Overcoming Intuition: Metacognitive Difficulty Activates Analytic Reasoning

Adam L. Alter and Daniel M. Oppenheimer
Princeton University

Nicholas Epley
University of Chicago

Rebecca N. Eyre
Harvard University

Humans appear to reason using two processing styles: System 1 processes that are quick, intuitive, and effortless and System 2 processes that are slow, analytical, and deliberate that occasionally correct the output of System 1. Four experiments suggest that System 2 processes are activated by metacognitive experiences of difficulty or disfluency during the process of reasoning. Incidental experiences of difficulty or disfluency—receiving information in a degraded font (Experiments 1 and 4), in difficult-to-read lettering (Experiment 2), or while furrowing one's brow (Experiment 3)—reduced the impact of heuristics and defaults in judgment (Experiments 1 and 3), reduced reliance on peripheral cues in persuasion (Experiment 2), and improved syllogistic reasoning (Experiment 4). Metacognitive experiences of difficulty or disfluency appear to serve as an alarm that activates analytic forms of reasoning that assess and sometimes correct the output of more intuitive forms of reasoning.

Keywords: fluency, disfluency, dual-system processing, reasoning, judgment

Few psychological theories enjoy the longevity of William James's (1890/1950) suggestion that human reasoning involves two distinct processing systems: one that is quick, effortless, associative, and intuitive and another that is slow, effortful, analytic, and deliberate. When deciding whether it is more dangerous to travel by car or airplane, for instance, people may quickly generate horrific images of airline disasters and (erroneously) conclude that it is more dangerous to fly than to drive. Alternatively, they may think more analytically about the total number of automobile versus airline accidents, the number of miles driven versus flown per accident, or the possibility that automobile accidents are underreported whereas airline accidents command media headlines and conclude (accurately) that they are safer in an airplane.

Although not without controversy (see Kruglanski & Thompson, 1999; Osman, 2004),[1] dual-process theories have been used widely by developmental, cognitive, and social psychologists to explain such diverse phenomena as persuasion (e.g., Chaiken,

1980; Petty & Cacioppo, 1986), social cognition (Epley, Keysar, Van Boven, & Gilovich, 2004; Keysar & Barr, 2002), self-perception (Schwarz, 1998), causal attribution (Gilbert, 1989), stereotyping (Bodenhausen, Macrae, & Sherman, 1999), overconfidence (Griffin & Tversky, 1992), higher order reasoning (Evans, 2003; Evans & Over, 1996; Sloman, 1996), various memory phenomena (Jacoby, Kelley, & McElree, 1999; Jones & Jacoby, 2001; Whittlesea & Leboe, 2003), and a long list of non-Bayesian biases in judgment and decision making (e.g., Kahneman & Frederick, 2002). These dual-process theories enable understanding of diverse phenomena because they predict qualitatively different judgments depending on which reasoning system is used. In particular, deliberate and analytical systems of reasoning (System 2) can override or undo intuitive and associative (System 1) responses. Understanding when System 2 reasoning is likely to be used is therefore critical for understanding human judgment and decision making.

Although most dual-process models provide an extensive description of each system, few devote much attention to exactly when people will adopt each approach to information processing. Those that do typically argue that System 2 will be activated when

[1] Osman (2004) proposed a unitary process model that nonetheless acknowledges that processing occurs along a gradient from explicit to implicit, which mirrors System 1 and System 2 processing, respectively. Kruglanski and Thompson (1999; the *unimodel*) adopted a subtly different approach. They recognized that people make use of different types of information but subsume them under a single umbrella of "persuasive cues." They emphasized that people are ultimately trying to make sense of the world, regardless of which type of cue they use in a given situation. However, like Osman's unitary process model, the unimodel acknowledges that some forms of processing are more complex than others.

people have both the capacity and the motivation to engage in effortful processing. Existing research demonstrates that errors in System 1 reasoning are less likely to be corrected when people are under cognitive load or respond quickly (e.g., Bless & Schwarz, 1999; Chaiken, 1980; Petty & Cacioppo, 1986), but they are more likely to be corrected when people are accountable for their decisions (Tetlock & Lerner, 1999) and when the outcome is personally relevant (Ajzen & Sexton, 1999; Chaiken, 1980; Petty & Cacioppo, 1986). This existing research does not address, however, when people will recognize that System 1 processes might be producing faulty output that requires more analytical thought.

Accordingly, we investigated the novel question of when people are compelled to use System 2 processes in the first place. We predicted that people's use of more elaborate reasoning processes would be based on experiential cues that more elaborate reasoning processes are required and that System 2 processes would therefore be activated by cues that suggest a simple System 1 judgment might be faulty.

Confidence in the accuracy of intuitive judgment appears to depend in large part on the ease or difficulty with which information comes to mind (Gill, Swann, & Silvera, 1998; Kelley & Lindsay, 1993) and the perceived difficulty of the judgment at hand. If information is processed easily or fluently, intuitive (System 1) processes will guide judgment. If information is processed with difficulty or disfluently, however, this experience will serve as a cue that the task is difficult or that one's intuitive response is likely to be wrong, thereby activating more elaborate (System 2) processing. The ease or difficulty experienced while processing information is therefore used as a cue to guide one's subsequent processing styles. We thus predicted that experienced difficulty or disfluency would function as a signal that a simple and intuitive judgment was insufficient and that more elaborate cognitive processing would be necessary, thereby increasing System 2 processing. Indeed, neuroscientific evidence suggests that disfluency triggers the anterior cingulate cortex (Boksman et al., 2005), an alarm that activates the prefrontal cortex responsible for deliberative and effortful thought (Botvinick, Braver, Carter, Barch, & Cohen, 2001; Lieberman, Gaunt, Gilbert, & Trope, 2002; see also Goel, Buchel, Frith, & Dolan, 2000).

Previous research has shown that people rate disfluent stimuli more negatively than fluent stimuli across a range of domains. For example, people believe that disfluently named stocks will perform more poorly than will fluently named stocks (Alter & Oppenheimer, 2006), that disfluent prose is written by an author that is less intelligent than an author of fluent prose (Oppenheimer, 2006), and that disfluent aphorisms are less likely to be true than fluent aphorisms (McGlone & Tofighbakhsh, 2000). However, the role of disfluency in the current research is novel. Unlike existing research in which disfluency served as a direct cue to judgment, we investigate disfluency as an indirect cue that serves as a metacognitive signal to prompt more systematic processing. We predicted that experiencing difficulty or disfluency during the course of reasoning would trigger System 2 processes and decrease the frequency of responses consistent with System 1 processes. We present four experiments, across a range of domains, that are consistent with this hypothesis.

Experiment 1—Intuitive Defaults

Experiment 1 was designed to provide initial evidence that people adopt a systematic approach to reasoning when they experience cognitive disfluency. Participants completed the Cognitive Reflection Test (CRT; Frederick, 2005). This test consists of three items for which the gut reaction, or intuitive default, is incorrect but that respondents can correctly answer through deliberate reconsideration. A correct answer on each item suggests that the respondent engaged systematic processing to correct the intuitive response. Frederick (2005) administered the CRT to 3,400 participants across 35 studies and 11 samples and showed that scores on the CRT were highly correlated with a variety of measures associated with analytic thinking, including intelligence (Stanovich & West, 2000). If disfluency initiates systematic processing, then participants should perform better on the test if they experience disfluency while generating their answers. We manipulated disfluency in this experiment by printing the questions in either a difficult-to-read font (disfluent condition) or an easy-to-read font (fluent condition). We predicted that participants would answer more of the CRT items correctly when they were printed in a difficult-to-read font than when they were printed in an easy-to-read font.

Method

We recruited 40 Princeton University undergraduate volunteers at the student campus center to complete the three-item CRT (Frederick, 2005). Participants were seated either alone or in small groups, and the experimenter ensured that they completed the questionnaire individually. Those in the fluent condition completed a version of the CRT written in easy-to-read black Myriad Web 12-point font, whereas participants in the disfluent condition completed a version of the CRT printed in difficult-to-read 10% gray italicized Myriad Web 10-point font. Participants were randomly assigned to complete either the fluent or the disfluent version of the CRT. Previous research has shown that similar font manipulations effectively influence fluency (e.g., Oppenheimer, 2006; Werth & Strack, 2003). Consistent with these studies, a separate sample of 13 participants rated (on a 5-point scale) the disfluent font ($M = 3.08$, $SD = 0.76$) as being more difficult to read than the fluent font ($M = 1.54$, $SD = 0.87$), $t(12) = 3.55$, $p < .01$, $\eta^2 = .51$.

Results and Discussion

As predicted, participants answered more items on the CRT correctly in the disfluent font condition ($M = 2.45$, $SD = 0.64$) than in the fluent font condition ($M = 1.90$, $SD = 0.89$), $t(38) = 2.25$, $p = .03$, $\eta^2 = .12$. Whereas 90% of participants in the fluent condition answered at least one question incorrectly, only 35% did so in the disfluent condition, $\chi^2(1, N = 40) = 12.91$, $p < .001$, Cramer's $V = .57$. Finally, participants in the fluent condition provided the incorrect and intuitive response more often (23% of responses) than did participants in the disfluent condition (10% of responses), $Z = 1.96$, $p = .05$, $\eta^2 = .07$. When the CRT was difficult to read, participants appeared to engage in systematic processing and overcame their invalid intuitions to answer more questions correctly. These results provide preliminary evidence

that disfluency initiates systematic processing. We sought converging evidence for this hypothesis in the subsequent experiments and also addressed a variety of alternative interpretations for the results of Experiment 1.

The most obvious alternative interpretation of Experiment 1 is that presenting the test items in a disfluent font simply slowed participants down, thereby forcing them to process the information more carefully. This exogenous, task-specific mechanism is somewhat less interesting than our proposed endogenous mechanism, so we attempted in the remaining experiments to rule out the possibility that these effects merely reflected task-imposed constraints on processing speed. In Experiment 2, we adopted a fluency manipulation that did not force participants to process the information slowly and sought evidence for our proposed mechanism that people interpret disfluency as a signal to exert extra cognitive effort to adequately complete a task. In addition, although the CRT is a domain-general test of systematic reasoning, we wanted to ensure that the effects of disfluency on processing depth also applied to other concrete domains of judgment. Accordingly, we examined whether disfluency would lead people to focus on systematic processing cues when reading a review of a new MP3 player.

Experiment 2—Persuasion

The dominant models of persuasion, the heuristic–systematic model (Chaiken, 1980) and the elaboration likelihood model (Petty & Cacioppo, 1986), propose two distinct types of processing cues. Systematic or central cues generally involve the evidentiary quality of an argument, whereas heuristic or peripheral cues generally involve contextual factors irrelevant to an argument's quality (e.g., a target's appearance of competence). In Experiment 2, we tested whether experienced disfluency would increase people's reliance on systematic processing cues when evaluating a persuasive communication.

Participants read a fabricated review of a new MP3 player accompanied by a picture of either a competent-looking person discussing unimportant features (positive heuristic–negative systematic condition) or an incompetent-looking person discussing important features (negative heuristic–positive systematic condition). We manipulated processing fluency by presenting the masthead of the review in either difficult-to-read (disfluent condition) or easy-to-read (fluent condition) type. We predicted that, consistent with the results of Experiment 1, participants in the disfluent condition would rely more heavily on the systematic cue than on the heuristic cue.

Method

Pilot Experiments

Stimulus selection: Heuristic cues. A separate sample of 42 participants (Willis & Todorov, 2006) judged the apparent competence of a series of faces taken from a database of faces (Lundqvist, Flykt, & Öhman, 1998). The faces judged, on average, to be the most competent and the most incompetent from the database served as the targets of our heuristic-cue manipulation. Given that the physical appearance of competence is considered a heuristic cue, the competent-looking face constituted a strong (convincing) heuristic cue, and the incompetent-looking face constituted a weak (unconvincing) heuristic cue.

Stimulus selection: Systematic cues. In partial fulfillment of a course requirement, 10 Princeton University undergraduates reported the three most important and the three least important features of an MP3 player. We manipulated the strength of the systematic cue by using either the three most commonly mentioned important features (strong systematic cue: price, storage capacity, and battery life) or the three most commonly mentioned unimportant features (weak systematic cue: variety of colors, popular with celebrities, used by "everyone"). The resulting reviews are depicted in Figure 1.

Assessed difficulty: Underlying mechanism. Twenty Princeton University undergraduate volunteers at the campus student center completed a questionnaire designed to investigate whether our manipulation of fluency did indeed serve as a cue that greater cognitive effort would or would not be needed in the task. The
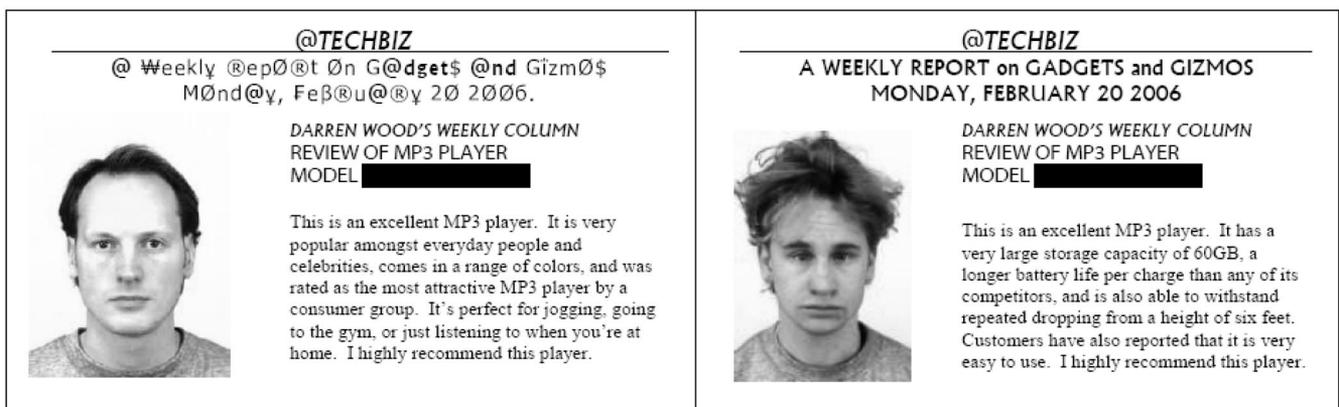

Figure 1. Stimuli from two of the four conditions in Experiment 1. The left panel shows the disfluent masthead and strong heuristic cue (competent-looking face) condition, and the right panel shows the fluent masthead and strong systematic cue (review listing important features) condition. The masthead fluency was switched in the other two conditions, which were otherwise identical.

questionnaire began with the fluent masthead for 10 participants and the disfluent masthead for the remaining 10 participants (see Figure 1 for mastheads). We asked participants to read the masthead and to imagine that they were going to read the remainder of the review. They rated how much effort they expected to have to expend to understand the contents of the review and the difficulty of reading the masthead (both on 5-point scales). As we expected, the fluency manipulation check showed that the disfluent masthead was considered more difficult to read ($M = 2.30$, $SD = 0.67$) than the fluent masthead ($M = 1.20$, $SD = 0.42$), $t(18) = 4.37$, $p < .001$, $\eta^2 = .52$. Perhaps more important, participants who read the disfluent masthead expected to need to expend more cognitive effort to understand the remaining contents of the review ($M = 2.30$, $SD = 0.67$) than did those who read the fluent masthead ($M = 1.40$, $SD = 0.70$), $t(18) = 2.93$, $p < .01$, $\eta^2 = .32$. This preliminary result suggests that people rely on the ease with which they process initial information to determine how much effort they will require to process subsequent information.

Main Experiment

Having shown that participants interpret disfluency as a cue to engage greater cognitive resources, we investigated whether disfluency also increases participants' reliance on systematic cues. Forty Princeton University volunteers at the student campus center read a short review of a new MP3 player, ostensibly printed from a Web site called Techbiz.com. Participants were seated alone and in small groups, and the experimenter ensured that participants in groups completed the questionnaire individually. The masthead on the review was written either in easy-to-read typeset (fluent condition) or in a difficult-to-read combination of letter-resembling symbols (disfluent condition; see Figure 1). Participants in the positive heuristic/negative systematic condition saw the highly competent-looking face used for the reviewer's photo paired with a review praising unimportant features of the MP3 player. Participants in the negative heuristic/positive systematic condition saw the opposite: the incompetent-looking face paired with a review praising important features of the MP3 player. In both conditions, the overall review of the MP3 player was positive. We randomly assigned participants to read one of the four versions of the review.

Note that, contrary to the fluency manipulation in Experiment 1, this fluency manipulation did not alter the information on which participants based their judgments. This manipulation instead altered the fluency of the masthead at the top of the review rather than the fluency of the information in the review itself. The actual content and format of the reviews were identical between conditions. Participants in the disfluent condition were therefore not compelled to read the review itself more slowly. Although it is possible that participants in the disfluent condition read the masthead more slowly and then maintained this slower processing speed through the rest of the materials, we suspect that the thousands of hours of practice our participants spent reading text in normal fonts outside our experimental context render this particular alternative quite unlikely. Nonetheless, this manipulation improves on many fluency manipulations as the actual content and format of the reviews were identical between conditions, and participants in the disfluent condition were therefore not required to read the review itself more slowly.

After reading the review, participants rated the competence of the reviewer, rated the quality of the MP3 player, and estimated how much they would like the experience of owning the MP3 player on separate 7-point scales. Scores on the three scales were highly correlated (Cronbach's $\alpha = .82$), so we averaged scores to create a composite *favorability* rating.

Results and Discussion

As predicted, participants' favorability ratings were more heavily influenced by the systematic cue (the quality of the arguments) in the disfluent condition than by the systematic cue in the fluent condition. Specifically, participants in the disfluent condition preferred the MP3 player when the systematic cue was persuasive ($M_{systematic} = 4.50$, $SD = 0.82$, vs. $M_{heuristic} = 3.47$, $SD = 1.62$), whereas those in the fluent condition preferred the MP3 player when the heuristic cue was persuasive ($M_{heuristic} = 3.63$, $SD = 0.72$, vs. $M_{systematic} = 3.00$, $SD = 1.14$), $F_{interaction}(1, 39) = 5.44$, $p < .03$, $\eta^2 = .13$ (see Table 1 for results of each component of the composite favorability rating).[2] Neither follow-up simple effect comparison reached significance, although participants' ratings were marginally more favorable toward the strong systematic cue review than the strong heuristic cue review when the masthead was disfluent, $F(1, 19) = 3.24$, $p < .10$, $\eta^2 = .15$. These results suggest that disfluency induces a more systematic processing style. It is important to note that this study used a manipulation that did not require participants to spend more time reading information in the disfluent condition than in the fluent condition. Furthermore, the pilot study shows that people expect to need additional cognitive resources to process information after the experience of disfluency. This strongly suggests that the apparent need for more systematic processing activated by the disfluent cues was at least a contributing if not the primary mechanism that led to System 2 processing.

Nonetheless, it is still possible that participants in Experiment 2 who read the review with the disfluent masthead also processed subsequent information more slowly. We therefore adopted a fluency manipulation in Experiment 3 that did not alter the properties of the stimuli at all, thereby eliminating the possibility that participants processed information more slowly because of changes in processing speed that originated in the stimuli. We also conducted Experiment 3 in a third domain of judgment to further demonstrate that the effects of disfluency on processing depth

---

[2] On closer inspection, the interaction in this experiment was driven by the strong systematic cue condition, in which participants gave significantly higher favorability ratings when the masthead was disfluent than when it was fluent, $F(1, 19) = 18.89$, $p < .001$, $\eta^2 = .51$. In contrast, the masthead had little effect on favorability ratings in the strong heuristic cue condition, $F < 1$. One possible explanation for this effect is that in the strong systematic cue condition, the disheveled appearance of the reviewer acted as a negative cue that directly contrasted with the compelling content of the actual review. Thus, when participants paid relatively more attention to the content of the review, their ratings were commensurately more favorable. In contrast, although the review in the strong heuristic condition was not particularly compelling, it still praised the MP3 player. As such, in the strong heuristic condition, the two cues did not run in opposition to one another—they were both positive. Thus, in the strong heuristic condition, participants' favorability ratings might not have been swayed strongly by whether they attended to the reviewer or the content of his review.

Table 1
Experiment 1: Mean Ratings of Predicted Liking of MP3 Player, Estimated Quality of MP3 Player, and Reviewer Competence

Rating dimension and masthead fluency	Strong heuristic cue		Strong systematic cue	
	M	*SD*	*M*	*SD*
Liking				
Fluent	3.90	1.37	3.50	0.97
Disfluent	4.10	2.08	5.00	1.25
Quality				
Fluent	4.30	1.42	3.40	0.84
Disfluent	3.60	1.90	4.40	1.17
Competence				
Fluent	2.70	1.16	2.10	0.88
Disfluent	2.70	1.16	4.10	1.37

Note. Ratings were made on separate 7-point scales. Although the data adhere to similar patterns across the three measures, the interaction is significant on the competence dimension, $F(1, 39) = 7.50$, $p = .01$, $\eta^2 = .17$; marginally significant on the quality dimension, $F(1, 39) = 3.75$, $p = .06$, $\eta^2 = .09$; but nonsignificant on the liking dimension, $F(1, 39) = 1.94$, $p = .17$, $\eta^2 = .05$.

occur across a broad array of domains, in this case, investigating reliance on simple heuristics in judgment.

Experiment 3—Representativeness Heuristic

In reviewing dual-process accounts in judgment and decision making, Kahneman and Frederick (2002) discussed the representativeness heuristic (Kahneman & Tversky, 1973) as a prototypical case of System 1 reasoning. Using this heuristic when making judgments of category inclusion entails categorizing exemplars primarily on the basis of the extent to which they are representative of a given category, ignoring more cognitively demanding cues such as base rates. Because representativeness is so strongly associated with the heuristic branch of dual-process reasoning, in Experiment 3, we examined whether the experience of disfluency might reduce people's reliance on the representativeness heuristic. Extending the generality of the effect, we manipulated fluency by asking participants to adopt a facial expression consistent with the exertion of cognitive effort (furrowed brow) or a control expression that did not imply the exertion of effort (puffed cheeks; see Stepper & Strack, 1993). As in Experiment 2, this fluency manipulation did not necessarily slow participants down, which eliminated the possibility that participants in the disfluent condition were merely forced to process the stimuli more slowly.

Method

Pilot Study

As in Experiment 2, we first investigated whether our manipulation of fluency—in this case, one's facial expression—did indeed serve as a cue that greater cognitive effort would or would not be needed in the task. Because heuristic reasoning represents a default intuitive judgment that systematic reasoning would overcome, we investigated whether furrowing one's brow influenced confidence in one's judgment. Low confidence in judgment would serve as a cue that more cognitive processing would be necessary

to arrive at the correct answer, and we expected those in the disfluent condition (who were furrowing their brows) would be less confident in their judgments than those in the fluent condition (who were puffing their cheeks).

Twenty Harvard University undergraduates participated in a lab study in which they answered a series of trivia questions and indicated their confidence that each of their answers was correct (from 0% to 100% confident). The experimenter told participants that the study was designed to examine how people answer questions under distracting conditions, and participants learned that they were going to answer questions while making various potentially distracting body movements. Participants were randomly assigned to adopt one of two facial expressions. Half of the participants were instructed to puff out their cheeks, whereas the other half were instructed to furrow their brows (Stepper & Strack, 1993). The experimenter demonstrated each expression until the participant held it correctly. Furrowing one's brow is associated with difficult mental effort and concentration, whereas puffing one's cheeks is a neutral expression that is equally difficult to maintain but is not associated with mental effort (Tourangeau & Ellsworth, 1979).

As predicted, participants who puffed their cheeks were significantly more confident in their responses ($M = 65\%$, $SD = 14\%$) than were participants who furrowed their brows ($M = 52\%$, $SD = 12\%$), $t(18) = 2.07$, $p < .05$, $\eta^2 = .21$. It is important to note that this difference in confidence did not reflect any difference in the actual accuracy of participants' responses among those who puffed their cheeks ($M = 36\%$ correct, $SD = 13\%$) versus those who furrowed their brows ($M = 38\%$ correct, $SD = 10\%$), $t(18) = 0.37$, *ns*; for the interaction between confidence and accuracy, $F(1, 18) = 4.85$, $p < .05$, $\eta^2 = .22$. These findings suggest that people are less confident in their judgments when they adopt facial expressions commonly associated with cognitive effort.

Main Experiment

One hundred fifty Harvard University undergraduates participated in a lab study in exchange for course credit. The materials for this experiment were taken from one of the original demonstrations of the representativeness heuristic: the Tom W. scenario (Kahneman & Tversky, 1973). Participants in this original demonstration read a description of a fictitious person—Tom W.—who was described in a way intended to seem similar to the widely recognized stereotype of an engineer.

As in the original experiment, one group of participants ($n = 51$) read a description of Tom W. and estimated on an 11-point scale how similar he was to a typical student with one of nine undergraduate majors (library science, social science and social work, business administration, computer science, humanities and education, law, medicine, engineering, and physical and life sciences). The mean similarity rating for each major functioned as a measure of how representative Tom W. was of a typical student with each major.

A second group of participants ($n = 55$) did not read a description of Tom W. but instead estimated the percentage of students on campus who were studying each of the nine majors. The mean proportion represented an estimate of the base rates associated with each of the nine majors.

Finally, a third group of students ($n = 44$) ranked the likelihood (from 1 = *most likely* to 9 = *least likely*) that Tom W. actually studied each of the nine majors. As in the pilot test, we randomly allocated half of these participants to puff their cheeks (fluent condition) and the other half to furrow their brows (disfluent condition) while rendering their judgments.

Results and Discussion

We correlated each participant's nine likelihood ratings with the mean representativeness and base-rate evaluations from the first two groups of participants. These two correlations provide an estimate of how strongly each participant in the third group relied on representativeness and base-rate information when assessing the likelihood that Tom W. studied each of the nine majors.

We began by reverse scoring participants' rankings so that the most likely major was scored a 9 and the least likely was scored a 1. A positive correlation between this recoded ranking variable and representativeness scores therefore indicates that the participant relied heavily on the representativeness heuristic, whereas a negative correlation indicates that the participant strongly neglected base rates.

As predicted, participants who furrowed their brows relied less on the representativeness heuristic (mean $r = .43$, $SD = 0.46$) than did participants who puffed their cheeks (mean $r = .74$, $SD = 0.19$), $t(42) = 2.84$, $p < .01$, $\eta^2 = .16$. Similarly, participants who furrowed their brows exhibited less base-rate neglect (mean $r = -.12$, $SD = 0.30$) than did participants who puffed their cheeks (mean $r = -.36$, $SD = 0.33$), $t(42) = 2.60$, $p < .01$, $\eta^2 = .14$. Thus, participants whose facial expressions induced a feeling of disfluency were less likely to rely on System 1 processes and were more inclined to use System 2. As the pilot study suggests, participants who furrowed their brows felt less confident in their judgments. These results suggest that participants dealt with the experience of lowered confidence by adopting a more careful, systematic approach to the task.

Although Experiments 1–3 demonstrate that experienced disfluency leads to more systematic processing, none of the experiments address the potential influence of mood. People in negative mood states—especially sadness—tend to process information more systematically (Schwarz, Bless, & Bohner, 1991), and it is at least possible that experienced disfluency worsens people's transient mood states. We therefore designed Experiment 4 to extend Experiments 1–3 into a new domain of reasoning while simultaneously investigating the possible role of mood in producing the observed results.

Experiment 4—Syllogistic Reasoning

Participants in Experiment 4 attempted to deduce logical conclusions from a series of two-statement syllogisms printed in either easy-to-read or difficult-to-read font. Syllogistic reasoning is one of the most widely studied processes in cognitive psychology (e.g., Johnson-Laird & Bara, 1984; Rips, 1994) and is often used as a case study in higher order reasoning for dual-process models of cognition (e.g., Evans & Over 1996; Sloman, 1996). We expected people to answer more questions correctly when the syllogisms were written in difficult-to-read font, which would be consistent with the results of Experiments 1–3. We also sought to investigate whether these effects were driven by differential mood effects of the fluency conditions.

Method

Pilot Study

As in Experiments 2 and 3, we first investigated whether our manipulation of fluency—in this case, the fonts of the items—could serve as a cue that greater cognitive effort would or would not be needed in the task. Accordingly, 69 Princeton University undergraduate volunteers at the student campus center read two syllogism questions printed in either an easy-to-read (fluent) font or a difficult-to-read (disfluent) font. The fonts were identical to those used in Experiment 1. Without actually answering the questions, participants rated how confident they were that they would be able to answer them correctly and estimated their difficulty (each on a 5-point scale). As we expected, participants who read the syllogisms printed in fluent font were more confident that they would be able to answer the questions correctly ($M = 4.86$, $SD = 1.39$, vs. $M = 4.00$, $SD = 1.39$), $t(67) = 3.03$, $p < .01$, $\eta^2 = .12$, and believed the task was less difficult ($M = 2.54$, $SD = 1.17$, vs. $M = 3.29$, $SD = 1.27$) than did participants who read the syllogisms printed in disfluent font, $t(67) = 2.58$, $p < .05$, $\eta^2 = .09$. Thus, the experience of reading the syllogisms printed in disfluent font lowered participants' confidence in their ability to answer the questions correctly and led them to believe that the questions were more difficult. Notice that in contrast to Study 3, these measures of confidence and difficulty were taken before answering the questions rather than after, thereby providing convergent evidence for our proposed mechanism.

Main Study

Forty-one Princeton University undergraduates at the student campus center volunteered to complete a questionnaire that contained six syllogistic reasoning problems. The experimenter approached participants individually or in small groups but ensured that they completed the questionnaire without the help of other participants. The syllogisms were selected on the basis of accuracy base rates established in prior research (Johnson-Laird & Bara, 1984; Zielinski, Goodwin, & Halford, 2006). Two were easy (answered correctly by 85% of respondents), two were moderately difficult (50% correct response rate), and two were very difficult (20% correct response rate). The easy and very difficult items were omitted from further analyses because the ceiling and floor effects obscured the effects of fluency on processing depth. Shallow heuristic processing enabled participants to answer the easy items correctly, whereas systematic reasoning was insufficient to guarantee accuracy on the difficult questions. Participants were randomly assigned to read the questionnaire printed in either an easy-to-read (fluent) or a difficult-to-read (disfluent) font, the same fonts that were used in Experiment 1.

Finally, participants indicated how happy or sad they felt on a 7-point scale (1 = *very sad*; 4 = *neither happy nor sad*; 7 = *very happy*). This is a standard method for measuring transient mood states (e.g., Forgas, 1995).

Results and Discussion

As expected, participants in the disfluent condition answered a greater proportion of the questions correctly ($M = 64\%$) than did

participants in the fluent condition ($M = 43\%$), $t(39) = 2.01$, $p < .05$, $\eta^2 = .09$. This fluency manipulation had no impact on participants' reported mood state ($M_{\text{fluent}} = 4.50$ vs. $M_{\text{disfluent}} = 4.29$), $t < 1$, $\eta^2 < .01$; mood was not correlated with performance, $r(39) = .18$, $p = .25$; and including participants' mood as a covariate did not diminish the impact of fluency on performance, $t(39) = 2.15$, $p < .05$, $\eta^2 = .11$. The performance boost associated with disfluent processing is therefore unlikely to be explained by differences in incidental mood states.

General Discussion

Dual-process models of human judgment are fundamental to much research in cognitive psychology, social psychology, and judgment and decision making. People make different decisions depending on whether they adopt systematic processing or rely on intuitive, heuristic processing. Understanding the output of human judgment and decision making therefore hinges on the ability to predict when people will activate these reasoning systems.

This research provides evidence that experienced difficulty or disfluency is one cue that leads people to adopt a systematic approach to information processing. This effect emerged with three manipulations of fluency across four domains of reasoning (a domain-general test of systematic reasoning, persuasion, person perception, and higher order reasoning). Our results suggest that participants in each experiment who experienced difficulty or disfluency while reasoning believed the tasks were more difficult and therefore engaged in more analytical processing than did those who did not. Pilot tests conducted for Experiment 2, 3, and 4 confirmed that our manipulations of fluency increased the apparent difficulty of the task and decreased participants' confidence in their accuracy of judgment. Although people are usually content to rely on heuristic processing, experienced difficulty or disfluency appears to act as a cue that the problem in question may require more elaborate thought and one's simple or intuitive response is likely to be wrong. These results are consistent with research demonstrating that people are less likely to choose a default option when their confidence is weakened (Simmons & Nelson, 2006).

In addition to describing the effect of fluency on processing depth, we also ruled out the alternative possibilities that stimulus-driven delays in processing speed (Experiments 2 and 3) and negative mood (Experiment 4) rather than disfluency per se induced systematic processing. Attending to stimuli more carefully and processing them more slowly are both hallmarks of System 2 processes, but our results were not created merely by constraints inherent in our stimuli that required more careful attention or slower processing. Instead, our results appear to have been produced because disfluency acts as a cue that more deliberate processing is required.

Existing fluency research has shown that people interpret stimuli depending on how easy those stimuli are to process (for a review, see Schwarz, 2004), but our research makes the distinct theoretical point that processing fluency can also influence judgment indirectly by serving as a cue to engage in deeper reasoning. Until now, fluency researchers have not provided participants with cues at a shallow, heuristic level that systematically contradict cues at a deeper, systematic level (e.g., an MP3 player reviewer who looks incompetent but writes a compelling review and vice versa). As a result, it has not been possible to determine whether fluency affects how deeply people process available information. Our findings are therefore novel because they show that fluency is used not only directly as a cue for judgment but also indirectly as a mechanism for strategy selection.

Moreover, these findings suggest a resolution for inconsistencies in the fluency literature. For instance, some research suggests that fluent stimuli are perceived as being more familiar than disfluent stimuli (Monin, 2003), whereas other research suggests that fluent stimuli are judged as being less familiar than disfluent stimuli (Guttentag & Dunn, 2003). In addition, although disfluency implies low quality at a superficial level (e.g., Reber, Winkielman, & Schwarz, 1998), it also activates System 2 processing that might lead people to look beyond a target's superficial characteristics. Fluency might exert opposing influences when people use cognitive difficulty as a direct cue (e.g., "I don't like the target") or as an indirect cue (e.g., "I should think more carefully about my judgment"). Thus, the direct effects of fluency on evaluation and the indirect effects of fluency on subsequent processing might, in certain contexts, have opposing influences on judgment.

More generally, these findings bring us closer to determining what activates System 2 processing. To be viable accounts of human reasoning, dual-system theories need to explain not only what the two systems are but also what activates the use of each system. This research suggests that processing fluency may be one important factor that determines when people will overcome their intuitive responses to engage more systematic reasoning.

References

Ajzen, I., & Sexton, J. (1999). Depth of processing, belief congruence, and attitude–behavior correspondence. In S. Chaiken & Y. Trope (Eds.), *Dual-process theories in social psychology* (pp. 117–138). New York: Guilford Press.

Alter, A. L., & Oppenheimer, D. M. (2006). Predicting short-term stock fluctuations using processing fluency. *Proceedings of the National Academy of Sciences, USA, 103,* 9369–9372.

Bless, H., & Schwarz, N. (1999). Sufficient and necessary conditions in dual-process models: The case of mood and information processing. In S. Chaiken & Y. Trope (Eds.), *Dual-process theories in social psychology* (pp. 423–440). New York: Guilford Press.

Bodenhausen, G. V., Macrae, C. N., & Sherman, J. W. (1999). On the dialectics of discrimination: Dual processes in social stereotyping. In S. Chaiken & Y. Trope (Eds.), *Dual-process theories in social psychology* (pp. 271–290). New York: Guilford Press.

Boksman, K., Théberge, J., Williamson, P., Drost, D. J., Malla, A., Densmore, M., et al. (2005). A 4.0-T fMRI study of brain connectivity during word fluency in first-episode schizophrenia. *Schizophrenia Research, 75,* 247–263.

Botvinick, M. M., Braver, T. S., Carter, C. S., Barch, D. M., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review, 108,* 624–652.

Chaiken, S. (1980). Heuristic versus systematic information processing and the use of source versus message cues in persuasion. *Journal of Personality and Social Psychology, 39,* 752–766.

Epley, N., Keysar, B., Van Boven, L., & Gilovich, T. (2004). Perspective taking as egocentric anchoring and adjustment. *Journal of Personality and Social Psychology, 87,* 327–339.

Evans, J. S. B. T. (2003). In two minds: Dual-process accounts of reasoning. *Trends in Cognitive Sciences, 7,* 454–459.

Evans, J. S. B. T., & Over, D. E. (1996). *Rationality and reasoning.* Hove, UK: Psychology Press.

Forgas, J. P. (1995). Mood and judgment: The affect infusion model (AIM). *Psychological Bulletin, 117,* 39–66.

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

Gilbert, D. T. (1989). Thinking lightly about others: Automatic components of the social inference process. In J. S. Uleman & J. A. Bargh (Eds.), *Unintended thought* (pp. 189–211). New York: Guilford Press.

Gill, M. J., Swann, W. B., Jr., & Silvera, D. H. (1998). On the genesis of confidence. *Journal of Personality and Social Psychology, 75,* 1101–1114.

Goel, V., Buchel, C., Frith, C., & Dolan, R. J. (2000). Dissociation of mechanisms underlying syllogistic reasoning. *Neuroimage, 12,* 504–518.

Griffin, D., & Tversky, A. (1992). The weighing of evidence and the determinants of confidence. *Cognitive Psychology, 24,* 411–435.

Guttentag, R., & Dunn, J. (2003). Judgments of remembering: The revelation effect in children and adults. *Journal of Experimental Child Psychology, 86,* 153–167.

Jacoby, L. L., Kelley, C. M., & McElree, B. D. (1999). The role of cognitive control: Early selection vs. late correction. In S. Chaiken & Y. Trope (Eds.), *Dual-process theories in social psychology* (pp. 383–400). New York: Guilford Press.

James, W. (1950). *The principles of psychology.* New York: Dover. (Original work published 1890)

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

Jones, T. C., & Jacoby, L. L. (2001). Feature and conjunction errors in memory: Evidence for dual-process theory. *Journal of Memory and Language, 45,* 82–102.

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

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review, 80,* 237–251.

Kelley, C. M., & Lindsay, D. S. (1993). Remembering mistaken for knowing: Ease of retrieval as a basis for confidence in answers to general knowledge questions. *Journal of Memory and Language, 32,* 1–24.

Keysar, B., & Barr, D. J. (2002). Self anchoring in conversation: Why language users do not do what they "should." In T. Gilovich, D. W. Griffin, & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 150–166). Cambridge, UK: Cambridge University Press.

Kruglanski, A. W., & Thompson, E. P. (1999). Persuasion by a single route: A view from the unimodel. *Psychological Inquiry, 10,* 83–109.

Lieberman, M. D., Gaunt, R., Gilbert, D. T., & Trope, Y. (2002). Reflection and reflexion: A social cognitive neuroscience approach to attributional inference. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 34, pp. 199–249). New York: Academic Press.

Lundqvist, D., Flykt, A., & Öhman, A. (1998). *The Karolinska directed emotional faces.* Stockholm, Sweden: Karolinska Institute, Department of Clinical Neuroscience, Psychology Section.

McGlone, M. S., & Tofighbakhsh, J. (2000). Birds of a feather flock conjointly (?): Rhyme as reason in aphorisms. *Psychological Science, 11,* 424–428.

Monin, B. (2003). The warm glow heuristic: When liking leads to familiarity. *Journal of Personality and Social Psychology, 85,* 1035–1048.

Oppenheimer, D. M. (2006). Consequences of erudite vernacular utilized irrespective of necessity: Problems with using long words needlessly. *Applied Cognitive Psychology, 20,* 139–156.

Osman, M. (2004). An evaluation of dual-process theories of reasoning. *Psychonomic Bulletin and Review, 11,* 988–1010.

Petty, R. E., & Cacioppo, J. T. (1986). The elaboration likelihood model of persuasion. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 19, pp. 123–205). New York: Academic Press.

Reber, R., Winkielman, P., & Schwarz, N. (1998). Effects of perceptual fluency on affective judgments. *Psychological Science, 9,* 45–48.

Rips, L. J. (1994). *The psychology of proof: Deductive reasoning in human thinking.* Cambridge, MA: MIT Press.

Schwarz, N. (1998). Accessible content and accessibility experiences: The interplay of declarative and experiential information in judgment. *Personality and Social Psychology Review, 2,* 87–99.

Schwarz, N. (2004). Metacognitive experiences in consumer judgment and decision making. *Journal of Consumer Psychology, 14,* 332–348.

Schwarz, N., Bless, H., & Bohner, G. (1991). Mood and persuasion: Affective states influence the processing of persuasive communications. In M. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 24, pp. 161–197). New York: Academic Press.

Simmons, J. P., & Nelson, L. D. (2006). Intuitive confidence: Choosing between intuitive and nonintuitive alternatives. *Journal of Experimental Psychology: General, 135,* 409–428.

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

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

Stepper, S., & Strack, F. (1993). Proprioceptive determinants of emotional and nonemotional feelings. *Journal of Personality and Social Psychology, 64,* 211–220.

Tetlock, P. E., & Lerner, J. S. (1999). The social contingency model: Identifying empirical and normative boundary conditions for the error-and-bias portrait of human nature. In S. Chaiken & Y. Trope (Eds.), *Dual-process theories in social psychology* (pp. 571–585). New York: Guilford Press.

Tourangeau, R., & Ellsworth, P. (1979). The role of the face in the experience of emotion. *Journal of Personality and Social Psychology, 37,* 1519–1531.

Werth, L., & Strack, F. (2003). An inferential approach to the knew-it-all-along phenomenon. *Memory, 11,* 411–419.

Whittlesea, B. W. A., & Leboe, J. P. (2003). Two fluency heuristics (and how to tell them apart). *Journal of Memory and Language, 49,* 62–79.

Willis, J., & Todorov, A. (2006). First impressions: Making up your mind after a 100-ms exposure to a face. *Psychological Science, 17,* 592–598.

Zielinski, T., Goodwin, G. P., & Halford, G. S. (2006). *Complexity of categorical syllogisms: An integration of two metrics.* Manuscript submitted for publication.