Journal of Experimental Psychology: General

ISSN: 0096-3445

2020, Vol. 149, No. 12, 2395–2405
http://dx.doi.org/10.1037/xge0000779

Consistency Just Feels Right: Procedural Fluency Increases Confidence in Performance

Elanor F. Williams
Washington University in St. Louis

Kristen E. Duke
University of Toronto

David Dunning
University of Michigan

Incidental features of a stimulus can increase how easily it is processed, which can then increase confidence in task performance. Here, we examine the impact of fluency stemming from procedural features embedded in a task rather than in the features of a stimulus. We propose that manipulating the consistency of procedural features over a series of stimuli can produce *procedural fluency*, a metacognitive sense of ease in processing that can inflate confidence without boosting accuracy. That is, even superficial consistency within a task can lead people to inaccurately believe they are performing better. As with fluency derived from features of individual stimuli, drawing attention to procedural consistency leads people to discount it, attenuating its impact on confidence. Further, the influence of procedural fluency on confidence relies on individuals' naïve theories about what fluency signals about their performance. Accordingly, manipulating these naïve theories mitigates the effects of procedural fluency on confidence.

Keywords: metacognition, fluency, confidence, consistency

A raft of research has shown that fluency—how quick and effortless a task feels to complete—predicts how confident people are in their performance (e.g., Gill, Swann, & Silvera, 1998; Hertzog, Dunlosky, Robinson, & Kidder, 2003; Reder, 1987; Schwarz, 2004). Often, the inferences people draw from fluency are accurate extensions of learned experience, in that people who perform tasks more quickly and more effortlessly typically are indeed more accurate (e.g., Herzog & Hertwig, 2013). As a result, using fluency as a cue can be an effective strategy for estimating confidence. In fact, fluency can be a useful cue for making a wide variety of judgments, including liking, truthfulness, fame, and value (e.g., Alter & Oppenheimer, 2009; Herzog & Hertwig, 2013; Schwarz & Clore, 2007).

Fluency, however, is not always relevant to the judgments people make, nor diagnostic of the accuracy of those judgments. Individuals might be led astray by extraneous features of judgment tasks that produce a misleading sense of fluency. Researchers have investigated whether fluency can influence judgments even when it is irrelevant. The most conservative of such tests modify the individual stimuli to create superficial and meaningless fluency; for example, an experimenter might make printed materials easier or harder to read without changing their content and then test how this manipulation affects participants' judgments. This research shows, for instance, that people report that legibly printed statements seem more familiar (Reber & Zupanek, 2002) and truer (Reber & Schwarz, 1999) than identical statements written in harder-to-read fonts, and that people have stronger preferences regarding choice options printed in a clear font than in an unclear one (Novemsky, Dhar, Schwarz, & Simonson, 2007). These results arise even though the font in which text is printed is not informative of the familiarity, truth, appeal, or other qualities of its content.

In general, this research suggests that fluency alone can lead to confidence and a variety of other judgments while bearing no relationship to actual performance or experience (for a review, see Alter & Oppenheimer, 2009). As such, studying superficial fluency and its operation has become an important undertaking to understand how people make judgments and assess the quality of those judgments. In particular, irrelevant fluency can illuminate how people make inferences about their own performance.

Stimulus-Driven Fluency

However, questions have been raised recently about the reliability of certain fluency effects, particularly those that embed fluency in the features of a stimulus. Consider the aforementioned popular approach of manipulating fluency by printing stimulus materials in an easy- or difficult-to-read font. Past research suggests that the disfluency caused by superficially hard-to-read text not only can make participants less confident in their initial judgments about

This article was published Online First April 23, 2020.
Elanor F. Williams, Olin Business School, Washington University in St. Louis; Kristen E. Duke, Rotman School of Management, University of Toronto; David Dunning, Department of Psychology, University of Michigan.

Correspondence concerning this article should be addressed to Elanor F. Williams, Olin Business School, Washington University in St. Louis, Campus Box 1156, 1 Brookings Drive, St. Louis, MO 63130. E-mail: elanorfwilliams@wustl.edu

those stimuli, but also can prompt them to engage in deeper, more analytic processing of the materials (Alter, Oppenheimer, Epley, & Eyre, 2007) or to better retain the information in them (Diemand-Yauman, Oppenheimer, & Vaughan, 2011). These intriguing findings have recently come into question, though, with numerous researchers reporting that manipulations using hard-to-read fonts lead to lower confidence but *worse* performance (Miele & Molden, 2010; Miele, Son, & Metcalfe, 2013). Relatedly, others report null font fluency effects on learning and performance (e.g., Kühl & Eitel, 2016; Meyer et al., 2015; Rummer, Schweppe, & Schwede, 2016; Thompson et al., 2013) or even on confidence itself (Magreehan, Serra, Schwartz, & Narciss, 2016; Sungkhasettee, Friedman, & Castel, 2011). Others have proposed that additional criteria may be necessary for font disfluency to have an effect on confidence and improve learning. These criteria range from moderators, such as individual differences in implicit theories of intelligence, to additional task features, such as obvious contrast between easy and difficult fonts (Magreehan et al., 2016; Miele, Finn, & Molden, 2011; Oppenheimer & Alter, 2014; Pieger, Mengelkamp, & Bannert, 2017; Wänke & Hansen, 2015). Overall, the inconsistency in the effects produced by one of the most commonly used fluency manipulations indicates that more and different evidence is needed to shore up our understanding of the link between fluency and confidence.

Among the potential drawbacks of using a difficult-to-read font as an instantiation of disfluency is that it can be so out of the ordinary that participants may spontaneously realize that the disfluency it produces is irrelevant to the task. Thus, they know to discount the difficulty induced by the font and may not let it influence their confidence or other judgments (e.g., Oppenheimer, 2004). In comparison, some markers of fluency may be more subtle and thus harder to identify and compensate for. We suggest that certain fluency manipulations may be undetected or overlooked by respondents because they appear to be a central part of the task itself, rather than transparently overlaid on them in the way that font fluency is. Yet despite being more interwoven into the task itself, embedded fluency manipulations may remain irrelevant to the task and its content. Some past work does support this assertion. For example, researchers using scrambled sentence tasks can present the words in more or less correct grammatical order, with more grammatical primes seeming more fluent even though the words and sentence solutions themselves remain constant (Greifeneder & Bless, 2010). Or, researchers can ask participants to complete a judgment task repeatedly over a period of time; as individuals gain practice performing the task, they may mistakenly use the experience of increasing ease over time as a signal that leads them to give more positive assessments of the stimuli they are judging (O'Connor & Cheema, 2018). Beyond these, there are several ways that one could superficially embed fluency into a task's stimuli in order to influence confidence, specifically. For instance, people are more confident they would be able to find anagrams from scrambled sets of letters that are pronounceable than from sets that are not (Topolinski, Bakhtiari, & Erle, 2016). They are also more confident that dot pattern equations are correct when those equations are organized in a symmetric, orderly fashion than in an asymmetrical fashion (Reber, Brun, & Mitterndorfer, 2008). These findings suggest that manipulations changing the structural approach to stimuli to make them harder or easier to process can change fluency-related judgments of confidence, even though the nature of this influence remains arguably less salient or detectable.

Procedural Fluency

Here, we investigated a different and novel source of fluency, one which flows not from the individual stimuli being processed but rather from the procedure that participants follow. This fluency is still unrelated to objective task performance even though it is embedded in the broader procedure of completing a task. In so doing, we sought to provide new evidence supporting the link between fluency and confidence that is, importantly, independent of actual performance and accuracy. Specifically, we examined how making a task's procedure superficially more consistent across stimuli can affect fluency and judgments of performance.

To foreshadow the results, we found that mere consistency in a task's construction and execution—how similarly the task stimuli are presented over a set of trials—can produce feelings of fluency and confidence, a finding that we term *procedural fluency*. In our studies, participants completed a reasoning task that followed a consistent routine over multiple trials, making it procedurally fluent, or a routine that included minor deviations across trials, which rendered the task more disfluent. This task-level fluency is distinct from the ease of processing associated with individual stimuli (e.g., whether they are presented in a hard- or easy-to-read font or have a symmetrical or asymmetrical form). Importantly, this superficial change did not affect participants' actual performance. Despite the fact that consistency was not a meaningful marker of performance and did not improve it, we found that procedural consistency increased participants' confidence and, in particular, confidence that was decoupled from actual achievement.

This work departs from existing research in several ways. Instead of producing fluency by altering the superficial appearance of a task's stimuli and thus making each individual stimulus more or less fluent, we produced fluency by manipulating consistency in the overall procedure that participants followed. It is only in how consistently the items are presented over the course of the task as a whole that solving them induces fluency and confidence. Said differently, each stimulus is more or less fluent only in the context of the stimuli that preceded it; the procedural fluency manipulation would evaporate and have no effect on confidence if an item were extracted from the task and presented on its own. Thus, the central contribution of the current research is providing evidence bolstering the link between fluency and confidence by introducing a previously undocumented source of fluency derived from the procedural consistency of a task. This novel finding demonstrates that manipulating how consistently a procedure is presented can make people feel more confident in their performance without affecting the quality of that performance, and allows us to examine how fluency affects confidence while sidestepping issues potentially raised by inducing fluency at a stimulus level.

The idea that consistency and repetition can produce fluency likely sounds familiar. Mere exposure, one of the original ways that researchers tested how incidental fluency can affect judgments, is the finding that repeated exposure (even nonconscious exposure) to a stimulus makes that stimulus easier to process (e.g., Zajonc, 2001). In research on mere exposure, participants are repeatedly exposed to the same stimulus (which is almost always intermixed with other nonrepeated stimuli; e.g., Dechêne, Stahl, Hansen, & Wänke, 2009). However, unlike in the mere exposure paradigm, in the current research it is consistency (or the lack thereof) that is built into a task procedure encompassing a variety of stimuli, rather than repetition of an individual stimulus, that ultimately induces procedural fluency or disfluency in our tasks.

Our work contributes to fluency research in an additional way. Much recent theorizing about how fluency operates has focused on the importance of context and discrepancy—that is, how fluent or disfluent something seems relative to one's expectations or the surrounding stimuli (e.g., Hansen & Wänke, 2013; Wänke & Hansen, 2015; Whittlesea & Williams, 2000; Wilcox & Song, 2011). This work even goes so far as to suggest that intermixed trials of fluent and disfluent stimuli may be necessary for fluency effects to appear (Dechêne et al., 2009; Dechêne, Stahl, Hansen, & Wänke, 2010). Procedural fluency adds to a small but growing body of research that reiterates how stimuli can feel fluent not just in contrast to other stimuli or to one's expectations, but also in concert with each other—that is, the holistic experience of a task matters as much as the individual stimuli themselves (e.g., O'Connor & Cheema, 2018; Susser, Panitz, Buchin, & Mulligan, 2017).

Other research provides indirect evidence to support that experiencing procedural consistency might lead people to believe they are performing well. For example, participants believe they have counted stimuli more accurately when they count with rhythmic (vs. arrhythmic) timing (Stevenson & Carlson, 2020). In deeper reasoning tasks, people observing others think of them as more skilled when they approach tasks consistently rather than in an ad hoc manner (Falk & Zimmermann, 2017). Finally, people are more confident in their judgments when applying consistent algorithms to reach them versus solving them in an ad hoc manner, even when those algorithms are wrong (Williams, Dunning, & Kruger, 2013). Yet this prior work focuses on the strategies that individuals employ when attempting to solve tasks, revealing that using a consistent strategy inflates confidence. The current research instead explores how constructing a task with superficially consistent features across trials influences perceived performance.

The Present Research

In two experiments, we tested whether consistency in the procedure that one follows across a series of stimuli induces a feeling of procedural fluency, which has corresponding effects on participants' confidence in their own task performance. Specifically, we tested whether superficial consistency in the presentation and execution of a task, as when stimuli are presented in the same order or the same color, can make people more confident they are completing the task correctly.

In addition, we tested whether procedural consistency is subject to the same boundary conditions as other forms of fluency. The first boundary we tested is whether people are made aware of the source of fluency. When the source of fluency is brought to light, particularly when that source is not diagnostic of performance, people often avoid drawing inferences from the fluency of their experience. In short, people "discount" fluency if its origin is made explicit to them (e.g., Schwarz & Clore, 2007). We tested for this discounting effect in Experiment 1.

The way fluency affects judgments also depends on what people believe that fluency means about their performance. The lenses through which people interpret fluency have been called "naïve theories"—in other words, people's lay intuitions about what fluency signals (e.g., Schwarz, 2004; Thomas & Morwitz, 2009). When people intuitively believe that having an easier time completing a task is a sign that one has completed it correctly, this should boost their confidence. However, if people believe instead that the feeling of fluency may not necessarily be diagnostic of accuracy, it should influence their confidence less. Thus, in Experiment 2, we tested whether providing an alternative naïve theory alters the effect of procedural consistency on confidence.

Experiment 1

This experiment was an initial investigation of procedural fluency using a word search game. In each trial, participants encountered a set of letters and were asked to find words that could be formed by rearranging those letters. These words each had to begin with a target letter in the set. Importantly, this target letter either remained consistently in the same relative position in the set (i.e., the third out of six letters) across all trials, or varied in its position across trials. Although this manipulation should not have influenced participants' actual performance, we proposed that participants who experienced superficial consistency in the procedure of the task would experience procedural fluency and accordingly report greater confidence in their performance.

We also investigated the effects of drawing attention to this procedural consistency (or lack thereof) during task performance. Warning participants about the potential for (in)consistency to influence their experience completing the task should encourage them to attribute any fluency or disfluency to its true source rather than to their own performance. This should consequently reduce the effect of procedural consistency on their confidence.

Method

Participants. Four hundred ten undergraduates at a large West Coast university (38.8% female, M_{age} = 21.1 years) participated for partial course credit. We collected data for two full weeks of laboratory participants with a target sample size of at least 50 per cell, which would provide 80% power to detect a small-to-medium size effect. In both experiments, we report all measures, manipulations, and data exclusions; analyses were not performed until all data were collected. Some participants did not complete all follow-up questions, leading to varying degrees of freedom in the analyses below.

Procedure. Participants were randomly assigned to one of four conditions in a 2 (consistent vs. inconsistent) × 2 (warning vs. no warning) between-participants design. In the word game, participants encountered sets of six letters, with one letter highlighted in orange. Their task was to find all the five-letter words that could be formed by rearranging those six letters and that began with the highlighted letter, and they could only use each letter in the set once per word. They were informed that there were between one and seven such words for each set of letters. Following an example and a correct solution (see Figure 1), participants encountered five letter sets. They had 90 s to search for words for each set.

During the task, participants in the consistent conditions viewed letter sets in which the highlighted letter was always the third letter

Figure 1. Example letter set as presented to participants in Experiment 1. In this example set of letters, the five-letter words starting with *L* (the highlighted letter) are LABEL/LIBEL.

in the set, while participants in the inconsistent conditions saw the (same) highlighted letter as the second, sixth, fifth, first, and fourth letter, respectively. Additionally, participants in the warning conditions viewed the following message underlined and embedded in the task instructions just before they started the task: "Note: the highlighted letter in each letter set will be in the same [a different] location. This may affect how easy or difficult the task seems to be." This warning was simply omitted in the no warning condition.

Measures. Participants completed measures regarding their confidence about each trial and about the task overall. After completing each individual set of letters (i.e., each trial), participants indicated how confident they were that they had found all available five-letter words in the set, from 1 (0%: *not at all confident*) to 11 (100%: *completely confident*), reported below as percentages. We averaged these ratings on each set of letters to form a single measure, "item-level confidence," in the analyses below. After completing all five sets, participants indicated how confident they felt overall in their ability to find the words (what we call "overall confidence" below), from 1 (0%: *not at all confident*) to 11 (100%: *completely confident*), reported below as percentages. In addition, participants reported how well they thought they had performed compared to fellow students of their age and gender ("relative performance beliefs"), from 0 (*much worse than most*) to 10 (*much better than most*), and how easy or difficult the task was ("perceived ease"), from 1 (*extremely difficult*) to 7 (*extremely easy*). Finally, they indicated the extent to which the location of the highlighted letters had affected their ability to find words ("awareness of procedural fluency"), from 1 (*not at all*) to 5 (*extremely*), and how often they generally play word games, from 1 (*never*) to 7 (*daily*).

Results

Performance. The consistency manipulation did not affect participants' performance (i.e., the total number of words participants found), and, if anything, inconsistent participants actually found slightly but not significantly more words ($M_{\text{consistent}} = 6.13$, $SD = 3.27$ vs. $M_{\text{inconsistent}} = 6.66$, $SD = 3.37$), $F(1, 408) = 2.63$, $p = .11$, $\eta_p^2 = .006$, 90% confidence interval (CI) of the effect size [.00, .03].[1] In addition, the warning manipulation also did not affect participants' performance ($M_{\text{warning}} = 6.28$, $SD = 3.47$ vs. $M_{\text{no warning}} = 6.51$, $SD = 3.20$), $F(1, 408) = .52$, $p = .47$, $\eta_p^2 = .001$, CI [.00, .01], and, further, did not interact with the consistency manipulation to influence performance, $F(1, 406) = .83$, $p = .36$, $\eta_p^2 = .002$, CI [.00, .02]. As such, any differences in confidence were unlikely to be due to actual differences in performance.

Item-level confidence. Unsurprisingly, participants who correctly solved more anagrams and thus performed better were more confident in their performance, $\beta = 1.13$, $SE = 0.35$, $t(385) = 3.25$, $p = .001$; to control for this relationship, in this and subsequent analyses, objective performance on the task (i.e., the total number of words correctly identified across all trials) was included as a covariate. As predicted, procedural consistency increased item-level confidence, as long as attention was not drawn to it. An analysis of covariance (ANCOVA) with performance as a covariate revealed no main effect of consistency, $F(1, 382) = .78$, $p = .38$, $\eta_p^2 = .002$, 90% CI [.00, .02], nor of warning, $F(1, 382) = .04$, $p = .84$, $\eta_p^2 < .001$, CI [.00, .006]. However, there was a significant interaction between these two factors, $F(1, 382) = 4.26$, $p = .04$, $\eta_p^2 = .01$, CI [.0003, .04].

As Figure 2 shows, within the unwarned conditions, consistent condition participants indeed reported significantly higher confidence (adj. $M = 33.21$, $SE = 2.32$) than did inconsistent condition participants (adj. $M = 26.46$, $SE = 2.30$), $F(1, 382) = 4.25$, $p = .04$, $\eta_p^2 = .01$, 90% CI [.0003, .03]. However, when participants' attention was drawn to the source of the fluency (i.e., within the warning conditions), this difference was attenuated (adj. $M_{\text{consistent}} = 28.95$, $SE = 2.30$ vs. adj. $M_{\text{inconsistent}} = 31.64$, $SE = 2.24$), $F(1, 382) = .71$, $p = .40$, $\eta_p^2 = .002$, CI [.00, .02]. An analysis of variance (ANOVA) without controlling for accuracy revealed the same pattern, although slightly weaker, as would be expected. Thus, participants were more confident that they had found all available words in each set when they experienced superficial consistency in the task's presentation; however, drawing participants' attention to this procedural consistency attenuated this effect.

Item-level overconfidence. We can also account more directly for participants' performance on each set by comparing their confidence on each anagram set to their actual accuracy on that set. In particular, we can compare their reported percentage confidence that they had found all the five-letter words in a given anagram set to the percentage of such words they had actually found. For example, if a participant indicated that they were 100% confident (a confidence score of 1.00) that they found all of the anagrams, but actually found only three of the four words in a set (an accuracy score of 0.75), they would receive an overconfidence score of 0.25 for that set. Likewise, had this same participant reported only 50% confidence that they found all the anagrams, they would receive an overconfidence score of −.25 for that set. These overconfidence scores were averaged across the five sets to form a single measure of overconfidence. An ANOVA revealed a marginal main effect of consistency, $F(1, 383) = 2.76$, $p = .10$, $\eta_p^2 = .007$, 90% CI [.00, .03], and no main effect of warning, $F(1, 383) = 0.52$, $p = .47$, $\eta_p^2 = .001$, CI [.00, .01], qualified by a significant interaction between the two factors, $F(1, 383) = 4.75$, $p = .03$, $\eta_p^2 = .01$, CI [.001, .04].

Within the unwarned conditions, consistent participants were somewhat more confident than their performances warranted ($M = .04$, $SE = .03$) and were confident to a greater degree than inconsistent participants were, who were underconfident ($M = -.06$, $SE = .03$), $F(1, 383) = 7.25$, $p = .007$, $\eta_p^2 = .02$, 90% CI [.003, .05]. However, this difference was attenuated within the warned conditions ($M_{\text{consistent}} = .003$, $SE = .03$ vs. $M_{\text{inconsistent}} = .02$, $SE = .03$), $F(1, 383) = .14$, $p = .71$, $\eta_p^2 < .001$, CI [.00, .01].

[1] Ninety-percent confidence intervals are appropriate here as F-tests are one-sided tests, and η_p^2 must be positive because it is squared. Ninety-five-percent confidence intervals can thus include zero even when the test is significant, so we report 90% confidence intervals, as is typical.

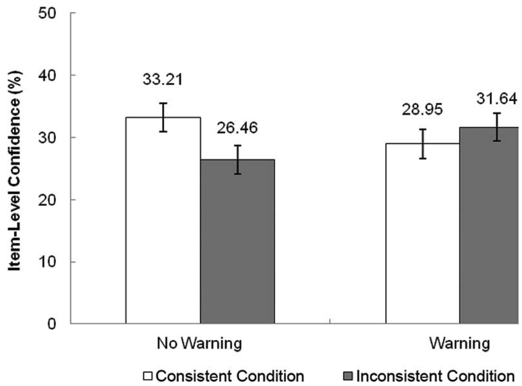

Figure 2. Item-level confidence ratings, Experiment 1. Estimated marginal means for item-level confidence ratings across items within the consistent and inconsistent conditions, with and without warning. Error bars represent standard errors of the means.

Thus, superficial consistency or inconsistency in the task's presentation affected the calibration of participants' assessments of their performances relative to their actual accuracy. However, drawing participants' attention to the consistency or inconsistency in the task reduced this discrepancy.[2]

Overall confidence. The single-item overall confidence measure showed a similar pattern to item-level confidence. Again, participants who correctly solved more anagrams were more confident in their overall performance, $\beta = 1.57$, $SE = .35$, $t(392) = 4.49$, $p < .001$. More importantly, an ANCOVA with performance (i.e., total number of anagrams solved correctly) as a covariate revealed that there was no main effect of consistency, $F(1, 389) = .03$, $p = .87$, $\eta_p^2 < .001$, 90% CI [.00, .005], nor of warning, $F(1, 389) = .12$, $p = .73$, $\eta_p^2 < .001$, CI [.00, .009]. However, as with averaged confidence, there was a significant interaction between consistency and warning, $F(1, 389) = 5.26$, $p = .02$, $\eta_p^2 = .01$, CI [.001, .04].

As shown in Figure 3, when they were not warned about the task, consistent participants were marginally more confident overall (adj. $M = 33.53$, $SE = 2.35$) than were inconsistent participants (adj. $M = 27.82$, $SE = 2.34$), $F(1, 389) = 2.96$, $p = .09$, $\eta_p^2 = .008$, 90% CI [.00, .03]. However, when participants' attention was drawn to the source of fluency, this difference was slightly but not significantly reversed (adj. $M_{\text{consistent}} = 27.42$, $SE = 2.34$ vs. adj. $M_{\text{inconsistent}} = 32.34$, $SE = 2.25$), $F(1, 382) = 2.30$, $p = .13$, $\eta_p^2 = .006$, CI [.00, .03]. As with averaged confidence, an ANOVA without controlling for actual performance revealed a similar but weaker pattern.

Relative performance beliefs. We also asked participants to compare their perceived performance to that of their peers. Consistency did not seem to affect these beliefs: participants' perceptions of their performance relative to their peers reflected a pattern similar to their confidence ratings, but it did not reach significance. When not warned, consistent participants thought they performed nonsignificantly better compared to their peers (adj. $M = 4.53$, $SE = .22$) than did inconsistent participants (adj. $M = 4.29$, $SE = .22$), $F(1, 388) = 0.59$, $p = .44$, $\eta_p^2 = .002$, 90% CI [.00, .01], but this pattern was slightly reversed when participants were warned (adj. $M_{\text{consistent}} = 4.64$, $SE = .22$ vs. adj. $M_{\text{inconsistent}} = 4.92$, $SE = .21$), $F(1, 388) = .81$, $p = .37$, $\eta_p^2 = .002$, CI [.00, .02]. However, in an ANCOVA with performance as a covariate, only performance was significant, $F(1, 388) = 22.88$, $p < .001$, $\eta_p^2 = .06$, CI [.02, .10]; there was no main effect of consistency, $F(1, 388) = .01$, $p = .93$, $\eta_p^2 < .001$, CI [.00, .002], nor of warning, $F(1, 388) = 2.78$, $p = .10$, $\eta_p^2 = .01$, CI [.00, .03]. Further, the interaction between consistency and warning was not significant, $F(1, 388) = 1.39$, $p = .24$, $\eta_p^2 = .004$, CI [.00, .02].

Perceived ease. Interestingly, the fluency manipulation did not seem to affect how easy or difficult participants explicitly reported the task to be. As in other measures, performance positively predicted perceived ease, $\beta = .04$, $SE = .02$, $t(392) = 2.48$, $p = .01$, so that the better participants did, the easier they thought the task was. An ANCOVA with performance as a covariate found no main effect of consistency, $F(1, 389) = 1.66$, $p = .20$, $\eta_p^2 = .004$, 90% CI [.00, .02], nor of warning, $F(1, 389) = .31$, $p = .58$, $\eta_p^2 = .001$, CI [.00, .01]. Further, the interaction between consistency and warning was not significant, $F(1, 389) = 1.09$, $p = .30$, $\eta_p^2 = .003$, CI [.00, .02]. When not warned, consistent and inconsistent participants thought the task was equally difficult (adj. $M_{\text{consistent}} = 2.26$, $SE = .11$ vs. adj. $M_{\text{inconsistent}} = 2.28$, $SE = .11$), $F(1, 389) = 0.03$, $p = .86$, $\eta_p^2 < .001$, CI [.00, .003], and inconsistent participants who had been warned thought the task was slightly but not significantly easier than consistent participants receiving the warning (adj. $M_{\text{consistent}} = 2.08$, $SE = .11$ vs. adj. $M_{\text{inconsistent}} = 2.38$, $SE = .11$), $F(1, 389) = 2.76$, $p = .10$, $\eta_p^2 = .007$, CI [.00, .03]. We discuss the potential meaning of this finding in the General Discussion.

Awareness of procedural fluency. Finally, participants did not believe that the location of the highlighted letter affected their ability to find the anagrams. The mean rating for this question was below the midpoint of the scale ($M_{\text{overall}} = 2.53$, $SD = 1.11$), and there were no significant effects of performance, $F(1, 386) = .10$, $p = .75$, $\eta_p^2 < .001$, 90% CI [.00, .01]; consistency, $F(1, 386) < .001$, $p = 1.00$, $\eta_p^2 < .001$, CI [.00, 1.00]; warning, $F(1, 386) = .91$, $p = .34$, $\eta_p^2 = .002$, CI [.00, .02]. Neither was there an interaction, $F(1, 386) = .01$, $p = .91$, $\eta_p^2 < .001$, CI [.00, .003].

Discussion

In summary, Experiment 1 demonstrated that individuals who encounter superficially consistent materials while completing a task feel more confident in their performance, to the point that their confidence may exceed their actual performance levels, than individuals who encounter superficial inconsistency, whose confidence may underestimate their performance levels. Thus, although they were no more successful in solving the anagrams, individuals who experienced procedural consistency (vs. inconsistency) believed that they performed better. However, drawing attention to how consistently or inconsistently the task materials were presented attenuated this difference. These effects are consistent with how fluency is understood to work, supporting the idea of super-

[2] Note that this finding converges with past research (e.g., Williams et al., 2013), which suggests that consistent participants may not only be more confident than inconsistent participants but also more confident than is warranted by their true level of performance. However, the consistency in prior investigations was in the strategy participants themselves chose to deploy to solve the problems, rather than superficial consistency in task features, as in the current research.

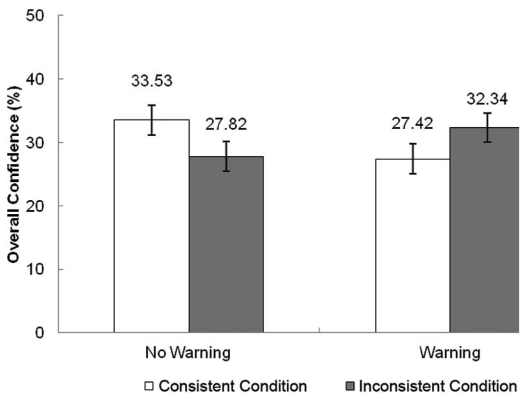

Figure 3. Overall confidence ratings, Experiment 1. Estimated marginal means for overall confidence ratings within the consistent and inconsistent conditions, with and without warning. Error bars represent standard errors of the means.

ficial procedural consistency as a new type of fluency induction, and procedural fluency itself a new type of fluency. We further tested this claim in Experiment 2.

Experiment 2

When people experience fluency, they interpret this feeling of ease through the lens of a naïve theory (Schwarz, 2004). In Experiment 2, we examined the role of naïve theories in procedural fluency by explicitly manipulating participants' beliefs about what "ease" means. Participants completed a number sequence game with superficial consistency or inconsistency in the task's presentation. We expected that participants would again show the procedural fluency effect, reporting higher confidence in their performance when the sequences proceeded in a consistent rather than inconsistent fashion. However, for half of our participants, we manipulated the lay belief they were operating under before they completed the task: we told these participants that experiencing *difficulty* is actually a signal of successful performance. We predicted that adding this naïve theory manipulation would reduce or eliminate the difference between consistent and inconsistent participants.

Method

Participants. Two hundred fifty-six individuals drawn from the Mechanical Turk (MTurk) platform (48.4% female, M_{age} = 34.0 years) participated for $0.30. We set a target sample size of at least 50 participants per cell, which would provide 80% power to detect a small-to-medium effect size, and collected data for approximately one full day on MTurk. Some participants did not complete all follow-up questions, leading to varying degrees of freedom in the analyses below.

Procedure. Participants were randomly assigned to one of four conditions in a 2 (consistent vs. inconsistent) × 2 (naïve theory: difficulty-success-manipulation vs. control) between-participants design. Participants encountered a series of five number sequences, each containing seven numbers and a blank space. Their task was to determine, from the available numbers in the sequence, which number belonged in the blank space to continue the sequence. They could use scratch paper as needed.

Participants in the consistent conditions solved for the seventh (i.e., second to last) number in each sequence, whereas inconsistent condition participants solved for the third, seventh, fourth, fifth, and eighth numbers, respectively. Participants in the difficulty-success-manipulation conditions read an additional passage after reading the instructions but before they began the task:

> Take note: Although it seems counterintuitive to some people, evidence suggests that feeling like you're working hard to solve a problem is actually a sign of better performance than feeling like a problem is easy. When facing difficulty in solving a problem, individuals engage in **deeper processing** to counteract it. This greater involvement with the task material improves concentration and fosters greater success. Please keep this in mind as you form an impression of your task performance on the number sequences today.

In other words, this message suggested that the experience of fluency during the task's progression may actually be indicative of poorer performance, and that disfluency can indicate deeper thought and analysis and thus may serve as a signal of better performance. This passage was simply omitted for participants in the *control* conditions.

Measures. Participants completed measures regarding their confidence about each number sequence and about the task overall. After solving each sequence, participants indicated how confident they were that they had completed that specific sequence correctly, from 1 (0%: *not at all confident*) to 11 (100%: *completely confident*), reported below as percentages. As in Experiment 1, we averaged these ratings on each sequence to form a single measure, "item-level confidence," in the analyses below. Following completion of all five sequences, participants indicated how confident they felt overall in their ability to find the solutions as they were working on the number sequences (what we call "overall confidence" below), from 1 (0%: *not at all confident*) to 11 (100%: *completely confident*), reported below as percentages. They also indicated how easy or difficult they found the task to be overall ("perceived ease"), from 1 (*very difficult*) to 7 (*very easy*). They then noted whether they had used anything other than paper and a pen or pencil to solve the sequences. Finally, all participants provided their age, ethnicity, and gender, and were told the solutions to the number sequences.

Results

Performance. Inconsistent participants actually performed slightly better (solving $M = 2.91$ sequences correctly, $SD = 1.29$) than consistent participants did ($M = 2.59$, $SD = 1.28$), $F(1, 254) = 3.78$, $p = .05$, $\eta_p^2 = .02$, 90% CI [.00, .05]. Furthermore, participants in both conditions spent similar amounts of time solving the number sequences ($M_{\mathrm{consistent}} = 321.98$ s, $SD = 238.47$; $M_{\mathrm{inconsistent}} = 332.65$ s, $SD = 259.66$); $F(1, 255) = 0.12$, $p = .73$, $\eta_p^2 < .001$, CI [.00, .01]; log-transformed: $F(1, 255) = 0.002$, $p = .96$, $\eta_p^2 < .001$, CI [.00, 1.00]. Thus, participants experiencing superficial consistency (vs. inconsistency) were neither more accurate nor faster at actually solving the problems, and any differences in their confidence cannot be attributed to procedural consistency making the task actually easier to complete. The naïve theories manipulation also did not affect participants' performance ($M_{\text{manipulated}} = 2.71$, $SD = 1.32$ vs. $M_{\text{control}} = 2.79$, $SD = 1.26$), $F(1, 254) = .21$, $p = .65$, $\eta_p^2 < .001$, CI [.00, .02], and further did not interact with the consistency manipulation to influence performance, $F(1, 252) = .27$, $p = .60$, $\eta_p^2 = .001$, CI [.00, .02]. Here again, differences in confidence are unlikely to be due to actual differences in performance.

Item-level confidence. As in Experiment 1, participants who performed better (i.e., solved more sequences correctly) were more confident in their performance, $\beta = 10.37$, $SE = 88$, $t(254) = 11.85$, $p < .001$; to control for this relationship, in this and subsequent analyses, performance was included as a covariate. As predicted, consistency led to confidence, unless participants learned of an alternate naïve theory to explain their performance. An ANCOVA with performance as a covariate revealed a main effect of consistency, $F(1, 251) = 9.00$, $p = .003$, $\eta_p^2 = .04$, 90% CI [.007, .08], and no main effect of the naïve theory manipulation, $F(1, 251) = .16$, $p = .69$, $\eta_p^2 = .001$, CI [.00, .02]. However, consistent with our predictions, there was a significant interaction between these two factors, $F(1, 251) = 4.12$, $p = .04$, $\eta_p^2 = .02$, CI [.0002, .05]. As shown in Figure 4, when naïve theories were not manipulated, consistent participants were more confident (adj. $M = 69.30$, $SE = 2.24$) than inconsistent participants were (adj. $M = 58.11$, $SE = 2.22$), $F(1, 251) = 12.52$, $p < .001$, $\eta_p^2 = .05$, 90% CI [.01, .10]. However, the difference between consistent and inconsistent participants was attenuated when we manipulated their theories about the meaning of procedural fluency (adj. $M_{\text{consistent}} = 63.92$, $SE = 2.20$; adj. $M_{\text{inconsistent}} = 61.72$, $SE = 2.22$), $F(1, 251) = .49$, $p = .48$, $\eta_p^2 = .002$, CI [.00, .02].

Item-level overconfidence. Interfering with participants' naïve theories also reduced the gap between how well-calibrated consistent and inconsistent participants' performance assessments were. In this experiment, participants completed five number sequences and solved each sequence either correctly or incorrectly. Therefore, confidence calibration in this experiment was computed by averaging participants' item-level confidence measures and subtracting the total proportion of sequences they had solved correctly. For example, if a participant reported 50% average confidence (a confidence score of 0.5) but only solved two of the five sequences correctly (an accuracy score of 0.4), they would receive an overconfidence score of 0.1.

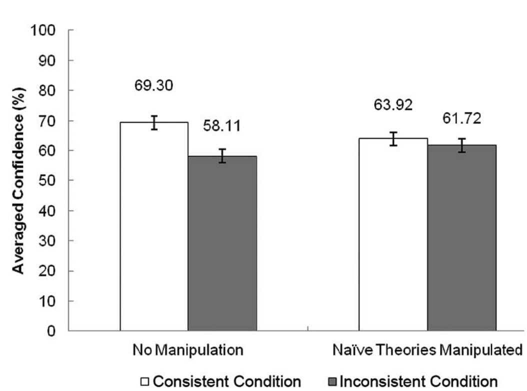

Figure 4. Averaged confidence ratings, Experiment 2. Estimated marginal means for averaged confidence ratings within the consistent and inconsistent conditions, with and without manipulating naïve theories. Error bars represent standard errors of the means.

An ANOVA revealed a main effect of consistency, $F(1, 252) = 12.92$, $p < .001$, $\eta_p^2 = .05$, 90% CI [.01, .10], and no main effect of the naïve theory manipulation, $F(1, 252) = .01$, $p = .93$, $\eta_p^2 < .001$, CI [.00, .002], but a significant interaction between these two factors, $F(1, 252) = 3.92$, $p = .049$, $\eta_p^2 = .02$, CI [.00, .05]. When naïve theories were not manipulated, consistent participants were more confident than their performance suggested they should be ($M = .16$, $SE = .03$), more so than were inconsistent participants ($M = .01$, $SE = .03$), $F(1, 252) = 15.42$, $p < .001$, $\eta_p^2 = .06$, CI [.02, .11]. However, manipulating naïve theories brought their calibration levels closer together ($M_{\text{consistent}} = .10$, $SE = .03$ vs. $M_{\text{inconsistent}} = .06$, $SE = .03$), $F(1, 252) = 1.31$, $p = .25$, $\eta_p^2 = .005$, CI [.00, .03].

Overall confidence. When asked to rate their overall confidence after completing all five sequences, consistent participants were also more confident, unless their naïve theories had been manipulated. Performance again predicted confidence, $\beta = 13.01$, $SE = 1.09$, $t(254) = 11.95$, $p < .001$, and is included as a covariate. An ANCOVA revealed a significant effect of consistency, $F(1, 251) = 8.26$, $p = .004$, $\eta_p^2 = .03$, 90% CI [.006, .08], no main effect of the naïve theory manipulation, $F(1, 251) = .24$, $p = .62$, $\eta_p^2 = .001$, CI [.00, .02], and a marginal interaction between these two factors, $F(1, 251) = 3.30$, $p = .07$, $\eta_p^2 = .01$, CI [.00, .05]. When naïve theories were not manipulated, consistent participants were more confident overall (adj. $M = 64.09$, $SE = 2.79$) than were inconsistent participants (adj. $M = 51.07$, $SE = 2.77$), $F(1, 251) = 10.89$, $p = .001$, $\eta_p^2 = .04$, CI [.01, .09]. However, the gap between consistent and inconsistent participants shrank substantially when we manipulated their naïve theories (adj. $M_{\text{consistent}} = 57.71$, $SE = 2.75$; adj. $M_{\text{inconsistent}} = 54.73$, $SE = 2.76$), $F(1, 251) = .58$, $p = .45$, $\eta_p^2 = .002$, CI [.00, .02]. See Figure 5.

Perceived ease. Here, consistency made the task feel easier. As in prior analyses, participants who performed better also perceived the task to be easier, $\beta = .56$, $SE = .07$, $t(254) = 8.31$, $p < .001$. An ANCOVA with performance as a covariate revealed a main effect of consistency, $F(1, 251) = 7.17$, $p = .008$, $\eta_p^2 = .03$, 90% CI [.004, .07], with no main effect of the naïve theory manipulation, $F(1, 251) = .75$, $p = .39$, $\eta_p^2 = .003$, CI [.00, .02]. The interaction between consistency and the naïve theory manipulation was not significant, $F(1, 251) = .69$, $p = .41$, $\eta_p^2 = .003$, CI [.00, .02]. Nevertheless, when participants' naïve theories were not manipulated, consistent participants believed it was significantly easier to complete (adj. $M = 3.62$, $SE = .17$) than inconsistent participants did (adj. $M = 3.02$, $SE = .17$), $F(1, 251) = 6.11$, $p = .01$, $\eta_p^2 = .02$, CI [.003, .06], while the gap in perceptions of ease shrank when we did manipulate their naïve theories (adj. $M_{\text{consistent}} = 3.33$, $SE = .17$ vs. adj. $M_{\text{inconsistent}} = 3.02$, $SE = .17$), $F(1, 251) = 1.75$, $p = .19$, $\eta_p^2 = .007$, CI [.00, .03].

Discussion

Overall, Experiment 2 replicated the finding that although they were no faster and no better at actually solving the problems, individuals experiencing superficial consistency expressed greater confidence in their performance. Furthermore, suggesting new naïve theories to participants that reframed the meaning of the ease or difficulty of the task, and could change their interpretation of what procedural fluency signals about their performance, elimi-

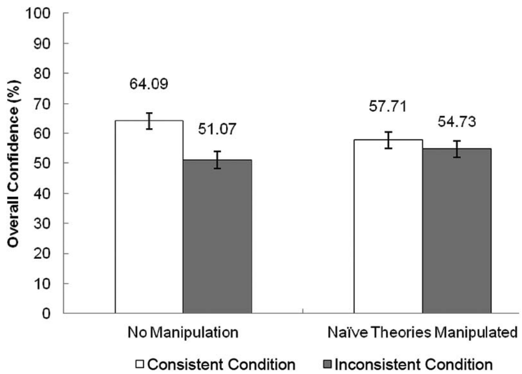

Figure 5. Overall confidence ratings, Experiment 2. Estimated marginal means for overall confidence ratings within the consistent and inconsistent conditions, with and without manipulating naïve theories. Error bars represent standard errors of the means.

nated this effect. Specifically, telling participants that experiencing difficulty is a sign of good performance (a proposal that runs counter to lay intuition) disrupted the influence of superficial consistency on their confidence.

General Discussion

Merely presenting participants with tasks that involved a superficially consistent rather than inconsistent procedure made them more confident that they had completed the task correctly, without actually leading them to perform any better. However, drawing attention to the procedural consistency reduced its impact on participants' confidence in their performance (Experiment 1). Furthermore, the effect of procedural fluency on confidence was also influenced by the naïve theories people held about what such fluency means (Experiment 2). When participants were led to believe that the experience of difficulty during task completion could actually be a sign of better performance, procedural fluency no longer led them to be more confident in their performance on the task.

It is interesting to note that although procedural consistency influenced confidence in the same ways as other types of fluency, and in similar ways across these studies, participants reported that the consistent version felt easier to complete only in Experiment 2, and this did not interact with the manipulation warning them that ease may not mean what they think it does. This may be because we asked participants about experienced ease after they had completed the task, well after the feeling of ease or difficulty may have faded. Further, fluency can have an effect even when people are not consciously aware of the source of the fluency (e.g., Thomas & Morwitz, 2009; Topolinski & Strack, 2009a), and our instantiations of it likewise may have been too subtle to explicitly register with our participants, even while influencing their confidence assessments.

Regardless, we would again point to our particular pattern of results, which shows that in the control conditions, the consistent and inconsistent participants report differing levels of confidence in their performance despite the fact that participants' actual accuracy in solving the trials was the same in both conditions. This finding means that their differing experience in completing the trials, rather than any differences in performance on the trials, was influencing their confidence estimates. We also found that this effect is attenuated when attention is drawn to the experience of consistency or inconsistency embedded in the trials or when participants are given an alternate interpretation of the fluency they experienced. These patterns are consistent with how fluency is known to affect confidence estimates in other contexts and experiments. Thus, we suggest they represent a demonstration of procedural fluency.

It is important to note the distinction between the superficial procedural consistency we investigated here (the form producing procedural fluency) and other forms of consistency that tasks may present, such as logical or cognitive consistency (see Winkielman, Huber, Kavanagh, & Schwarz, 2012 for a review). Research investigating these other forms of consistency has found, for example, that people implicitly prefer sets of sentences that are logically "consistent" (i.e., are logically valid combinations) with one another (Trippas, Handley, Verde, & Morsanyi, 2016); are uncomfortable endorsing ideas that are inconsistent with their preexisting preferences (Elliot & Devine, 1994); and feel better about their choices and believe others share their preferences when their method of choosing resembles the options themselves, as when choosing liked options and rejecting disliked options (Perfecto, Galak, Simmons, & Nelson, 2017). Logical and cognitive consistency are also similar to semantic coherence, in which items are semantically linked (as in a remote associates task) and thus facilitate faster recognition of each other (Topolinski & Strack, 2009b). Importantly, cognitive consistency and semantic coherence are both forms of congruity: a sense of items, objects, or ideas "meshing well" together. Note in the current research that we did nothing to manipulate the semantic or logical relatedness of stimuli. Instead, our procedure is more a form of constancy: items sharing the shape or form of preceding and following stimuli. Accordingly, our effects diverge from the consistency studied in research on cognitive consistency and sematic coherence.

Our work adds to the literature on fluency in other ways. There has been a movement in metacognition research to look beyond how people assess performance on relatively straightforward memory and information retrieval tasks, and to instead examine how people perceive their performance in more complex reasoning contexts (e.g., Ackerman & Thompson, 2017). Less is known about whether typical fluency effects work similarly with metareasoning and problem-solving tasks and, if so, how the strength of such effects differs from established findings with simpler tasks. As an initial attempt at this endeavor, Thompson et al. (2013) pitted perceptual fluency (instantiated via font readability) against answer fluency (measured by speed in initial response). They found that, although neither type of fluency predicted accuracy of judgment, answer fluency predicted participants' confidence while perceptual fluency did not. Our work is a preliminary addition to the literature on fluency effects in complex reasoning tasks.

Our results also suggest that researchers may wish to take the possibility of procedural fluency into consideration when selecting a research design. Specifically, we find that when presented with multiple trials of a task, the relative consistency or inconsistency in superficial features across trials may induce a feeling of (dis)fluency. Thus, we suggest that researchers might want to keep in mind that within-subjects designs have the potential to be affected by procedural (dis)fluency, with consequent effects on participants' judgments.

Our conceptual framework points to interesting future directions for research. As an example, although this article focused on the effects of procedural fluency on confidence, past work has documented numerous types of judgments that follow from fluency. Drawing on this prior work, superficial procedural consistency might also make the task itself appear more likable, feel more familiar, seem of greater value, and more. The connection between consistency and confidence found in other work (e.g., Williams et al., 2013) made confidence a natural starting place to test for procedural fluency, but subsequent work examining other judgments (e.g., liking or value) would be valuable.

This work also has several real-world implications. One domain where consistency-induced confidence could be problematic is testing and assessment. It is possible that students might take more care to check their answers or more carefully approach questions when exams are not procedurally consistent throughout—that is, when the question formats are varied more often, breaking up multiple choice and essay sections so that different question types are interspersed. This suggestion is in tune with educational research on "desirable difficulties" (see Bjork & Bjork, 2011 for a review), which broadly touts the benefits of providing variety on attaining educational outcomes. For example, this work advocates for varying the environmental settings (e.g., multiple rooms) in which learning takes place to produce more transferable knowledge (Smith, Glenberg, & Bjork, 1978) and for presenting high-knowledge learners with less semantically coherent text to boost recall (McNamara, Kintsch, Songer, & Kintsch, 1996). Although this work attributes the positive benefits of what we would call procedural inconsistency to its ability to activate greater cognitive processing that enhances performance, our research suggests that these types of "desirable difficulties" may also work to improve learning by attenuating overconfidence directly and encouraging more careful work.

Further, our results are consistent with the idea that fluency as a class could be expanded to include not only cognitive domains (e.g., Alter & Oppenheimer, 2009) but also physical ones. Investigations of metacognitive fluency have largely focused on how they influence assessments of cognitive operations. Our results suggest, however, that akin to motoric fluency (Susser et al., 2017), procedural fluency may also arise for purely physical tasks. This can explain why it can be hard to tell until it is too late, for example, that you have skipped a stitch while knitting. This notion is also consistent with work on desirable difficulties in learning movement skills, in which confident perceptions of learning on physical tasks are produced more by blocked (i.e., repetitive) practice than by interleaved (i.e., more distributed) practice, although long-term performance is enhanced more by the latter than the former (e.g., Simon & Bjork, 2002).

As another example, forms like government benefit applications, insurance claims, and tax returns often contain numerous fields with uniform appearances. This consistency in field formats may induce procedural fluency and inflate users' confidence that they have correctly filled out the forms, reducing the likelihood that they will double-check their work or seek advice from professionals. For the same reasons, this consistency may also increase the likelihood that professionals who repeatedly fill out the same forms might miss mistakes. Policymakers might consider making these forms superficially less consistent to reduce mistakes on such forms (of course, taking care not to overcomplicate the forms and reduce compliance and accuracy). In the policy domain, an additional potential application concerns nutrition labels on foods. Despite encountering them frequently, consumers still often misinterpret these labels (Cowburn & Stockley, 2005). Reminding consumers of how easy labels may *seem* to use, drawing their attention to the procedural fluency accompanying them, may encourage more careful calculations and choices.

Our research also suggests that procedural consistency can foster confidence even when it may not be warranted. For instance, when making a series of similar predictions, as when admissions officers consider the likelihood of success of a series of potential students, superficially consistent presentation of candidate information might lead forecasters to overestimate how likely their predictions are to come true. However, altering how decision makers interpret the sense of ease that can be induced by consistency (i.e., their naïve theories) may improve the calibration of their confidence estimates. Miscalibrated confidence also has negative effects on driving safety. Cities and towns in the United States and Europe, for example, have discovered that carefully planned interventions can momentarily reduce driver (over)confidence, curtailing speeding and thus traffic accidents and fatalities. One successful intervention involves creating irregular and unpredictable obstacles for drivers to maneuver around: trees, planters, speed bumps, and, in one Dutch town, fake road work sites (e.g., Vanderbilt, 2008). Our research suggests that the success of this tactic has even more basis in driver psychology than originally thought—one factor contributing to the success of such interventions may perhaps be how such interventions introduce procedural inconsistency in how drivers navigate, encouraging drivers to take greater care.

In general, feeling correct can be a good sign of correct judgment, but this feeling can also potentially lead people astray. Completing a task with superficially consistent procedures leads to a sense that one has also completed the task correctly, but our research shows that this feeling may be independent of one's true success. Procedural fluency, and its antidotes, can help us understand why people make the choices they do and help them to more consistently make better ones.

References

Ackerman, R., & Thompson, V. A. (2017). Meta-reasoning: Monitoring and control of thinking and reasoning. *Trends in Cognitive Sciences, 21,* 607–617. http://dx.doi.org/10.1016/j.tics.2017.05.004

Alter, A. L., & Oppenheimer, D. M. (2009). Uniting the tribes of fluency to form a metacognitive nation. *Personality and Social Psychology Review, 13,* 219–235. http://dx.doi.org/10.1177/1088868309341564

Alter, A. L., Oppenheimer, D. M., Epley, N., & Eyre, R. N. (2007). Overcoming intuition: Metacognitive difficulty activates analytic reasoning. *Journal of Experimental Psychology: General, 136,* 569–576. http://dx.doi.org/10.1037/0096-3445.136.4.569

Bjork, E. L., & Bjork, R. A. (2011). Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. In M. A. Gernsbacher, R. W. Pew, L. M. Hough, & J. R. Pomerantz (Eds.), *Psychology and the real world: Essays illustrating fundamental contributions to society* (pp. 56–64). New York, NY: Worth.

Cowburn, G., & Stockley, L. (2005). Consumer understanding and use of nutrition labelling: A systematic review. *Public Health Nutrition, 8,* 21–28. http://dx.doi.org/10.1079/PHN2004666

Dechêne, A., Stahl, C., Hansen, J., & Wänke, M. (2009). Mix me a list: Context moderates the truth effect and the mere-exposure effect. *Journal of Experimental Social Psychology, 45,* 1117–1122. http://dx.doi.org/10.1016/j.jesp.2009.06.019

Dechêne, A., Stahl, C., Hansen, J., & Wänke, M. (2010). The truth about the truth: A meta-analytic review of the truth effect. *Personality and Social Psychology Review, 14,* 238–257. http://dx.doi.org/10.1177/1088868309352251

Diemand-Yauman, C., Oppenheimer, D. M., & Vaughan, E. B. (2011). Fortune favors the bold (and the italicized): Effects of disfluency on educational outcomes. *Cognition, 118,* 111–115. http://dx.doi.org/10.1016/j.cognition.2010.09.012

Elliot, A. J., & Devine, P. G. (1994). On the motivational nature of cognitive dissonance: Dissonance as psychological discomfort. *Journal of Personality and Social Psychology, 67,* 382–394. http://dx.doi.org/10.1037/0022-3514.67.3.382

Falk, A., & Zimmermann, F. (2017). Consistency as a signal of skills. *Management Science, 63,* 2197–2210. http://dx.doi.org/10.1287/mnsc.2016.2459

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. http://dx.doi.org/10.1037/0022-3514.75.5.1101

Greifeneder, R., & Bless, H. (2010). The fate of activated information in impression formation: Fluency of concept activation moderates the emergence of assimilation versus contrast. *British Journal of Social Psychology, 49,* 405–414. http://dx.doi.org/10.1348/014466609X479699

Hansen, J., & Wänke, M. (2013). Fluency in context: Discrepancy makes processing experiences informative. In C. Unkelbach & R. Greifeneder (Eds.), *The experience of thinking* (pp. 80–94). New York, NY: Psychology Press.

Hertzog, C., Dunlosky, J., Robinson, A. E., & Kidder, D. P. (2003). Encoding fluency is a cue used for judgments about learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29,* 22–34. http://dx.doi.org/10.1037/0278-7393.29.1.22

Herzog, S. M., & Hertwig, R. (2013). The ecological validity of fluency. In R. Greifeneder & C. Unkelbach (Eds.), *The experience of thinking: How the fluency of mental processes influences cognition and behavior* (pp. 190–219). London, England: Psychology Press.

Kühl, A., & Eitel, T. (2016). Effects of disfluency on cognitive and metacognitive processes and outcomes. *Metacognition and Learning, 11,* 1–13. http://dx.doi.org/10.1007/s11409-016-9154-x

Magreehan, D. A., Serra, M. J., Schwartz, N. H., & Narciss, S. (2016). Further boundary conditions for the effects of perceptual disfluency on judgments of learning. *Metacognition and Learning, 11,* 35–56. http://dx.doi.org/10.1007/s11409-015-9147-1

McNamara, D. S., Kintsch, E., Songer, N. B., & Kintsch, W. (1996). Are good texts always better? Interactions of text coherence, background knowledge, and levels of understanding in learning from text. *Cognition and Instruction, 14,* 1–43. http://dx.doi.org/10.1207/s1532690xci1401_1

Meyer, A., Frederick, S., Burnham, T. C., Guevara Pinto, J. D., Boyer, T. W., Ball, L. J., . . . Schuldt, J. P. (2015). Disfluent fonts don't help people solve math problems. *Journal of Experimental Psychology: General, 144,* e16–e30. http://dx.doi.org/10.1037/xge0000049

Miele, D. B., Finn, B., & Molden, D. C. (2011). Does easily learned mean easily remembered? It depends on your beliefs about intelligence. *Psychological Science, 22,* 320–324. http://dx.doi.org/10.1177/0956797610397954

Miele, D. B., & Molden, D. C. (2010). Naive theories of intelligence and the role of processing fluency in perceived comprehension. *Journal of Experimental Psychology: General, 139,* 535–557. http://dx.doi.org/10.1037/a0019745

Miele, D. B., Son, L. K., & Metcalfe, J. (2013). Children's naive theories of intelligence influence their metacognitive judgments. *Child Development, 84,* 1879–1886. http://dx.doi.org/10.1111/cdev.12101

Novemsky, N., Dhar, R., Schwarz, N., & Simonson, I. (2007). Preference fluency and consumer choice. *Journal of Marketing Research, 44,* 347–356. http://dx.doi.org/10.1509/jmkr.44.3.347

O'Connor, K., & Cheema, A. (2018). Do evaluations rise with experience? *Psychological Science, 29,* 779–790. http://dx.doi.org/10.1177/0956797617744517

Oppenheimer, D. M. (2004). Spontaneous discounting of availability in frequency judgment tasks. *Psychological Science, 15,* 100–105. http://dx.doi.org/10.1111/j.0963-7214.2004.01502005.x

Oppenheimer, D. M., & Alter, A. L. (2014). The search for moderators in disfluency research. *Applied Cognitive Psychology, 28,* 502–504. http://dx.doi.org/10.1002/acp.3023

Perfecto, H., Galak, J., Simmons, J. P., & Nelson, L. D. (2017). Rejecting a bad option feels like choosing a good one. *Journal of Personality and Social Psychology, 113,* 659–670. http://dx.doi.org/10.1037/pspa0000092

Pieger, E., Mengelkamp, C., & Bannert, M. (2017). Fostering analytic metacognitive processes and reducing overconfidence by disfluency: The role of contrast effects. *Applied Cognitive Psychology, 31,* 291–301. http://dx.doi.org/10.1002/acp.3326

Reber, R., Brun, M., & Mitterndorfer, K. (2008). The use of heuristics in intuitive mathematical judgment. *Psychonomic Bulletin & Review, 15,* 1174–1178. http://dx.doi.org/10.3758/PBR.15.6.1174

Reber, R., & Schwarz, N. (1999). Effects of perceptual fluency on judgments of truth. *Consciousness and Cognition: An International Journal, 8,* 338–342. http://dx.doi.org/10.1006/ccog.1999.0386

Reber, R., & Zupanek, N. (2002). Effects of processing fluency on estimates of probability and frequency. In P. Sedlmeier & T. Betsch (Eds.), *Frequency processing and cognition* (pp. 175–188). Oxford, England: Oxford University Press. http://dx.doi.org/10.1093/acprof:oso/9780198508632.003.0011

Reder, L. M. (1987). Strategy selection in question answering. *Cognitive Psychology, 19,* 90–138. http://dx.doi.org/10.1016/0010-0285(87)90005-3

Rummer, R., Schweppe, J., & Schwede, A. (2016). Fortune is fickle: Null-effects of disfluency on learning outcomes. *Metacognition and Learning, 11,* 57–70. http://dx.doi.org/10.1007/s11409-015-9151-5

Schwarz, N. (2004). Metacognitive experiences in consumer judgment and decision making. *Journal of Consumer Psychology, 14,* 332–348. http://dx.doi.org/10.1207/s15327663jcp1404_2

Schwarz, N., & Clore, G. L. (2007). Feelings and phenomenal experiences. In E. T. Higgins & A. Kruglanski (Eds.), *Social psychology: Handbook of basic principles* (2nd ed., pp. 385–407). New York, NY: Guilford Press.

Simon, D. A., & Bjork, R. A. (2002). Models of performance in learning multisegment movement tasks: Consequences for acquisition, retention, and judgments of learning. *Journal of Experimental Psychology: Applied, 8,* 222–232. http://dx.doi.org/10.1037/1076-898X.8.4.222

Smith, S. M., Glenberg, A., & Bjork, R. A. (1978). Environmental context and human memory. *Memory & Cognition, 6,* 342–353. http://dx.doi.org/10.3758/BF03197465

Stevenson, L. M., & Carlson, R. A. (2020). Consistency, not speed: Temporal regularity as a metacognitive cue. *Psychological Research, 84,* 88–98. http://dx.doi.org/10.1007/s00426-018-0973-z

Sungkhasettee, V. W., Friedman, M. C., & Castel, A. D. (2011). Memory and metamemory for inverted words: Illusions of competency and desirable difficulties. *Psychonomic Bulletin & Review, 18,* 973–978. http://dx.doi.org/10.3758/s13423-011-0114-9

Susser, J. A., Panitz, J., Buchin, Z., & Mulligan, N. W. (2017). The motoric fluency effect on metamemory. *Journal of Memory and Language, 95,* 116–123. http://dx.doi.org/10.1016/j.jml.2017.03.002

Thomas, M., & Morwitz, V. G. (2009). The ease of computation effect: The interplay of metacognitive experiences and naïve theories in judgments of price difference. *Journal of Marketing Research, 46,* 81–91. http://dx.doi.org/10.1509/jmkr.46.1.81

Thompson, V. A., Turner, J. A. P., Pennycook, G., Ball, L. J., Brack, H., Ophir, Y., & Ackerman, R. (2013). The role of answer fluency and perceptual fluency as metacognitive cues for initiating analytic thinking. *Cognition, 128,* 237–251. http://dx.doi.org/10.1016/j.cognition.2012.09.012

Topolinski, S., Bakhtiari, G., & Erle, T. M. (2016). Can I cut the Gordian tnok? The impact of pronounceability, actual solvability, and length on intuitive problem assessments of anagrams. *Cognition, 146,* 439–452. http://dx.doi.org/10.1016/j.cognition.2015.10.019

Topolinski, S., & Strack, F. (2009a). Scanning the "fringe" of consciousness: What is felt and what is not felt in intuitions about semantic coherence. *Consciousness and Cognition: An International Journal, 18,* 608–618. http://dx.doi.org/10.1016/j.concog.2008.06.002

Topolinski, S., & Strack, F. (2009b). The analysis of intuition: Processing fluency and affect in judgements of semantic coherence. *Cognition and Emotion, 23,* 1465–1503. http://dx.doi.org/10.1080/02699930802420745

Trippas, D., Handley, S. J., Verde, M. F., & Morsanyi, K. (2016). Logic brightens my day: Evidence for implicit sensitivity to logical validity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 42,* 1448–1457. http://dx.doi.org/10.1037/xlm0000248

Vanderbilt, T. (2008). *Traffic*. New York, NY: Vintage.

Wänke, M., & Hansen, J. (2015). Relative processing fluency. *Current Directions in Psychological Science, 24,* 195–199. http://dx.doi.org/10.1177/0963721414561766

Whittlesea, B. W. A., & Williams, L. D. (2000). The source of feelings of familiarity: The discrepancy-attribution hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26,* 547–565. http://dx.doi.org/10.1037/0278-7393.26.3.547

Wilcox, K., & Song, S. (2011). Discrepant fluency in self-customization. *Journal of Marketing Research, 48,* 729–740. http://dx.doi.org/10.1509/jmkr.48.4.729

Williams, E. F., Dunning, D., & Kruger, J. (2013). The hobgoblin of consistency: Algorithmic judgment strategies underlie inflated self-assessments of performance. *Journal of Personality and Social Psychology, 104,* 976–994. http://dx.doi.org/10.1037/a0032416

Winkielman, P., Huber, D. E., Kavanagh, L., & Schwarz, N. (2012). Fluency of consistency: When thoughts fit nicely and flow smoothly. In B. Gawronski & F. Strack (Eds.), *Cognitive consistency: A fundamental principle in social cognition* (pp. 89–111). New York, NY: Guilford Press.

Zajonc, R. (2001). Mere exposure: A gateway to the subliminal. *Current Directions in Psychological Science, 10,* 224–228. http://dx.doi.org/10.1111/1467-8721.00154

Received February 23, 2017
Revision received March 15, 2020
Accepted March 18, 2020 ■