation is that the audience or opponent brings up counterarguments. Some theorists argue that reasoning developed to enable argumentation with other people (Haidt, 2012; Mercier, 2016). This can be contrasted with DPTs, which largely construe reasoning as if the function of reasoning were to correct intuition when it is factually incorrect. Research shows consistently that the role that an individual holds in a debate – for or against a proposition – significantly affects the ability to recognize fallacies as being weak, and that reasoning performance on a variety of tasks improves if they are framed in the context of a dialogue (Mercier, 2016; Neuman et al., 2006). Thus, future studies could benefit from incorporating social factors into the experimental paradigms alongside the modes of thinking that were the focus of the present study, to build a more comprehensive picture of how and why some people fall for fallacies while others do not.

4.2 Conclusions and future directions

How people understand informal argumentation, which is ubiquitous in all communication, has not been studied extensively from a cognitive perspective. This study presented evidence supporting the notion that falsely accepting fallacies or rejecting reasonable arguments is more complicated than is correctly rejecting fallacies and acknowledging reasonable arguments as being strong. These findings point to fundamental similarities between how people understand informal logic and recent accounts of how they understand formal logic (De Neys, 2018). Importantly, the results were not compatible with corrective dual-process

models. There were no indications that responding correctly would be laborious and require analytic intervention, or that boosting efforts to think analytically would make people more likely to change their mind after forming an initial hunch. Thus, the widespread trust in bad argumentation that is obvious outside the reasoning laboratory may to some extent not be due to people not using enough mental effort to evaluate the arguments they encounter, but to a process in which they are at least implicitly aware that the argumentation is flawed but for reasons that are beyond the present paper, they choose to accept the poor arguments and reject the strong ones.

5 References

Angell, R. B. (1964). *Reasoning and logic*. New York: Appleton-Century-Crofts.

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

Bonnefon, J.-F. (2012). Utility conditionals as consequential arguments: A random sampling experiment. *Thinking & Reasoning, 18*, 379–393. http://dx.doi.org/10.1080/13546783.2012.670751.

Bonner, C., & Newell, B. R. (2010). In conflict with ourselves? An investigation of heuristic and analytic processes in decision making. *Memory & Cognition, 38*, 186–196. http://dx.doi.org/10.3758/MC.38.2.186.

Brisson, J., de Chantal, P.-L., Lortie Forgues, H., & Markovits, H. (2014). Belief bias is stronger when reasoning is more difficult. *Thinking & Reasoning*. http://dx.doi.org/10.1080/13546783.2013.875942.

Copi, I. M., & Burgess-Jackson, K. (1996). *Informal logic*. Englewood Cliffs, NJ: Prentice Hall.

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

De Neys, W. (Ed.) (2018). *Dual process theory 2.0*. London: Routledge.

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

De Neys, W., Moyens, E., & Vansteenwegen, D. (2010). Feeling we're biased: Autonomic arousal and reasoning conflict. *Cognitive, Affective & Behavioral Neuroscience, 10,* 208–216. http://dx.doi.org/10.3758/CABN.10.2.208.

De Neys, W., & Pennycook, G. (2019). Logic, fast and slow: advances in dual-process theorizing. *Current Directions in Psychological Science, 28*, 503–509. http://dx.doi.org/10.1177/0963721419855658.

Downes, S. (1995). Stephen's guide to the logical fallacies. Retrieved from http://www.fallacies.ca/aa.htm.

Evans, J. St. B. T. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Annual Review of Psychology, 59*, 255–278. http://dx.doi.org/10.1146/annurev.psych.59.103006.093629.

Evans, J. St. B. T., & Stanovich, K. E. (2013). Dual-process theories of higher cognition: Advancing the debate. *Perspectives on Psychological Science, 8*, 223–241. http://dx.doi.org/10.1177/1745691612460685.

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

Freeman, J. B., & Ambady, N. (2010). MouseTracker: Software for studying real-time mental processing using a computer mouse-tracking method. *Behavior Research Methods, 42*, 226–241. http://dx.doi.org/10.3758/BRM.42.1.226.

Fumagalli, R. (2020). Slipping on slippery slope arguments. *Bioethics, 34,* 412–419. http://dx.doi.org/10.1111/bioe.12727.

Galotti, K. M. (1989). Approaches to studying formal and everyday reasoning. *Psychological Bulletin, 105*, 331–351. http://dx.doi.org/10.1037/0033-2909.105.3.331.

Gill, M. J., & Ungson, N. D. (2018). How much blame does he truly deserve? Historicist narratives engender uncertainty about blameworthiness, facilitating motivated cognition in moral judgment. *Journal of Experimental Social Psychology, 77*, 11–23. http://dx.doi.org/10.1016/j.jesp.2018.03.008.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics* (Vol. 1). New York: Wiley.

Gürçay, B., & Baron, J. (2017). Challenges for the sequential two-system model of moral judgement. *Thinking & Reasoning, 23,* 49–80. http://dx.doi.org/10.1080/13546783.2016.1216011.

Hahn, U. (2020). Argument quality in real world argumentation. *Trends in Cognitive Sciences, 24*, 363–374. http://dx.doi.org/10.1016/j.tics.2020.01.004.

Hahn, U., & Oaksford, M. (2007). The rationality of informal argumentation: a Bayesian approach to reasoning fallacies. *Psychological Review, 114*, 704–732. http://dx.doi.org/10.1037/0033-295X.114.3.704.

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

Hamblin, C. L. (1970). *Fallacies*. London: Methuen.

Handley, S. J., Newstead, S. E., & Trippas, D. (2011). Logic, beliefs, and instruction: A test of the default interventionist account of belief bias. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*, 28–43. http://dx.doi.org/10.1037/a0021098.

Hornikx, J., & Hahn, U. (2012). Reasoning and argumentation: towards an integrated psychology of argumentation. *Thinking & Reasoning, 18*, 225–243. http://dx.doi.org/10.1080/13546783.2012.674715.

Johnson, E. D., Tubau, E., & De Neys, W. (2016). The doubting system 1: Evidence for automatic substitution sensitivity. *Acta Psychologica, 164,* 56–64. http://dx.doi.org/10.

1016/j.actpsy.2015.12.008.

Kahneman, D. (2003). A perspective on judgment and choice: Mapping bounded rationality. *American Psychologist, 58*, 697–720. http://dx.doi.org/10.1037/0003-066X.58.9.697.

Kahneman, D. (2011). *Thinking, fast and slow*. London: Penguin Books.

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

Kuhn, D. (1991). *The skills of argument*. Cambridge, NY: Csambridge University Press.

Kuhn, D., & Modrek, A. (2017). Do reasoning limitations undermine discourse? *Thinking & Reasoning, 24*, 97–116. http://dx.doi.org/10.1080/13546783.2017.1388846.

Lawson, M. A., Larrick, R. P. & Soll, J. B. (2020). Comparing fast thinking and slow thinking: The relative benefits of interventions, individual differences, and inferential rules. *Judgment and Decision Making, 15,* 660–684.

Makowski. (2018). The psycho package: an efficient and publishing-oriented workflow for psychological science. *Journal of Open Source Software, 3*, 470. http://dx.doi.org/10.21105/joss.00470.

Maldonado, M., Dumbar, E., & Chemla, E. (2019). Mouse tracking as a window into decision making. *Behavior Research Methods, 51*, 1085–1101. http://dx.doi.org/10.3758/s13428-018-01194-x.

Mercier, H. (2016). The argumentative theory: predictions and empirical evidence. *Trends in Cognitive Sciences, 20*, 689–700. http://dx.doi.org/10.1016/j.tics.2016.07.001.

Morewedge, C. K., & Kahneman, D. (2010). Associative processes in intuitive judgment. *Trends in Cognitive Sciences, 14*, 435–440. http://dx.doi.org/10.1016/j.tics.2010.07.004.

Münchov, H., Richter, T., von der Mühlen, S., & Schmid, S. (2019). The ability to evaluate arguments in scientific texts: Measurement, cognitive processes, nomological network, and relevance for academic success at the university. *British Journal of Educational Psychology, 89*, 501-523. http://dx.doi.org/10.1111/bjep.12298.

Neuman, Y., Glassner, A., & Weinstock, M. (2004). The effect of a reason's truth-value on the judgement of a fallacious argument. *Acta Psychologica, 116*, 173–184. http://dx.doi.org/10.1016/j.actpsy.2004.01.003.

Neuman, Y., Weinstock, M. P., & Glasner, A. (2006). The effect of contextual factors on the judgement of informal reasoning fallacies. *The Quarterly Journal of Experimental Psychology, 59*, 411–425. http://dx.doi.org/10.1080/17470210500151436.

Newman, I. R., Gibb, M., & Thompson, V. A. (2017). Rule-based reasoning is fast and belief-based reasoning can be slow: Challenging current explanations of belief-bias and base-rate neglect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 43,* 1154-1170. http://dx.doi.org/10.1037/xlm0000372.

Novak, T. P., & Hoffman, D. L. (2009). The fit of thinking style and situation: New measures of situation-specific experiential and rational cognition. *Journal of Consumer Research, 36*, 56–72. http://dx.doi.org/10.1086/596026.

Pennycook, G., Fugelsang, J. A., & Koehler, D. J. (2015). Everyday consequences of analytic thinking. *Current Directions in Psychological Science, 24*, 425–432. http://dx.doi.org/10.1177/0963721415604610.

Petty, R. E., Briñol, P., Loersch, C., & McCaslin, M. J. (2009). The Need for Cognition. In M. R. Leary & R. H. Hoyle (Eds.), *Handbook of individual differences in social behavior* (pp. 318-329). NY & London: Guilford Press.

Petty, R. E., & Cacioppo, J. T. (1986). The Elaboration Likelihood Model of persuasion. *Advances in Experimental Social Psychology, 19*, 123-205. http://dx.doi.org/10.1016/S0065-2601(08)60214-2.

Phillips, W. J., Fletcher, J. M., Marks, A. D. G., & Hine, D. W. (2016). Thinking styles and decision making: A meta-analysis. *Psychological Bulletin, 142*, 260–290. http://dx.doi.org/10.1037/bul0000027.

R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Revelle, W. (2021). psych: procedures for psychological, psychometric, and personality research (Version R package version 2.1.3): Northwestern University, Evanston, Illinois. Retrieved from https://CRAN.R-project.org/package=psych.

Sagan, C. (1997). *The demon-haunted world: science as a candle in the dark*. London: Headline Book Publishing.

Schellens, P. J., Šorm, E., Timmers, R., & Hoeken, H. (2017). Laypeople's evaluation of arguments: Are criteria for argument quality scheme-specific? *Argumentation, 31*, 681–703. http://dx.doi.org/10.1007/s10503-016-9418-2.

Schneider, I. K., & Schwarz, N. (2017). Mixed feelings: the case of ambivalence. *Current Opinion in Behavioral Sciences, 15*, 39–45. http://dx.doi.org/10.1016/j.cobeha.2017.05.012.

Schneider, I. K., van Harreveld, F., Rotteveel, M., Topolinski, S., van der Pligt, J., Schwarz, N., & Koole, S. L. (2015). The path of ambivalence: tracing the pull of opposing evaluations using mouse trajectories. *Frontiers in Psychology, 6,* 1–12. http://dx.doi.org/10.3389/fpyg.2015.00996.

Shaw, V. F. (1996). The cognitive processes in informal reasoning. *Thinking & Reasoning, 2*, 51–80. http://dx.doi.org/10.1080/135467896394564.

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

Stillman, P. E., & Ferguson, M. J. (2018). Decisional conflict predicts impatience. *Journal of the Association for Consumer Research, 4,* 47-56. http://dx.doi.org/10.1086/700842.

Stillman, P. E., Shen, X., & Ferguson, M. J. (2018). How mouse tracking can advance social cognitive theory. *Trends in Cognitive Sciences, 22*, 531–543. http://dx.doi.org/10.1016/j.tics.2018.03.012.

Szaszi, B., Palfi, B., Szollosi, A., Kieslich, P. J., & Aczel, B. (2018). Thinking dynamics and individual differences: Mouse-tracking analysis of the denominator neglect task.

Judgment and Decision Making, 13, 23–32.

Tenhunen, V. (1998). Argumentionnin virheet. Retrieved from https://skepsis.fi/jutut/virhelista.html.

Thompson, V., & Evans, J. St. B. T. (2012). Belief bias in informal reasoning. *Thinking & Reasoning, 18,* 278-301. http://dx.doi.org/10.1080/13546783.2012.670752.

Tindale, C. W. (2007). *Fallacies and argument appraisal*. New York: Cambridge University Press.

Toplak, M., West, R. F., & Stanovich, K. E. (2014). Assessing miserly information processing: An expansion of the Cognitive Reflection Test. *Thinking & Reasoning, 20*, 147–168. http://dx.doi.org/10.1080/13546783.2013.844729.

Toplak, M., West, R. F., & Stanovich, K. E. (2011). The Cognitive Reflection Test as a predictor of performance on heuristics-and-biases tasks. *Memory & Cognition, 39*, 1275–1289. http://dx.doi.org/10.3758/s13421-011-0104-1.

Travers, E., Rolison, J. J., & Feeney, A. (2016). The time course of conflict on the Cognitive Reflection Test. *Cognition, 150*, 109-118. http://dx.doi.org/10.1016/j.cognition.2016.01.015.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185*, 1124–1131. http://dx.doi.org/10.1007/978-94-010-1834-0_8.

van Eemeren, F., Garssen, B., & Meuffels, B. (2009). *Fallacies and judgments of reasonableness. Empirical research concerning the Pragma-Dialectical Discussion Rules*. Dordrecht Heidelberg London New York: Springer.

Voss, J. F., Blais, J., Means, M. L., Greene, T. R., & Ahwesh, E. (1986). Informal reasoning and subject matter knowledge in the solving of economics problems by naive and novice individuals. *Cognition & Instruction, 3,* 269-302. http://dx.doi.org/10.1207/s1532690xci0303_7.

Voss, J. F., Fincher-Kiefer, R., Wiley, J., & Silfies, L. N. (1993). On the processing of arguments. *Argumentation, 7,* 165–181. http://dx.doi.org/10.1007/Bf00710663..

Voss, J. F., & Means, M. L. (1991). Learning to reason via instruction in argumentation. *Learning & Instruction, 1,* 337–350. http://dx.doi.org/10.1016/0959-4752(91)90013-X.

Weinstock, M. P., Neuman, Y., & Tabak, I. (2004). Missing the point or missing the norm? Epistemological normals as predictors of students' ablity to identify fallacious arguments. *Contemporary Educational Psychology, 29*, 77–94. http://dx.doi.org/10.1016&S0361-476X(03)00024-9.

Appendix A

Arguments used in the study, translated from Finnish. Item numbers refer to the original item numbers in the raw data.

Fallacies included in the analyses

1. Climate change activist are wrong – they want people to stop having children.
2. Vegans cannot be trusted for their opinions, because they care more about the wellbeing of veals than about kids getting enough calcium.
3. Drinks with a higher alcohol content than beer should not be sold in grocery stores, because it would lead to youth being corrupted and drinking breezers.
4. The defendant is guilty, because there is no evidence of his innocence.
5. Schools should start serving three vegetarian meals a week – those against are welcome to feed their kids at home.
6. The death penalty should not be abolished. Supporting abolishment means letting monstrous deeds go unpunished.
7. Cannabis should not be legalized. If it is legalized, the use of other drugs will increase, too.
8. There should be no free coffee in workplaces, because free coffee leads to people taking more breaks and working less.
9. Computers cannot really think, because if they did, it would mean we are nothing but machines.
11. Don't eat meat. Eating meat means you don't care about animals.
14. Adoption should not be made any easier, because if it is, many will start adopting children from overseas and our gene pool and culture will be impoverished.
15. The authorities should not allow abortion even in the first weeks of pregnancy, because if they do that, there is no limit to how late they may next allow it.
16. We should not accept credit cards that contain microchips, because they will lead us all to being under surveillance and losing all privacy.
17. Violence on television causes youth violence. The fact that youth violence exists is proof that television is to blame.
18. Petting cats relieves stress. Hence, cat owners are more easygoing than other people.
19. It's better not to vaccinate children, because vaccinations cause serious illnesses.
20. The intelligence of the Americans is low because half of them have IQs below the average.
21. The average height of Finnish women can't be 166 cm, because I don't know any Finnish women of that height.
22. Apparently people are generally in favor of lowering taxes, because all of my friends are in favor of it.
23. Homeopathy is a good cure for the flu – it cured Pekka's flu.

24. We should give up switching back and forth to Daylight Savings Time because some people have heart attacks after these switches.
25. I don't believe that more people have become vegetarians, because I still don't know any vegetarians.
26. Vaccines are dangerous because my nephew and others got a high fever after having their Swine Flu shots.
28. Coffee improves exam performance. I did poorly on my exam because I didn't have coffee the morning before my exam.
30. Since Pekka has almost no education, his political opinions should be ignored.
31. Sauli Niinistö [President of Finland] supports shortening the presidential term. Thus, the presidential term should be shortened.
32. Juha Sipilä [Prime Minister] has stated that Finland should not receive ISIS-prisoners from Syria. Thus, we should not do it.
33. Many convicts have been tried for fraud, so their complaints about prisons are not to be trusted.
34. Many psychologists have abandoned psychoanalysis because of its inanity, so psychoanalysis must be abandoned.
35. In recent years, dozens of respected scientists have denied climate change. Thus, we should have some reservations about it.
36. In many countries, it's normal to punish kids by isolating them. This goes to show that isolation teaches kids to behave.
37. All cultures tell tales of supernatural phenomena, which indicates that the phenomena are real.
39. No extraterrestrial life has been found. Thus, no other creature like humans can exist.
40. There is not a single piece of evidence for dark matter – thus, it does not exist.
41. Many mental health problems cannot be measured through studying brain functions, so they are not real.
42. Using drugs causes crime. If we can end drug use, we will end crime.
43. Smoking causes lung cancer. If people wouldn't smoke, they wouldn't get lung cancer.
44. Western civilization was born under Christianity, so giving up faith would destroy Western civilization.
45. Heikki is taller than Ilkka and Ilkka is taller than Jussi, which makes Heikki shorter than Jussi.
46. Tuomas is older than Juhani and Juhani is older than Aapo. Thus we can conclude that Tuomas is younger than Aapo.
47. Stuff can't appear from nowhere, so natural selection can't have given rise to a complicated organ such as the eye.
48. Most heroin users have also tried cannabis, so most cannabis users end up using heroin.
50. It has rained on the two previous Christmases, so it will likely also rain this Christmas.
51. A poll of fifty university students shows that more young adults are voting.

52. The horrible flight accident yesterday proves that flying is dangerous.
53. Because Social Services pays the bills of people who have no money, it should pay everyone else's bills, too.
54. Smoking and cancer are related. Thus, cancer is usually caused by smoking.
55. People with a low educational level are more often marginalized. Therefore marginalization is due to a lack of education.
56. Children who are irritated and restless in kindergarten often go down with a cold the following day. Thus, by tackling these problem behaviors, we can prevent diseases.
59. My husband claims that a three-month-old sleeps better in light places. That's something only a dad would say; just ignore it.
61. States are built on laws. Laws are true, because they have been decided upon by the governments.
62. By participating in the lottery regularly for ten years, the chances of winning the jackpot on an individual round increase.
63. Extraterrestrials have visited the Earth, but we can't see them, because we lack the appropriate equipment.

Reasonable arguments included in the analyses

2. The opinions of the teachers is important, because teachers are the ones who know most about the consequences of large classes.
4. Alcoholic beverages with a higher alcohol content than beer should not be sold in grocery stores because most studies show that selling them there increases alcohol-related problems.
7. The defendant might be guilty even if no evidence is found.
9. You shouldn't keep smoking. If you do, you are subjecting yourself to numerous health risks.
10. Fringe benefits should be improved, because it would likely increase employee satisfaction.
11. Recycling should be made easier, because then people would likely recycle more.
12. Public transportation should be improved because it would make it possible to decrease private motoring.
13. Selling alcohol should not be prohibited. Prohibiting it in Finland would likely just lead to people going abroad to buy it.
15. It is a good idea to vaccinate children because vaccinations protect from serious illnesses.
16. When gambling, one should always stop when one is winning, because the chances of a winning streak continuing are slim.
19. Shoes that fit well cause no blisters, because they conform to the shape of the foot as one wears them.
21. You should not drive while intoxicated because alcohol lowers reaction speed, and driving requires a good ability to react quickly.

24. Because Stephen Hawking was an expert in physics, his theories in the field of physics deserve attention.
28. 97 % of the world's climate scientists consider climate change to be caused by humans. Therefore it is likely that climate change is caused by humans.
29. Because a large proportion of high school students are exhausted due to schoolwork, the curriculum plans should take this issue into account.
31. Unemployment is a risk factor for marginalization, so by supporting the unemployed, we can decrease marginalization.
33. A diet with too much salt heightens blood pressure. Thus it can increase the risk of heart diseases.
34. Kaisa is shorter than Miina, and Miina is shorter than Oona, so Oona is taller than Kaisa.
35. Pentti is richer than Mauri. Mauri is richer than Yrjö; thus we can conclude that Yrjö is poorer than Pentti.
36. Female workers are equally hardworking and productive as their male colleagues, and therefore deserve equal pay for equal work.
39. Most heroin users have also tried cannabis. Thus, some cannabis users end up being heroinists.
43. Even if I once beat my friend at a game of chess, I cannot know if I am a better player than she is.
44. The weather is bad for driving and the risk for car crashes is high in this kind of weather, so it is safer to take the train.
45. Lack of sleep lowers the ability to concentrate. Children who are having problems concentrating should make sure to sleep enough.
46. Hot weather increases ice cream consumption. Thus people are more likely to eat ice cream in the summer than at other times.
47. Using a phone while driving lowers alertness and thus increases the risk for car crashes.
49. Down syndrome can only form if an individual has an extra chromosome. Without this extra chromosome the syndrome cannot appear.
56. I did not get the job because I am not qualified for it.
57. Because hairs are not preserved on fossils, we cannot use fossils to prove that dinosaurs did not have hair.

Ambiguous arguments not included in analyses (originally meant as fallacies)

10. Don't drink alcohol. If healthy living is important to you, you can't drink alcohol.
12. You can eat as much pasta as you want because according to many Italian nutritionists, pasta is healthy.
13. You are probably going to fail your exam, because students like you always fail.
27. You shouldn't have unprotected sex, because it will give you an STD.

29. The reason people behave strangely is that they are driven by their unconscious instincts.
38. Most people think cannabis is bad for health, so it should not be legalized.
49. A poll of two hundred Helsinkians showed that in the next elections, the Green League will get the most seats.
57. Findings show that people in leading positions have large vocabularies, so those who want to succeed in their careers would do well to increase their vocabularies.
58. Drugs produced by criminal organizations cause substance abuse. Thus, the recent increase in substance abuse indicates that criminal organizations have become more active.
60. Human behavior is controlled by the soul, which no measurement devices can observe.

Ambiguous arguments not included in analyses (originally meant as reasonable arguments)

1. The Secretary is familiar with the issue, because she is the official in charge of the matter.
3. Trump's opinions are often questionable, because his factual knowledge of issues is often lacking.
5. The Presidential term should be shortened, because a majority of the Finnish people are in favor of shortening it.
6. You should be able to find the book in the library, because the database shows that it is not on loan.
8. Supernatural phenomena may exist, because it is impossible to prove that they do not.
14. The tax on sweets should not be lowered. Lowering it would not necessarily lead to an increased consumption of sweets.
17. Support for the Social Democratic Party is likely to increase, because a large national poll indicates it will.
18. We cannot know what kind of life there is outside our Solar system, because there is no equipment sophisticated enough to find out.
20. This law should not be passed, because the majority of MPs oppose it.
22. Because Jaana has not studied medicine, her anti-vaccine opinions should not receive attention.
23. Emma only has a low education. Thus, her language skills might be weak.
25. Because Jussi Halla-Aho [a politician] has a PhD in the humanities, he is not an expert on technical questions.
26. Donald Trump claims that large news companies are spreading fake news. Despite this, you can trust established media.
27. Most health experts say that cannabis is detrimental to health. Therefore it should not be legalized.
30. Because the value of money is determined by the market, the value of money fluctuates as the market situation changes.
32. Exercise lowers the risk for Type 2 diabetes. If everyone exercised enough, there would be fewer cases of diabetes.

37. Because the mean height of men is higher than that of women, a randomly selected man is likely to be taller than a randomly selected woman.
38. Even though I tossed a coin and got three tails in a row, the odds of getting heads on the next toss are no higher than the odds of getting tails.
40. Even though everyone I know can swim, there are also people who can't swim.
41. Even though the men I know support the idea of a standing army, it is not necessarily a good idea.
42. Even if a self-driving car is involved in an accident, it does not mean that self-driving cars are dangerous.
48. Hot summers are associated with an increase in death by drowning, which means that, on average, more people drown on hot days.
50. Exercise increases physical fitness. Even if you don't exercise every day, it is still possible that you are in good shape physically.
51. Based on human rights, same-sex marriages must be allowed. Those who oppose, can sign a petition.
52. It is possible to do well in school even if one has dyslexia.
53. If you have celiac disease, you can't eat wheat.
54. Taking the stairs instead of the elevator is good exercise for everyone, except for persons with a physical disability.
55. My breath became fresher when I used mouthwash.

Appendix B

Instructions at the beginning of the experiment

"Your task is to assess a series of claims by the quality of the justifications given, are they strong or weak. You should not let your own opinions affect your assessments. For example, you might disagree on whether studying Swedish should be compulsory in school, but in the following sentence, it is well motivated. In other words, the following is a STRONG argument: 'Swedish should remain a compulsory subject in school, because studies have shown that language skills in Swedish increase the employment opportunities of young people'. In contrast, the following argument is WEAK, because it is poorly motivated: 'Swedish should remain a compulsory subject in school, because it is also compulsory in Sweden'. The number of claims is large, and you should not linger on your responses, but answer fairly quickly."

Intuitive instructions (for half of the participants)

Halfway through the experiment:

"NEW INSTRUCTIONS FOR ANSWERING THE NEXT TASKS

People evaluate sentences like the ones you have just seen in two main ways, either intuitively or by thinking about them analytically and rationally. Preliminary evidence indicates that when evaluating these types of arguments, the best accuracy is usually achieved by trusting one's first impressions. We want to investigate this further and we therefore ask you to evaluate the next arguments purely based on your intuitive first impressions. Please do not scrutinize the arguments carefully. Instead, base your answers on your instincts."

After 20 trials in the intuitive condition:

"REMINDER OF INSTRUCTIONS:

Please remember to base your responses on the first impression that springs to your mind."

Analytic instructions (for half of the participants)

Halfway through the experiment:

"NEW INSTRUCTIONS FOR ANSWERING THE NEXT TASKS

People evaluate sentences like the ones you have just seen in two main ways, either intuitively or by thinking about them analytically and rationally. Preliminary evidence indicates that when evaluating these types of arguments, the best accuracy is usually achieved by thinking carefully and analytically about the sentences. We want to investigate this further and we therefore ask you to evaluate the next arguments as reasonably and carefully as you can. Please do not respond as you would based on your first impressions. Instead, base your answers on your logical reasoning."

After 20 trials in the analytic condition:

"REMINDER OF INSTRUCTIONS:

Please remember to evaluate the sentences and the reasons given in them rationally."