

A hard-to-read font reduces the framing effect in a large sample

Christoph W. Korn¹ · Juliane Ries² · Lennart Schalk^{3,4} · Yulia Oganian⁵ · Henrik Saalbach⁶

Published online: 30 October 2017
© Psychonomic Society, Inc. 2017

Abstract How can apparent decision biases, such as the framing effect, be reduced? Intriguing findings within recent years indicate that foreign language settings reduce framing effects, which has been explained in terms of deeper cognitive processing. Because hard-to-read fonts have been argued to trigger deeper cognitive processing, so-called cognitive disfluency, we tested whether hard-to-read fonts reduce framing effects. We found no reliable evidence for an effect of hard-to-read fonts on four framing scenarios in a laboratory (final $N = 158$) and an online study ($N = 271$). However, in a preregistered online study with a rather large sample ($N = 732$), a hard-to-read font reduced the framing effect in the classic “Asian disease” scenario (in a one-sided test). This suggests that hard-read-fonts can modulate decision biases—albeit with rather small effect sizes. Overall, our findings stress the importance of large samples for the reliability and replicability of modulations of decision biases.

Keywords Decision-making biases · Apparent irrationalities · Cognition · Syllogistic reasoning

✉ Christoph W. Korn
c.korn@uke.de

¹ Institute for Systems Neuroscience, University Medical Center Hamburg-Eppendorf, Martinistrasse 52, 20246 Hamburg, Germany

² Saarland University, Saarbrücken, Germany

³ ETH Zurich, Zurich, Switzerland

⁴ PH Schwyz, Goldau, Switzerland

⁵ School of Medicine, University of California San Francisco, San Francisco, CA, USA

⁶ Leipzig University, Leipzig, Germany

Introduction

Humans often succumb to seemingly irrational decision biases (Gigerenzer & Gaissmaier, 2011; Kahneman, 2003; Tversky & Kahneman, 1981), but it remains relatively underexplored which manipulations alter such biases. Several recent studies have provided evidence that manipulating language contexts affects prominent decision biases, most notably framing effects (Costa, Foucart, Amon, Aparici, & Apesteguia, 2014; Gao, Zika, Rogers, & Thierry, 2015; Hayakawa, Costa, Foucart, & Keysar, 2016; Keysar, Hayakawa, & An, 2012; Oganian, Korn, & Heekeren, 2016; Pavlenko, 2012).

The framing effect describes the well-replicated observation that humans opt more often for the safe versus the risky alternative when choice options are framed as gains (e.g., number of people surviving) but show the reverse pattern when options are framed as losses (e.g., number of people dying)—despite identical outcome distributions in both frames (Kühberger, 1998; Tversky & Kahneman, 1981). To explain decision biases, such as framing effects and their modulations, scholars often draw on dual-process accounts of decision biases (Hayakawa et al., 2016; Kahneman, 2003; Pavlenko, 2012). According to the rather generic “two systems metaphor,” humans often process information in a quite superficial and heuristic way (system 1), which may explain why the framing of decision problems exerts such a consistent influence on risky choices. Deeper, more careful processing (system 2) prevails under specific conditions, for example, when cognitive control is enhanced. In line with this overall notion, we have interpreted our previous finding that language-switching reduces the framing effect, in terms of enhanced cognitive control and thus reduced intuitive processing induced by language-switching (Oganian et al., 2016). Specifically, participants in our earlier study read the general introduction to the experiment in either their native tongue

(German) or in a foreign language (English or French) and then answered framing questions in either their native or a foreign language. Framing effects were diminished when the presentation language switched between instructions and framing questions (from native to foreign or from foreign to native language) versus when the presentation language remained constant (only in the native or a foreign language).

Thus, motivated by previous reports that manipulating the format, i.e., the language of presentation, reduced framing effects (Costa et al., 2014; Keysar et al., 2012; Oganian et al., 2016), we investigated another manipulation of presentation format supposed to deepen cognitive processing. Specifically, we tested the effect of “cognitive disfluency”: Researchers have argued that presenting stimuli in hard-to-read fonts might trigger cognitive disfluency (Alter, 2013), as reflected in improved memory performance (Diemand-Yauman, Oppenheimer, & Vaughan, 2011) and less intuitive answers, e.g., in the cognitive reflection test (CRT) (Alter, Oppenheimer, Epley, & Eyre, 2007). Cognitive disfluency is theorized to induce deeper, more careful processing (i.e., engage system 2), because people reading a problem in a hard-to-read font misattribute the difficulty of reading to the difficulty of the problem itself (Alter, 2013). Consequently, cognitive disfluency induced by hard-to-read fonts seemed a likely candidate for a nonlanguage manipulation that could reduce framing effects in a similar way as switching between languages (Oganian et al., 2016).

Therefore, we tested whether hard-to-read fonts diminish framing effects as one of the most prominent and well-replicated decision biases. We deemed this question especially pertinent given that several recent reports could not, or only partly, replicate disfluency effects in experiments investigating learning and memory performance (Eitel, Kühl, Scheiter, & Gerjets, 2014; Kühl & Eitel, 2016; Meyer et al., 2015; Rummer, Schweppe, & Schwede, 2015; Thompson et al., 2013). To foreshadow our main results, we did not find conclusive evidence in Experiments 1 and 2. However, we found a significant albeit small effect in the preregistered Experiment 3, which used a larger sample size to specifically test the directional hypothesis that a hard-to-read font diminishes the framing effect.

Experiment 1

Materials and Methods

Participants A total of 160 participants (all university students) were recruited at Saarland University and tested in groups of up to 15 participants. We excluded two participants, because German was not their mother tongue (final sample: $N = 158$, mean age = 23.7 years, $SD = 4.48$; 83% female). Sample size was in a similar range as in previous studies on

framing effects and foreign language use (Costa et al., 2014; Keysar et al., 2012). Participants received monetary compensation or course credit. The study was conducted in accord with the Declaration of Helsinki. All participants gave written, informed consent.

Procedure and test materials Participants received the test material, including general instructions, either printed in a hard-to-read, disfluent font “Monotype Corsiva” (printed in gray; RGB 190, 190, 190; 12-point; $N = 78$) or in an easy-to-read, fluent font “Arial” (printed in black, 16-point; $N = 80$). The same fonts were used in a previous study on disfluency effects in memory (Diemand-Yauman et al., 2011). We asked an independent group of students to rate the readability of the two fonts in a within-participants design ($N = 32$, mean age = 24.5 years, $SD = 3.49$; 66% female). Ratings were obtained on a 7-point scale (1 = *not easy-to-read at all* and 7 = *very easy-to-read*). These ratings differed significantly between the easy- (mean $\pm SD$: 6.6 ± 0.9) and the hard-to-read conditions (4.2 ± 1.6 ; $t(31) = 7.5$, $p < 10^{-7}$).

In the main sample of Experiment 1, age and sex did not differ between the two groups: both $p > 0.4$. Each question was printed on one separate page, and participants had to answer each question within 1 min. The experimenter indicated the time to turn the pages.

All participants completed the following questions:

- (a) Four framing questions, including two with rather high and two with rather low emotional content. We included the additional factor emotional content to test its potential influence in a supplementary analysis since foreign language effects have been related to the often increased emotional distance conferred by a foreign language (Caldwell-Harris, 2014; Pavlenko, 2012). One of the two high-emotional content questions was the classic “Asian disease” scenario used in previous demonstrations of foreign language effects (Costa et al., 2014; Keysar et al., 2012; Oganian et al., 2016). The second high-emotional content question was an adapted version of the “Asian disease” scenario. The two low-emotional content questions (computer virus and damaged paintings) were adapted from a previous study (Oganian et al., 2016).
- (b) One framing control question in which the two options differed in expected value. In this control question, the majority of participants typically chooses the option with the higher expected value—regardless of the frame (we used the “unemployment” scenario, for details see Costa et al., 2014; Keysar et al., 2012; Oganian et al., 2016).
- (c) Seven classic logical questions, which prompt heuristic but incorrect answers (De Neys & Bonnefon, 2013; Meyer et al., 2015). These included three questions of the CRT with an open answer format and four questions

with a binary answer format (regarding base rate neglect; conjunction fallacy, ratio bias, syllogistic reasoning).

The two framing questions with high- and low-emotional content were balanced for order. The first two framing questions were presented before the seven logical questions and the second two framing questions afterwards, immediately followed by the framing control question. Participants were asked whether they knew any of the presented questions and were excluded from the analyses of the relevant items, if they affirmed this question.

A subgroup of the participants ($N = 59$) additionally performed a memory test in either hard-to-read or easy-to-read font. This memory test was modeled along a previous demonstration of disfluency effects (Diemand-Yauman et al., 2011). At the beginning of the experiment, participants received a table listing eight characteristics of each of three uncommon bird species. They had 2 min to learn this material and were asked to answer 12 questions about these characteristics at the end of the experiment (i.e., after a 15-min retention interval).

Analysis Framing tasks were analyzed with logistic regression models. We first tested the four framing questions individually (using the function *mnrfit* in MATLAB corresponding to the function *glm* in R with a binomial logit link function) and then all four combined in a hierarchical model (using the function *lmer* in R). Specifically, we compared a model that only included the between-participants factors frame (gain versus loss) with a model that included the factor frame and the between-participants factor font (easy- vs. hard-to-read) as well as their interaction [answer = gain-loss * easy-hard + (1 | participant) + (1 | question)]. Additionally, we explored possible effects of the between-participants factor order and the within-participants factor emotional content.

We used χ^2 statistics to compare models. We also provide Bayes factors based on the Bayesian Information Criterion in Table 1 (Jarosz & Wiley, 2014; Masson, 2011). The influence of the factor font was assessed with χ^2 tests in the seven logical questions and with a *t* test in the memory task.

Results and Discussion

After Bonferroni-correction for four tests, the framing effect was significant in three of the four framing questions (see Table 1 for percentages and Table 2 for full statistics). However, no effect of font or of the frame X font interaction reached significance after using Bonferroni correction to account for multiple comparisons.

Similarly, when we used a hierarchical model as an analysis approach that combines all four questions, we found a highly significant framing effect, $p < 10^{-4}$, but no effect of font. That is, the model, including the additional factor font and the frame X font interaction, did not explain significantly more variance than the simple model that only included the factor frame, $p = 0.115$.

Additionally, we found no significant influence of emotional content. A model, including this factor, along with the frame X emotional-content interaction did not explain significantly more variance than the simple model with the factor frame only, $p = 0.182$. However, a model, including the factor order as well as the frame X order interaction, significantly increased explained variance compared to the simple frame only model, $p = 0.005$. Accordingly, participants chose the sure option less often in the framing questions presented first, effect of order: $p = 0.001$, and exhibited a larger framing effect, effect of frame X order: $p = 0.021$. That is, when averaging over the type of framing question, a

Table 1 Framing and hard-to-read font (Experiments 1, 2, and 3): Percentages

Experiment	Framing question	% sure answer					
		ER Gain	Loss	HR Gain	Loss	ER Gain-Loss	HR Gain-Loss
1 (Laboratory)	Asian disease	71	46	55	45	25	11
	Adapted Asian disease	79	51	55	42	28	13
	Computer virus	58	40	55	26	18	29
	Damaged paintings	76	57	82	29	19	53
2 (Online)	Asian disease	54	46	59	31	8	28
	Adapted Asian disease	54	52	72	44	2	28
	Computer virus	52	34	56	34	18	22
	Damaged paintings	74	52	85	45	22	40
3 (Online)	Asian disease	68	36	62	42	32	20

ER easy-to-read font; HR hard-to-read-font

Table 2 Framing and hard-to-read font (Experiments 1, 2, and 3): effects of frame and font in logistic regressions

Experiment	Framing question	Statistics								Bayes factors	
		Intercept		Frame		Font		Frame X font		Model with main effect of frame versus interaction model	Model with main effects of frame and font versus interaction model
		Beta	Exact <i>p</i> value	Beta	Exact <i>p</i> value	Beta	Exact <i>p</i> value	Beta	Exact <i>p</i> value		
1 (Laboratory)	Asian disease	-0.18	0.282	0.37	0.027	0.18	0.282	-0.16	0.338	53.3	7.7
	Adapted Asian disease	-0.32	0.068	0.45	0.010	0.37	0.033	-0.18	0.292	9.3	7.0
	Computer virus	0.23	0.186	0.49	0.004	0.18	0.286	0.13	0.450	66.8	9.2
	Damaged paintings	-0.51	0.007	0.82	0.001	0.22	0.249	0.38	0.046	6.3	1.6
2 (Online)	Asian disease	0.10	0.441	0.37	0.004	0.11	0.411	0.21	0.107	54.8	4.5
	Adapted Asian disease	-0.24	0.060	0.32	0.014	-0.12	0.367	0.27	0.035	19.7	1.7
	Computer virus	0.26	0.043	0.41	0.002	-0.03	0.791	0.04	0.754	247.7	15.7
	Damaged paintings	-0.67	0.001	0.73	0.001	-0.01	0.457	0.24	0.097	59.9	4.0
3 (Online)	Asian disease	-0.08	0.305	0.53	<10 ⁻¹¹	0.00	0.967	-0.13	0.044	170.1	6.2

For overview, *p* values are given without Bonferroni-correction. For statistical inference in Experiments 1 and 2, we correct for the number of the four framing questions within each experiment. In the preregistered Experiment 3, we had an explicit directional hypothesis for the interaction effect and therefore performed a one-sided test (with the indicated *p* value divided by 2)

robust framing effect emerged in the first scenario presented, $\beta = 0.65$, $p < 0.001$, but it did not reach significance when the order was swapped, $\beta = 0.21$, $p = 0.212$. The control framing question indicated that participants overall read and understood the questions: 83% of the participants chose the response with the higher expected value and this percentage did not depend on font type, $p = 0.863$.

None of the seven logical questions or the memory task showed a significant effect of font (both with and without Bonferroni-correction; see Table 3 for statistics and accuracy). This pattern is in line with several recent reports (Eitel et al., 2014; Köhl & Eitel, 2016; Meyer et al., 2015; Rummer et al., 2015; Thompson et al., 2013) that did

not replicate the initial findings on disfluency effects (Alter et al., 2007; Diemand-Yauman et al., 2011).

Experiment 2

Given the absence of an effect of hard-to-read font on the framing effect in the first experiment, we performed an additional online study with a larger sample size to obtain higher statistical power. We focused on presenting the framing scenarios, because we were mainly interested in whether disfluency influences the framing effect and because null-effects of disfluency have already been reported for the CRT and for

Table 3 Seven logical questions and a memory test (Experiment 1)

Question	<i>N</i>		Proportion correct answer		Effect statistic	Exact <i>p</i> value
	Easy-to-read	Hard-to-read	Easy-to-read	Hard-to-read		
Bat and ball	66	66	0.26	0.26	0.00 ^a	1.000
Keyboard	63	63	0.32	0.44	2.15 ^a	0.199
Lake with flowers	68	66	0.56	0.44	191 ^a	0.226
Bill	64	59	0.39	0.46	0.57 ^a	0.471
Alex	66	62	0.53	0.53	0.00 ^a	1.000
Beans	73	71	0.92	0.94	0.37 ^a	0.745
Roses	68	70	0.69	0.70	0.01 ^a	1.000
Memory test ^b	27	32	0.55 (0.24)	0.55 (0.23)	0.15 ^b	0.882

^a $\chi^2(1)$ statistic

^b Mean proportion correct in 12 memory questions; *t* test with $df = 57$

memory tasks (Eitel et al., 2014; Kühn & Eitel, 2016; Meyer et al., 2015; Rummer et al., 2015; Thompson et al., 2013).

Materials and Methods

Participants We recruited participants via University mailing lists and social media and directed them to a German online survey system (<https://www.soscisurvey.de>), which we used previously for data collection (Oganian et al., 2016). Of the 293 participants who completed the questionnaire, 271 were included for analyses (final sample: mean age = 24.0 years, $SD = 4.34$; 75% female). Participants were included if a) German was their only mother tongue, b) age was between 18 and 60 years, c) they indicated their sex, d) they completed all framing questions, and e) comments at the end of questionnaire indicated no knowledge of framing or disfluency effects.

Procedure and test materials We opted for “Impact” printed in gray as the hard-to-read font (12-point), because this was another font used by a previous report on disfluency effects (see Studies 7 and 13 in Meyer et al., 2015). The fluent font was “Arial” (printed in black, 14-point). Participants were randomly allocated to conditions (easy-to-read font: $N = 131$; hard-to-read font: $N = 140$) and answered the same four framing questions and the same control question as in Experiment 1. Ratings by the sample of Experiment 2 confirmed that the readability of the easy-to-read font (mean $\pm SD$: 6.3 ± 1.2) was significantly higher than the readability of the hard-to-read condition (3.2 ± 2.1 ; $t(254) = 14.0$, $p < 10^{-32}$).

We were concerned that participants zoomed the screen or aborted the experiment when faced with hard-to-read text. We therefore explicitly wrote in the instructions that the font may be difficult to read in some cases and that they should not zoom or abort because of this. In the final sample, the percentages of participants in the easy- and the hard-to-read font conditions were balanced (with 52% of participants in the hard-to-read font condition).

Results and Discussion

The framing effect was significant for all four questions individually (see Table 1 for percentages and Table 2 for statistics) and for the combined analysis within a hierarchical model, $p < 10^{-7}$. Again, neither the main effects of font nor the frame X font interactions reached significance in the individual questions: all Bonferroni-corrected $p > 0.1$. Model comparison showed that neither font, $p = 0.083$, nor emotional-content, $p = 0.228$, explained additional variance beyond the influence of frame. Order showed the same influence as in the laboratory study, model comparison: $p < 10^{-7}$; effect of order: $p < 10^{-7}$; effect of frame X order: $p = 0.008$. That is, the framing effect was strong in the first scenario presented, $\beta = 0.54$, $p < 10^{-4}$, and relatively reduced for the swapped order, $\beta =$

0.28 , $p = 0.029$. We found no significant interaction effects with the time spent on the framing questions: all $p > 0.1$. In the control question, the majority of participants (67%) chose the option with the higher expected value and this did not depend on font type, $p = 0.277$. Converging with Experiment 1, Experiment 2 did not provide conclusive evidence for a reduction of the framing effect by hard-to-read fonts.

Experiment 3

Following recommendations by anonymous reviewers, we conducted Experiment 3 as an extended replication of Experiment 2 with the following key differences. First, we conducted a formal power analysis for detecting a reduction of the framing effect (see below). Second, in contrast to Experiment 2, we refrained from warning participants about possibly hard-to-read fonts because such warnings may likely offset potential effects of hard-to-read fonts, i.e., participants tend to discount disfluency when they are made aware that it results from an irrelevant source (Oppenheimer & Frank, 2008; Schwarz, 2004). Third, to provide high transparency, we preregistered all details of Experiment 3 on the online platform of the Open Science Framework (OSF) before starting data collection (<https://osf.io/aqnmq/register/565fb3678c5e4a66b5582f67>). Fourth, we focused on the best-validated framing question, the classic “Asian disease” scenario (along with the control scenario). Presenting only one scenario precludes influences of presentation order. Fifth, to explore individual differences in need for cognition and statistical abilities, we administered the German versions of the Need for Cognition Scale (NFC) (Bless, Wänke, Bohner, Fellhauer, & Schwarz, 1994) and the Berlin Numeracy Test (BNT) (Cokely, Galesic, Schulz, Ghazal, & Garcia-Retamero, 2012).

Materials and Methods

Power analysis We used the program G*Power (<http://www.gpower.hhu.de>) (Faul, Erdfelder, Buchner, & Lang, 2009). Our goal was to obtain 0.85 power at the standard 0.05 alpha error probability. We tested the directional hypothesis that a hard-to-read font reduces the framing effect and therefore used a one-tailed test for the interaction effect in a logistic regression (under the large sample approximation). The directional hypothesis was motivated first by theoretical notions and initial evidence that hard-to-read fonts might elicit deeper (i.e., less biased) reasoning (Alter, 2013), and second by previous demonstrations that language-switching reduced the framing effect (Oganian et al., 2016). The expected effect should entail at least a reduction by 0.1 in the portion of sure answers in the

gain versus the loss frame. The required sample size was calculated as 712 participants.

Participants As in Experiment 2, we used the online system SoSci Survey, which offers the opportunity to recruit participants of an associated panel. Participants of this panel participate out of interest and receive no monetary reimbursement. In accordance with the preregistration, we checked after 7 days and then again in intervals of 2 days whether the number of participants meeting the inclusion criteria (which were the same as in Experiment 2) was above 712. After 9 days, 835 participants had started the questionnaire and 732 participants (mean age = 39.0 years, $SD = 11.93$; 63% female) met the inclusion criteria.

Procedure and test materials As in Experiment 1, we used “Monotype Corsiva” (printed in gray; RGB 190, 190, 190, 9-point) as hard-to-read font. Because this font is not installed on many systems, we implemented all relevant texts as pictures. The fluent font was “Arial” (printed in black, 9-point). Participants were randomly allocated to conditions and drop-outs were balanced (final sample: easy-to-read font: $N = 378$; hard-to-read font: $N = 354$). Participants answered the classic “Asian disease” scenario and the control scenario, followed by questions about demography, readability, and prior knowledge of the scenarios. As expected, participants rated the easy-to-read font (mean $\pm SD$: 6.1 ± 1.3) as easier than the hard-to-read font (3.4 ± 1.6 ; $t(729) = 25.1$, $p < 10^{-99}$). Participants in the easy-to-read font condition indicated less often that they zoomed into the text (proportion indicating zooming = 0.05) than those in the hard-to-read font condition (0.12; $\chi^2(1) = 9.7$, $p < 0.005$; zooming was not an exclusion criterion). Afterwards, participants were asked to complete the 33-items NFC scale and the 7-items BNT questionnaire. For the main analyses of the framing question, we also included participants who did not complete these two scales (as detailed in the preregistration). The exploratory analyses regarding NFC and BNT rely on smaller numbers of participants than the main analyses (NFC: $N = 673$; BNT: $N = 575$). For NFC, we calculated mean responses on a 7-point scale ($-3 = \textit{not applicable at all}$ and $+3 = \textit{very applicable}$; mean $\pm SD$: 1.1 ± 0.8), and for BNT, we assessed the number of correct answers (4.4 ± 1.7).

Results and Discussion

The registered, confirmatory hypothesis testing indicated that the interaction effect of frame X font was significant, $p = 0.044$ one-sided, indicating that the hard-to-read font reduced the framing effect (Tables 1 and 2). The main effect of frame was highly significant, but the main effect of font did not reach significance.

In exploratory follow-up analyses, we tested for potential influences of NFC or BNT but found no significant interaction effects in our sample: all $p > 0.1$. We concede that these analyses may suffer from limited power to detect a potentially significant triple interaction of NFC or BNT X frame X font. Another exploratory analysis did not provide evidence for an influence of time spent on answering the framing question: all $p > 0.3$. Most participants (73%) chose the option with the higher expected value in the control question, and the main effect of font was not significant: $p = 0.832$.

Combined analyses of Experiments 1, 2, and 3 In an exploratory meta-analysis, we pooled data for the “Asian disease” scenario across all three experiments. While the main effect of frame was highly significant, $p < 10^{-14}$, the main effect of frame and the frame X font interaction did not reach significance: all $p > 0.4$. While the absence of a significant interaction effect in the combined analyses may be an indicator of the overall noise level of using hard-to-read fonts, an alternative conjecture points to a crucial difference between Experiment 2 versus Experiments 1 and 3. In Experiment 2, we alerted participants in the instructions about the use of a hard-to-read font, which may have undermined potential disfluency effects. Indeed, when we included the presence of the warning about a hard-to-read font as an additional factor, we found a significant triple interaction of warning X frame X font, $p = 0.018$, in the pooled data, which suggests that such warnings may abolish or even reverse the effects of hard-to-read fonts (see numerical values of the “Asian disease” scenario in Table 1).

General Discussion

Recent findings of foreign language effects on decision biases, most notably framing effects, have been explained in part by deeper cognitive processing (Costa et al., 2014; Geipel, Hadjichristidis, & Surian, 2016; Hayakawa et al., 2016; Keysar et al., 2012; Oganian et al., 2016). Hard-to-read fonts also have been suggested to deepen cognitive processing by triggering cognitive disfluency. We therefore hypothesized that presenting participants with framing effect tasks in hard-to-read fonts also may reduce the framing effect (Alter, 2013). To our knowledge, this is the first report to test for an impact of hard-to-read fonts on framing.

We did not find any significant influence of hard-to-read fonts on the framing effect in Experiments 1 and 2. Only specifically testing for a directional effect in Experiment 3 with a large sample size ($N = 732$) provided evidence for a weak modulating influence such that the hard-to-read font reduced the framing effect in the classic “Asian disease” scenario (i.e., the difference in the framing effect between the two

conditions was 12% points; Table 1). Even larger sample sizes will be needed to obtain a clear-cut picture of how individual differences might modulate this rather small effect. Exploratory analyses compiling data from all three experiments did not provide clear-cut evidence for a straightforward effect of hard-to-read fonts. Instead, these analyses suggest that the influence of hard-to-read fonts might depend on whether participants receive prior information regarding the readability of upcoming fonts. This is in line with the theoretical notion that disfluency only becomes relevant when participants confuse the difficulty engendered by the hard-to-read font with the difficulty engendered by the decision problem per se but not when they can attribute the difficulty to an extraneous source, i.e., a hard-to-read font (Oppenheimer & Frank, 2008; Schwarz, 2004). Taken together, the large sample size needed to detect an effect of a hard-to-read font in Experiment 3 and the potential influence of explicit information about such font manipulations limit their relevance for real-world applications.

We would like to mention the additional finding that presentation order of the scenarios influenced the framing effect such that it was absent or diminished in the later presented scenarios. This may indicate that experience with framing scenarios leads to more careful processing within the course of an experiment. Because effects of presentation order consistently emerged in Experiments 1 and 2 with smaller sample sizes suggests that such order effects are stronger (and may thus potentially overshadow) disfluency effects.

In Experiment 1, we did not replicate two types of disfluency effects reported previously. First, the initial suggestion of disfluency effects included the CRT and syllogistic reasoning questions (Studies 1 and 4 in Alter et al., 2007). Our null findings on the CRT are in line with a recent high-powered meta-analysis (with diverse samples of more than 7,000 participants) that does not support disfluency effects on the CRT (Meyer et al., 2015). Second, our findings did not replicate that hard-to-read fonts lead to better memory performance on knowledge questions (Study 1 in Diemand-Yauman et al., 2011). Our sample size ($N = 59$) more than doubled the sample size of the initial study ($N = 27$), and participants in both studies were university students (although entry requirements are more restrictive at the top-ranked US university at which the study by Diemand-Yauman et al., 2011 was conducted). Recent evidence suggests moderators for influences of hard-to-read fonts on memory performance. Most notably, more robust effects are observed with longer delays between learning and test (Seufert, Wagner, & Westphal, 2017; Weissgerber & Reinhard, 2017). Studies with rather short delays, such as the present study, usually do not find the effect (Eitel & Köhl, 2016; Magreehan, Serra, Schwartz, & Narciss, 2016; Rummer et al., 2015).

The need for replications and for transparent, confirmatory hypothesis testing has come into the spotlight (Camerer et al.,

2016; Open Science Collaboration, 2015). We therefore report all of the experiments that we conducted; the initial experiments indicated null effects and the preregistered experiment provided evidence for a weak reduction of the framing effect under a condition of disfluency.

It is intriguing that a simple manipulation of font readability could diminish a prominent decision bias, such as the framing effect. Nevertheless, given the large sample needed to detect this small modulating effect, we argue for a nuanced and careful consideration of the overall robustness, domain generality, and potential real-world applications of disfluency manipulations (Dunlosky & Mueller, 2016).

Acknowledgments The authors thank Alica Thissen and Christian Hahn for their help with data collection and discussion. They thank Dr. Dominik Leiner, the administrator of SoSci Survey, without whom Experiment 3 would not have been possible. Funding was provided by a research grant from Saarland University. CWK was supported by the SFB TRR 169 during the final stages of manuscript preparation. The funders had no role in the study.

Compliance with ethical standards

Declaration of interest No conflict of interest.

References

- Alter, A. L. (2013). The benefits of cognitive disfluency. *Current Directions in Psychological Science*, 22(6), 437–442. <https://doi.org/10.1177/0963721413498894>
- 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(4), 569–576. <https://doi.org/10.1037/0096-3445.136.4.569>
- Bless, H., Wänke, M., Bohner, G., Fellhauer, R. F., & Schwarz, N. (1994). Need for Cognition: Eine Skala zur Erfassung von Engagement und Freude bei Denkaufgaben. *Zeitschrift Für Sozialpsychologie*, (1994), 147–154. <https://doi.org/10.1111/j.1467-839X.2008.00269.x>
- Caldwell-Harris, C. L. (2014). Emotionality differences between a native and foreign language: Theoretical implications. *Frontiers in Psychology*, 5, 1–4. <https://doi.org/10.3389/fpsyg.2014.01055>
- Camerer, C. F., Dreber, A., Forsell, E., Ho, T., Huber, J., Kirchler, M., ... Wu, H. (2016). Evaluating replicability of laboratory experiments in economics. *Science*, 351(6280), 1433–1436. <https://doi.org/10.1126/science.aaf0918>
- Cokely, E. T., Galesic, M., Schulz, E., Ghazal, S., & Garcia-Retamero, R. (2012). Measuring Risk Literacy: The Berlin Numeracy Test. *Judgment and Decision Making*, 7(1), 25–47.
- Costa, A., Foucart, A., Amon, I., Aparici, M., & Apesteguia, J. (2014). “Piensa” twice: On the foreign language effect in decision making. *Cognition*, 130(2), 236–254. <https://doi.org/10.1016/j.cognition.2013.11.010>
- De Neys, W., & Bonnefon, J.-F. (2013). The “whys” and “whens” of individual differences in thinking biases. *Trends in Cognitive Sciences*, 17(4), 172–178. <https://doi.org/10.1016/j.tics.2013.02.001>
- 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(1), 114–118. <https://doi.org/10.1016/j.cognition.2010.09.012>
- Dunlosky, J., & Mueller, M. L. (2016). Recommendations for exploring the disfluency hypothesis for establishing whether perceptually degrading materials impacts performance. *Metacognition and Learning*, 11(1), 123–131. <https://doi.org/10.1007/s11409-016-9155-9>
- Eitel, A., & Kühn, T. (2016). Effects of disfluency and test expectancy on learning with text. *Metacognition and Learning*, 11, 107–121.
- Eitel, A., Kühn, T. I. M., Scheiter, K., & Gerjets, P. (2014). Disfluency meets cognitive load in multimedia learning: Does harder-to-read mean better-to-understand? *Applied Cognitive Psychology*, 28, 488–501. <https://doi.org/10.1002/acp.3004>
- Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, 41(4), 1149–1160. <https://doi.org/10.3758/BRM.41.4.1149>
- Gao, S., Zika, O., Rogers, R. D., & Thierry, G. (2015). Second language feedback abolishes the “hot hand” effect during even-probability gambling. *Journal of Neuroscience*, 35(15), 5983–5989. <https://doi.org/10.1523/JNEUROSCI.3622-14.2015>
- Geipel, J., Hadjichristidis, C., & Surian, L. (2016). Foreign language affects the contribution of intentions and outcomes to moral judgment. *Cognition*, 154, 34–39. <https://doi.org/10.1016/j.cognition.2016.05.010>
- Gigerenzer, G., & Gaissmaier, W. (2011). Heuristic decision making. *Annual Review of Psychology*, 62, 451–82. <https://doi.org/10.1146/annurev-psych-120709-145346>
- Hayakawa, S., Costa, A., Foucart, A., & Keysar, B. (2016). Using a foreign language changes our choices. *Trends in Cognitive Sciences*, 20:791-793. doi:<https://doi.org/10.1016/j.tics.2016.08.004>
- Jarosz, A. F., & Wiley, J. (2014). What Are the Odds? A Practical Guide to Computing and Reporting Bayes Factors. *The Journal of Problem Solving*, 7(1). doi:<https://doi.org/10.7771/1932-6246.1167>
- Kahneman, D. (2003). A perspective on judgment and choice: Mapping bounded rationality. *The American Psychologist*, 58(9), 697–720. <https://doi.org/10.1037/0003-066X.58.9.697>
- Keysar, B., Hayakawa, S. L., & An, S. G. (2012). The foreign-language effect: Thinking in a foreign tongue reduces decision biases. *Psychological Science*, 23(6), 661–668. <https://doi.org/10.1177/0956797611432178>
- Kühberger, A. (1998). The influence of framing on risky decisions: A meta-analysis. *Organizational Behavior and Human Decision Processes*, 75(1), 23–55. <https://doi.org/10.1006/obhd.1998.2781>
- Kühn, T., & Eitel, A. (2016). Effects of disfluency on cognitive and metacognitive processes and outcomes. *Metacognition and Learning*, 11(1), 1–13.
- 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. <https://doi.org/10.1037/a0034407>
- Masson, M. E. J. (2011). A tutorial on a practical Bayesian alternative to null-hypothesis significance testing. *Behavior Research Methods*, 43(3), 679–690. <https://doi.org/10.3758/s13428-010-0049-5>
- Meyer, A., Frederick, S., Burnham, T. C., Guerva 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(2), e16–30. <https://doi.org/10.1037/xge0000049>
- Oganian, Y., Korn, C. W., & Heekeren, H. R. (2016). Language switching—but not foreign language use per se—reduces the framing effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 42(1), 140–148. <https://doi.org/10.1037/xlm0000161>
- Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, 349(6251). <https://doi.org/10.1126/science.aac4716>
- Oppenheimer, D. M., & Frank, M. C. (2008). A rose in any other font would not smell as sweet: Effects of perceptual fluency on categorization. *Cognition*, 106(3), 1178–1194. <https://doi.org/10.1016/j.cognition.2007.05.010>
- Pavlenko, A. (2012). Affective processing in bilingual speakers: Disembodied cognition? *International Journal of Psychology*, 47(6), 405–28. <https://doi.org/10.1080/00207594.2012.743665>
- Rummer, R., Schweppe, J., & Schwede, A. (2015). Fortune is fickle: Null-effects of disfluency on learning outcomes. *Metacognition and Learning*, 11(1), 57–77. <https://doi.org/10.1007/s11409-015-9151-5>
- Schwarz, N. (2004). Metacognitive Experiences in Consumer Judgment and Decision Making. *Journal of Consumer Psychology*, 14(4), 332–348. https://doi.org/10.1207/s15327663jcp1404_2
- Seufert, T., Wagner, F., & Westphal, J. (2017). The effects of different levels of disfluency on learning outcomes and cognitive load. *Instructional Science*, 45, 221–238.
- Thompson, V. A., Ackerman, R., Sidi, Y., Ball, L. J., Pennycook, G., & Prowse Turner, J. A. (2013). The role of answer fluency and perceptual fluency in the monitoring and control of reasoning: Reply to Alter, Oppenheimer, and Epley (2013). *Cognition*, 128(2), 256–258. <https://doi.org/10.1016/j.cognition.2013.03.003>
- Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, 211(4481), 453–458.
- Weissgerber, S. C., & Reinhard, M.-A. (2017). Is disfluency desirable for learning? *Learning and Instruction*, 49, 199–217.