Research Article

Psychological Science
2017, Vol. 28(4) 427–436

DOI: 10.1177/0956797616685770
www.psychologicalscience.org/PS

If It's Difficult to Pronounce, It Might Not Be Risky: The Effect of Fluency on Judgment of Risk Does Not Generalize to New Stimuli

Štěpán Bahník[1,2] and Marek A. Vranka[3,4]

[1]Department of Management, Faculty of Business Administration, University of Economics, Prague; [2]Department of Psychology, Faculty of Human Sciences, University of Würzburg; [3]Department of Psychology, Faculty of Arts, Charles University; and [4]Department of Marketing Communication and Public Relations, Faculty of Social Sciences, Charles University

Abstract

Processing fluency is used as a basis for various types of judgment. For example, previous research has shown that people judge food additives with names that are more difficult to pronounce (i.e., that are disfluent) to be more harmful. We explored the possibility that the association between disfluency and perceived harmfulness might be in the opposite direction for some categories of stimuli. Although we found some support for this hypothesis, an improved analysis and further studies indicated that the effect was strongly dependent on the stimuli used. We then used stimulus sampling and showed that the original association between fluency and perceived safety was not replicable with the newly constructed stimuli. We found the association between fluency and perceived safety using the newly constructed stimuli in a final study, but only when pronounceability was confounded with word length. The results cast doubt on generalizability of the association between pronounceability and perceived safety and underscore the importance of treating stimulus as a random factor.

Keywords

fluency, risk, judgment, stimulus sampling, replication, open data, open materials, preregistered

Received 9/28/15; Revision accepted 12/2/16

The effects of processing fluency (i.e., a metacognitive feeling of ease of processing) have been shown in many different domains. Stimuli that are processed more fluently are usually judged to be more valuable (Alter & Oppenheimer, 2008), more likeable (Reber, Winkielman, & Schwarz, 1998), more frequent (Tversky & Kahneman, 1973), and more likely to be true (McGlone & Tofighbakhsh, 1999) than stimuli that are less fluent. One of the studies of judgmental effects of processing fluency showed that people perceive food additives and amusement-park rides with names that are hard to pronounce to be riskier (Song & Schwarz, 2009). People tend to avoid risk; therefore, a possible explanation for the association between processing fluency and perceived safety is that people encounter safe objects more often, which increases their familiarity and leads to more fluent processing. People thus form naive theories (Alter & Oppenheimer, 2009) regarding how fluency is associated with safety, which they use in judgment. In line with this explanation, items that are hard to pronounce are also judged to be more novel, and that novelty mediates the effect of pronounceability on judgment of risk (Song & Schwarz, 2009).

Some studies have shown that fluency can have different effects on judgment depending on context (Galak & Nelson, 2011; Pocheptsova, Labroo, & Dhar, 2010). These findings suggest that naive theories about the association

Corresponding Author:
Štěpán Bahník, Faculty of Business Administration, University of Economics, Prague, náměstí Winstona Churchilla 4, Prague 130 67, Czech Republic
E-mail: bahniks@seznam.cz

between fluency and a judged attribute may differ for different categories of objects. In our first five studies, we attempted to build on these findings and explored the hypothesis that the association between fluency and perceived safety is context dependent. That is, we tested whether people associate fluency with risk under some circumstances. We expected that this might be the case for categories whose more frequently encountered exemplars are associated with greater risk.

Following the methods used by Song and Schwarz (2009), the initial studies (Studies 1–4) treated stimulus as a fixed factor. Consequently, we found some initial support for the context dependence of the association between fluency and perceived safety, but the results were highly variable and seemed to depend less on categories of objects and more on the particular stimuli used. This led us to explore generalizability of Song and Schwarz's findings to newly constructed stimuli in three further studies (Studies 5–7), in which we used randomly created and sampled stimuli, which ensured that there was no possibility of bias in their selection. In Studies 5 and 6, we replicated the original association between fluency and perceived safety using the original stimuli created by Song and Schwarz. However, we did not replicate the effect with newly constructed stimuli. We used materials from a different experiment in Song and Schwarz in Study 7, and it showed a different pattern of results—we found the association between pronounceability and perceived safety with newly constructed stimuli, but not with the original stimuli used by Song and Schwarz. However, pronounceability was confounded with word length, and the association between pronounceability and perceived safety disappeared after we controlled for word length. Therefore, in line with Studies 5 and 6, even Study 7 cast doubt on the existence of a generalizable association between pronounceability and perceived risk and illustrated the importance of stimuli sampling.

Studies 1 Through 4

In the first four studies, we examined the hypothesis that the association between processing fluency and perceived safety may be dependent on the category of an evaluated object. Although familiarity may be a valid cue of safety for some categories of objects, this may not be the case for other categories. For example, people encounter names of more dangerous criminals in the news more often than names of less dangerous criminals. Likewise, cities in a war zone are more likely to be mentioned in the news if fighting has occurred there than if there has been no fighting. People may therefore learn the opposite association between fluency and risk for these categories of objects. They might then use it when judging risk just as they use the more common association between fluency and safety for other categories of objects.

Method

Participants. Before exclusion,[1] 89 Czech undergraduates participated in Study 1; 181 Czech university students participated in Study 2; 198 German participants took part in Study 3; and 607 workers from Amazon's Mechanical Turk (MTurk) participated in Study 4.[2]

Procedure. All four studies shared the same general procedure, adopted from Song and Schwarz (2009). Participants were given one or two scenarios describing a hypothetical situation in which they encountered 10 exemplars of a certain category. Then, they judged dangerousness of the exemplars on a 7-point scale (1 = *very safe*, 7 = *very dangerous*). The exemplars were introduced only by their names, and participants had no additional information about them. All names were 12 letters long and were selected such that half were relatively easy to pronounce (e.g., Allotoneline, Magnalroxate) and the other half were hard to pronounce (e.g., Ribozoxtlitp, Nxungzictrop).

We used a total of four hypothetical situations and categories of stimuli.[3] In the "food additives" scenario (adopted from Song & Schwarz, 2009), participants were told to imagine reading names of food additives on a food label. The "cities in a war zone" and "criminals" scenarios were created such that we expected that people would judge items that were easier to pronounce to be more dangerous than those that were hard to pronounce. In the "cities in a war zone" scenario, participants were told to imagine traveling through war-stricken Syria and to judge the dangerousness of cities they traveled through. In the "criminals" scenario, participants were told to rate the dangerousness of criminals considered for an amnesty. The wording and stimuli of the "beach resorts" scenario were the same as for the "cities in a war zone" scenario, but "war-stricken Syria" was replaced by "Turkish Riviera," and beach resorts were rated instead of cities. The names of food additives used in the present studies were created by Song and Schwarz, and the names of criminals were surnames from languages other than the participants' native languages (six Dutch, two Finnish, two English). Finally, the names of cities and beach resorts were actually the names of small cities in Lebanon.

Results

The results of the four studies are shown in Figure 1. We did not replicate the results of Song and Schwarz (2009) in Study 1; however, using their original materials, we replicated their results in Studies 2 and 3. Although the results of Study 2 suggested that the effect of pronounceability on judgment of riskiness might be reversed for some categories of objects (Fig. 1, left), we did not obtain the same effect in Studies 3 and 4. In fact, we obtained the effect in the original direction (perceived safety was

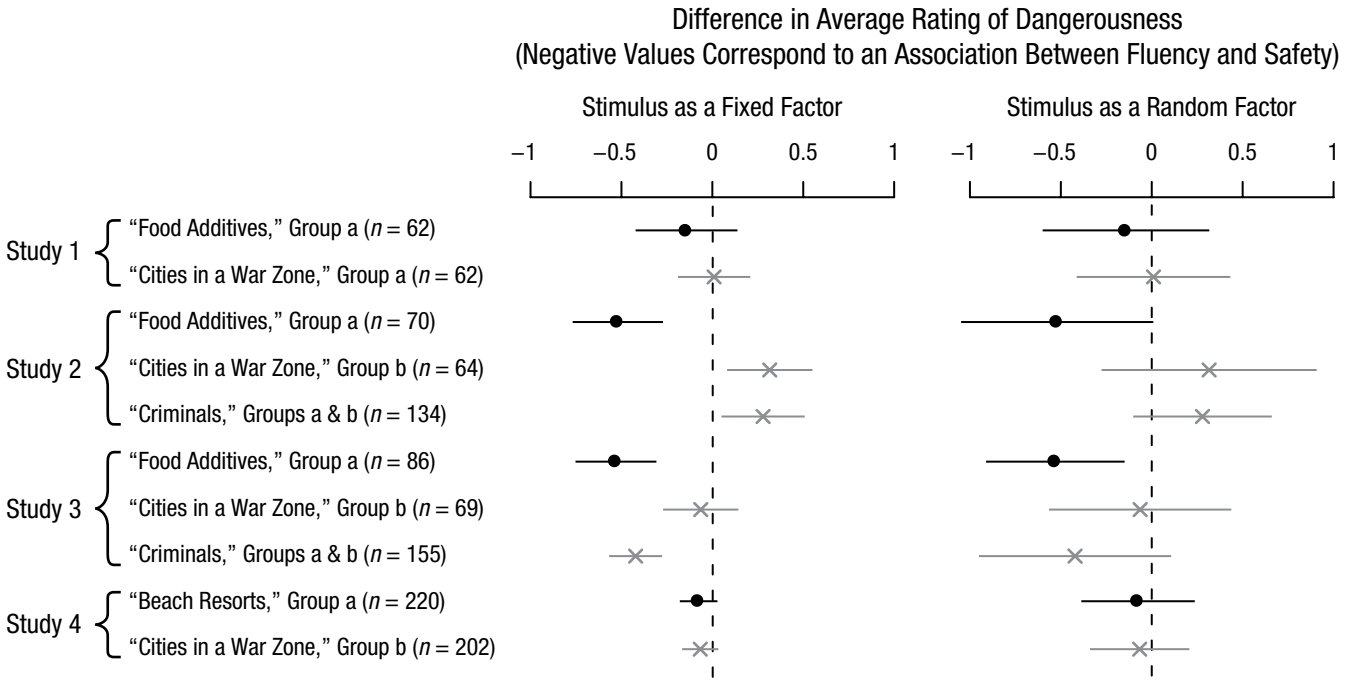

Fig. 1. Results of Studies 1 through 4, by experimental group (a, b) and hypothetical scenario. Points and crosses represent the mean differences in dangerousness ratings between easy- and hard-to-pronounce items (food additives, cities, criminals, or beach resorts). Error bars represent 95% confidence intervals for the estimated differences. Negative values correspond to an association between fluency and safety (i.e., easy-to-pronounce items were judged to be less dangerous than hard-to-pronounce items, the original association observed by Song and Schwarz, 2009). Positive values correspond to an association between fluency and risk (i.e., easy-to-pronounce items were judged to be more dangerous than hard-to-pronounce items). The left graph shows results for stimulus treated as a fixed factor, and the right graph shows results for stimulus treated as a random factor. Black points are used for scenarios in which we expected the original association to be replicated; gray crosses are used for scenarios in which we expected the opposite association.

greater for easy-to-pronounce items than for hard-to-pronounce items; Study 3) even when using the same scenario in which we observed the reversed effect in Study 2. We changed two items in the "criminals" scenario between Studies 2 and 3, so one possible explanation for the reversal is that the effect may depend on the particular items used. Furthermore, as in Song and Schwarz, in our analyses in Studies 1 through 4, we incorrectly treated stimulus as a fixed factor, which precludes the possibility of generalizing the results of these studies. When the analysis was conducted correctly, treating stimulus as a random factor (Fig. 1, right), Song and Schwarz's results were still replicated, but no effect in the opposite direction remained significant. In Studies 5 to 7, we tried to clarify these results and overcome shortcomings of the first four studies by randomly creating and sampling new stimuli.

Study 5

The results of the first four studies indicated that fluency effects can strongly depend on the particular items used. Therefore, in Study 5, in addition to studying the possibility of reversal of the association between fluency and perceived safety, we directly compared results obtained with the items originally used by Song and Schwarz (2009) with results from similar newly constructed items.

Method

We recruited participants from MTurk by posting a human intelligence task for 600 workers; ultimately, 616 MTurk workers participated in the study. We excluded 44 participants who had more than one missing data point or who used one rating more than seven times, in accordance with our preregistered exclusion criteria.

To explore the influence of particular items, we used the 10 items from the original "food additives" scenario and added 50 new items (e.g., Enzalutmmide, Griseofplvin). Each participant was given 10 randomly selected items from among the 60 items. The new items were created by taking existing 12-letter medication names, randomly changing one letter[4] in the names, and then removing names that sounded too similar to well-known substances (e.g., Tedtosterone). We used a new scenario in which participants imagined that they were members of a team of scientists searching through the archives of a laboratory that had researched either poisons or medicines (depending on participant's condition). The

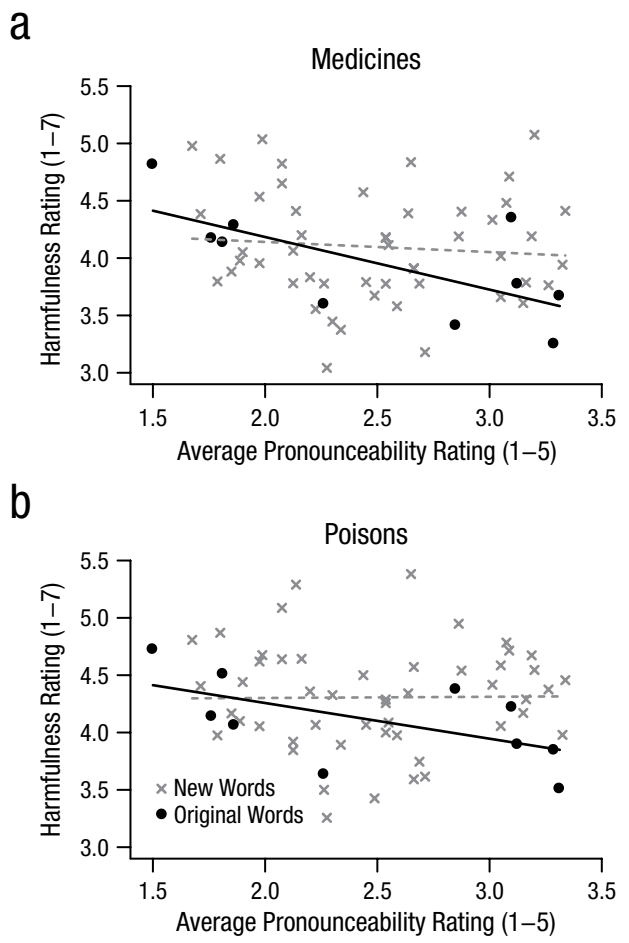

Fig. 2. Results from Study 5. The scatterplots (with best-fitting regression lines) show the relationship between pronounceability rating and harmfulness rating for (a) medicines and (b) poisons.

participants were asked to judge the substances' harmfulness on a scale from 1 (*harmless*) to 7 (*very harmful*) on the basis of their names.

Pronounceability of names used in the study was rated by an independent sample of 80 MTurk workers on a scale from 1 (*easy to pronounce*) to 5 (*hard to pronounce*). To make it easier to compare the results with those of the first four studies, we reversed the average ratings such that the pronounceability score was higher for easier-to-pronounce items. A negative slope for pronounceability therefore indicates the association between fluency and safety. We also centered the scores by subtracting the mean of pronounceability ratings from all values.

Results

A mixed-effects model with harmfulness rating as the dependent variable showed that harder-to-pronounce items were not judged to be significantly more harmful, $t(65.4) = -0.91$, $p = .37$, $b = -0.12$, 95% confidence interval (CI) = [−0.38, 0.14]. Even though neither of the effects was significant, the effect of pronounceability on harmfulness rating was weaker for poisons than for medicines, $t(5435.7) = 1.68$, $p = .09$, $b = 0.13$, 95% CI = [−0.02, 0.28], and was stronger for items used in the original study by Song and Schwarz (2009) than for the newly constructed items, $t(56.2) = -1.45$, $p = .15$, $b = -0.35$, 95% CI = [−0.81, 0.12]. The intercept was higher for poisons than for medicines, $t(568.0) = 3.91$, $p < .001$, $b = 0.21$, 95% CI = [0.11, 0.32], and it did not differ between the original items and the newly constructed items, $t(56.0) = -1.20$, $p = .24$, $b = -0.18$, 95% CI = [−0.47, 0.11].[5]

Although the interaction between item source and pronounceability was not significant, we were interested in whether the effect of pronounceability on judgment of harmfulness might differ between the original items and the newly constructed items. Therefore, we conducted separate analyses for the original items and the newly constructed items. The results showed a significant pronounceability effect for the original items, $t(12.8) = -3.02$, $p = .01$, $b = -0.48$, 95% CI = [−0.79, −0.17], but not for the newly constructed items, $t(58.8) = -0.79$, $p = .43$, $b = -0.11$, 95% CI = [−0.39, 0.16]. The interaction between category of judgment and pronounceability was not significant for either of the two item sources. Figure 2 shows the effect of pronounceability on harmfulness ratings at the item level.

Discussion

Although we found that the effect of disfluency on judgment of harmfulness was somewhat stronger for medicines than for poisons, the interaction was not significant and the effect was not in the opposite direction for poisons. The hypothesis that the effect of fluency on judgment of harmfulness might be reversed for some categories of stimuli was therefore not supported even in a study using random sampling of stimuli.

The results suggest that the effect of pronounceability on judgment of harmfulness might be limited only to the original items used by Song and Schwarz (2009). As in Studies 2 and 3, we replicated the effect for the original items, even when using different scenarios. However, we did not find a significant effect for the newly constructed items. Because the interaction of item source and pronounceability was not significant, we conducted an additional study that explored this result further.

Study 6

In Study 6, we again used the original "food additives" scenario along with the original items to compare results

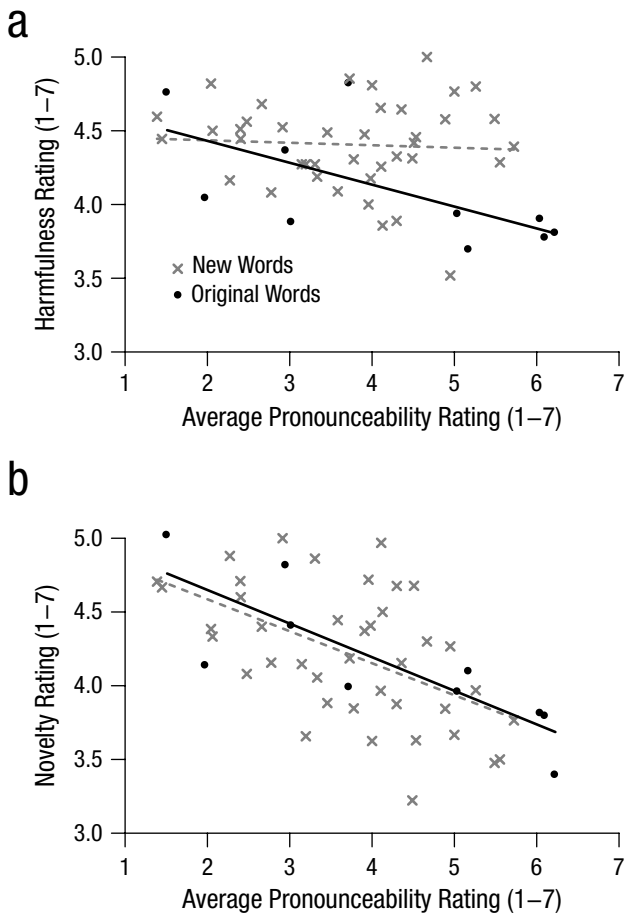

Fig. 3. Results from Study 6. The scatterplots (with best-fitting regression lines) show the relationship between (a) harmfulness ratings and pronounceability ratings and between (b) novelty ratings and pronounceability ratings.

of a direct replication of Song and Schwarz (2009) and a replication using the same sort of materials but using newly constructed items.

Method

We recruited 200 Czech participants for the study. We excluded 14 participants who used the same rating 12 or more times out of 15 possible opportunities, in accordance with our preregistered exclusion criterion. The experiment was conducted in a lab using a custom-written Python program using groups of up to 13 people as a part of a larger set of unrelated studies.

Participants were given the "food additives" scenario adopted from Song and Schwarz (2009) and judged the harmfulness of 15 food additives on the 7-point harmfulness scale used in Study 5. We used 10 items from the original study and additionally created 40 new items. The new items were created using the list of 12-letter medicine names from Study 5. We randomly substituted one letter in each word, randomly selected a 10-letter continuous string from this new name, and appended a 2-letter suffix from a list of suffixes that we based on a list of Czech names of food additives. The created names varied in pronounceability and were similar to the original names used in Song and Schwarz (2009).[6]

All participants rated the harmfulness of 10 newly constructed food-additive items and 5 original food-additive items. Next, participants were divided into two groups. The first group (n = 105; 100 after exclusion) rated the novelty of 10 newly constructed food-additive items and the 5 remaining original food-additive items on a 7-point scale from 1 (*very old*) to 7 (*very new*). The second group (n = 95, 86 after exclusion) rated the pronounceability of 20 newly constructed items and the 5 remaining original items on a 7-point scale from 1 (*hard to pronounce*) to 7 (*easy to pronounce*). All items were randomly selected, and participants did not rate the same item twice. We used average pronounceability ratings as a predictor in the analysis.

Results

We found no overall effect of pronounceability on judgment of harmfulness, $t(52.9) = -0.71$, $p = .48$, $b = -0.03$, 95% CI = [−0.12, 0.06]. As in Study 5, the interaction between item source and pronounceability was nonsignificant, $t(39.9) = -1.70$, $p = .10$, $b = -0.12$, 95% CI = [−0.26, 0.02], but the effect of pronounceability was again stronger for the original items than for the newly constructed items. We again conducted separate analyses for the original items and the newly constructed items: Although participants judged harder-to-pronounce items to be more harmful when judging the original items, $t(7.9) = -2.54$, $p = .03$, $b = -0.15$, 95% CI = [−0.27, −0.04], there was no significant effect for the newly constructed items, $t(39.5) = -0.59$, $p = .56$, $b = -0.03$, 95% CI = [−0.11, 0.06].

A similar analysis for novelty ratings showed that easier-to-pronounce items were judged to be less novel, $t(46.9) = -3.81$, $p < .001$, $b = -0.21$, 95% CI = [−0.31, −0.10]. However, the analysis did not show a significant interaction between item source and pronounceability, $t(41.7) = -0.30$, $p = .77$, $b = -0.03$, 95% CI = [−0.20, 0.14]. When we conducted separate analyses for the original items and the newly constructed items, we found that easier-to-pronounce items were judged to be less novel for both the original items, $t(7.3) = -4.14$, $p = .004$, $b = -0.23$, 95% CI = [−0.34, −0.12], and the newly constructed items, $t(38.2) = -3.65$, $p < .001$, $b = -0.20$, 95% CI = [−0.31, −0.09]. Figure 3 shows the effect of pronounceability on harmfulness and novelty ratings at the item level.

Discussion

These results, together with the results of Study 5, suggest that the effect of pronounceability on judgment of harmfulness may be limited to the items used by Song and Schwarz (2009). There was no evidence for the effect in the newly constructed items. Although the interaction between item source and pronounceability was not significant in either of the studies, this might have been caused by limited statistical power due to the small number of original items (Westfall, Kenny, & Judd, 2014). Furthermore, when the interaction effects from both studies were meta-analytically combined, the result was significant, $z = 2.14$, $p = .03$, $r = .22$, 95% CI = [.02, .40]. On the other hand, the effect of pronounceability on perceived novelty was evident even for the newly constructed items, which suggests that a different robust fluency effect can be replicated even with the newly constructed items.

Study 7

In addition to the "food additives" scenario (used in current Studies 1–3 and 6), two other scenarios were used by Song and Schwarz (2009). In these two scenarios, people had to judge the extent to which amusement-park rides were adventurous or risky. In Study 7, we replicated this experiment, again using the original items and the newly constructed items.

Method

We posted a human intelligence task for 1,000 workers on MTurk; ultimately, 1,042 MTurk workers participated. We excluded 60 participants who used the same rating for all items in a scenario, in accordance with a preregistered exclusion criterion. We also excluded 32 participants who did not complete the whole study, resulting in the final sample of 950 participants.

We used two scenarios adopted from Study 3 in Song and Schwarz (2009). Participants were asked to imagine that they were visiting an amusement park and reading a brochure with names of amusement-park rides. In the "desirable risk" scenario, they imagined looking for the most adventurous ride, and they judged all presented rides on a scale ranging from 1 (*very dull*) to 7 (*very adventurous*). In the "undesirable risk" scenario, they imagined not feeling well that day, and they were told that they wanted to avoid rides that were too adventurous, which could make them sick. In this scenario, participants judged all presented rides on a scale ranging from 1 (*very safe*) to 7 (*very risky*). Participants were given the two scenarios in random order; each scenario required the evaluation of 11 randomly selected items.

We used a total of 206 items, 6 of which were Native American names used as stimuli in Study 3 by Song and Schwarz (2009). The original names were 6 to 13 letters long, so we randomly selected 25 Native American names for each of the lengths within this range from an Internet database. Because there were not enough names 11 to 13 letters long, we created the remaining names (to reach a total of 25) by randomly combining 3- to 5-letter names. We thus obtained 200 names. Next, we randomly changed one letter in half of the names to introduce more variability in pronounceability and to reduce the association between pronounceability and word length, which were confounded in the study by Song and Schwarz.

An independent sample of 303 MTurk workers was given a random sample of 50 names out of the 206 and asked to rate pronounceability on a scale from 1 (*easily pronounceable*) to 7 (*hard pronounceable*). To make it easier to compare the results with those of Studies 1 to 6, we reversed the average ratings such that the pronounceability variable was higher for easier-to-pronounce items. We also centered the variable by subtracting the mean of pronounceability ratings from all values. The name length was recoded on a scale from −0.5 to 0.5. The analysis was conducted using only the newly constructed items, and the mixed-effects model included random slopes for participants for name length, pronounceability, and scenario.

Results

A preregistered analysis of the ratings of the rides showed that rides with longer names were perceived as riskier, $t(230.3) = 5.18$, $p < .001$, $b = 0.71$, 95% CI = [0.44, 0.98] (Fig. 4, right graphs).[7] There was no significant effect of pronounceability on riskiness ratings, $t(198.1) = -0.28$, $p = .78$, $b = -0.01$, 95% CI = [−0.09, 0.07]. Participants gave higher ratings in the scenario presented second than in the scenario presented first, $t(947.7) = 3.82$, $p < .001$, $b = 0.16$, 95% CI = [0.08, 0.24], and somewhat lower ratings in the "undesirable risk" scenario than in the "desirable risk" scenario, $t(949.4) = -2.02$, $p = .04$, $b = -0.05$, 95% CI = [−0.10, −0.00]. These effects were qualified by their interaction, $t(953.3) = -4.31$, $p < .001$, $b = -0.21$, 95% CI = [−0.31, −0.12], showing that the difference between adventurousness and riskiness ratings was higher for the scenario presented first. The effect of pronounceability did not differ on the basis of name length, $t(197.9) = 0.42$, $p = .67$, $b = 0.04$, 95% CI = [−0.13, 0.20], but it was stronger for the "undesirable risk" scenario, $t(18{,}749.3) = -5.53$, $p < .001$, $b = -0.11$, 95% CI = [−0.14, −0.07], than for the "desirable risk" scenario.[8]

Given the scenario and order effects and the strong confounding association between pronounceability and name length, $r(204) = -.79$, 95% CI = [−.84, −.74], $p < .001$ (see also Fig. 4), we next analyzed the data for the two

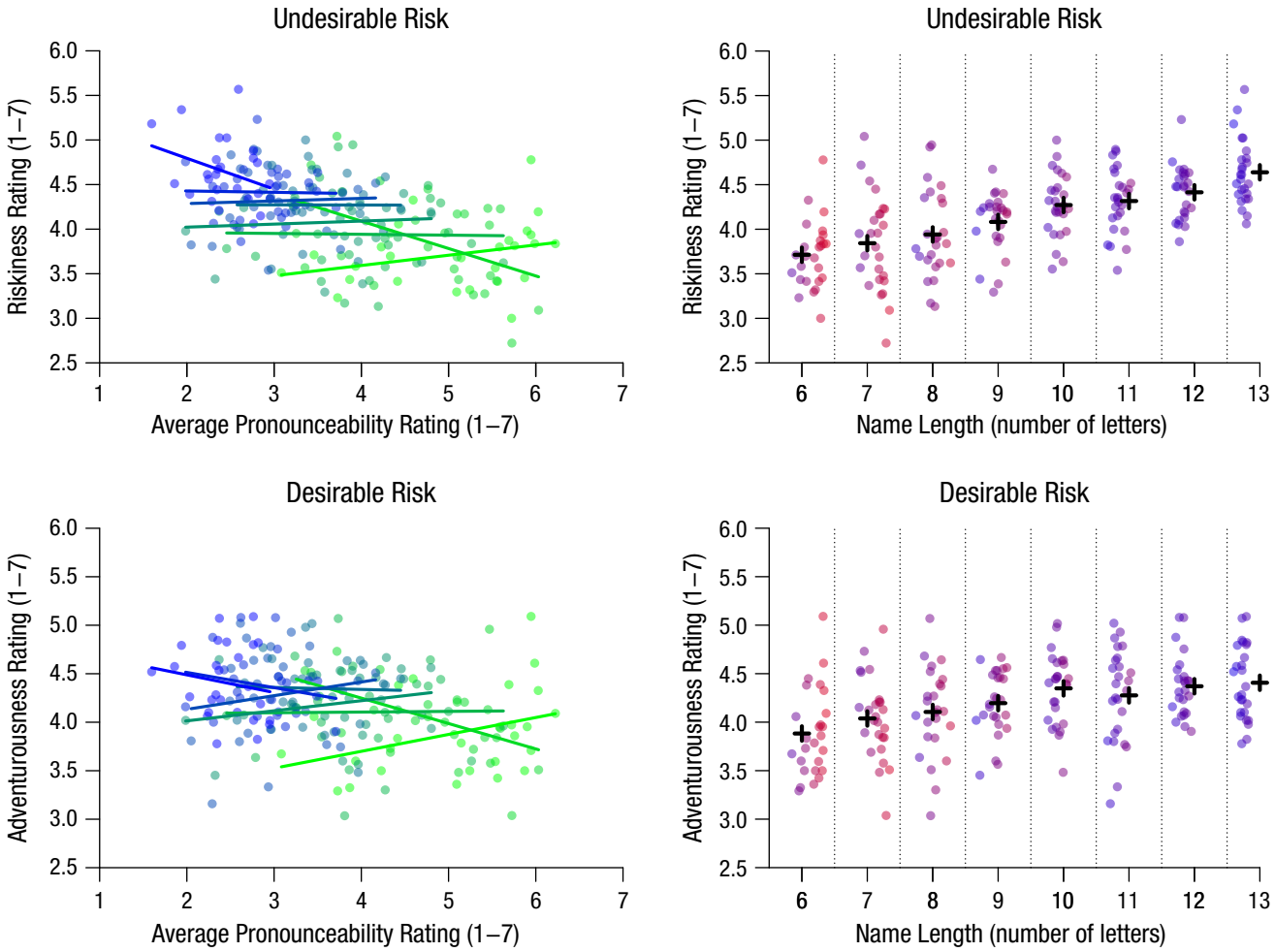

Fig. 4. Results from Study 7. The scatterplots (with best-fitting regression lines) on the left show the relationship between pronounceability ratings and riskiness ratings for an undesirable risk (top) and the relationship between pronounceability ratings and adventurousness ratings for a desirable risk (bottom). Words of different lengths are represented by different colors, from green (6 letters) to blue (13 letters). The graphs on the right show the association between name length and riskiness ratings for an undesirable risk (top) and the relationship between name length and adventurousness ratings for a desirable risk (bottom). Black crosses indicate the average riskiness and adventurousness ratings for names of each length. For each word-length group, the points are shifted according to the items' pronounceability rating, recoded on a scale from –0.5 to 0.5 (mean pronounceability = 0). The pronounceability of the names is also represented by the color of the plotted points, from blue (1; hard to pronounce) to red (7; easy to pronounce).

scenarios separately, using only the data from each participant's first scenario. There was no significant association between pronounceability and perceived risk for either of the scenarios when name length was taken into account, whereas name length still significantly predicted the ratings. Furthermore, adding pronounceability as a predictor to a model with name length did not improve the model, whereas adding name length to a model with pronounceability resulted in a significantly better fit. Therefore, the pronounceability effect that was seen without inclusion of name length as a predictor for both riskiness ratings, $t(275.0) = -7.03$, $p < .001$, $b = -0.23$, 95% CI = [–0.29, –0.17], and adventurousness ratings, $t(286.0) = -6.73$, $p < .001$, $b = -0.25$, 95% CI = [–0.32, –0.17], could be entirely driven by the association between pronounceability and name length. Figure 4 (left graphs) suggests that the relationship between pronounceability and perceived risk might be present for names of 7 and 13 letters. However, analysis estimating the effect of pronounceability separately for each name length yielded no significant pronounceability effects. In summary, the pronounceability of names of amusement-park rides was associated with their predicted riskiness and adventurousness; however, this effect disappeared when length of the names was taken into account.

Study 3 by Song and Schwarz (2009) used only three fluent and three disfluent items, and all the disfluent items were longer than the fluent items. It was therefore

not possible to evaluate the difference between the original items and the newly constructed items with adequate statistical power. However, we found no significant association between pronounceability and riskiness ratings for the original items for either of the scenarios. Mean riskiness ratings averaged across both scenarios were 4.98, 4.19, and 4.35 for the fluent items and 4.04, 4.69, and 4.57 for the disfluent items, respectively. Curiously, although we replicated the results of Song and Schwarz in Studies 5 and 6 using only their original items, in Study 7 we obtained the opposite pattern of results—we found the association between pronounceability and perceived risk only with the newly constructed items, not with the original items used by Song and Schwarz.

General Discussion

In the current research, we originally tested whether the association between processing fluency and judgment of risk differed depending on the category of evaluated objects (Studies 1–4). Although we initially found some support for the hypothesis, we obtained opposite results when trying to replicate the findings. The unexpected variability in the outcomes might have resulted from treating stimulus as a fixed factor and using different stimuli for each scenario: When we gave participants the same stimuli for two scenarios in which we expected opposite associations, we found no significant effect of the scenario (Study 4). Moreover, the effects that we believed supported our hypothesis were no longer statistically significant when we conducted a more appropriate statistical analysis in which we treated stimulus as a random factor.

Next, we randomly created new stimuli to eliminate any possible bias in the stimuli selection. Using these newly constructed stimuli, we found no significant effect of pronounceability on judgment of harmfulness. On the other hand, when analyzing only the original stimuli used by Song and Schwarz (2009), we found the original effect even for new scenarios (Study 5). This pattern of results was replicated even when using Song and Schwarz's original "food additives" scenario. Although we found no overall effect of pronounceability on judgment of harmfulness, we found an effect of pronounceability on judgment of novelty for both the original stimuli and the newly constructed stimuli (Study 6). In the final study, using another scenario employed by Song and Schwarz, we found the association between pronounceability and perceived safety for the newly constructed stimuli. However, the effect might have been driven completely by word length, which was confounded with pronounceability in Song and Schwarz's study. Furthermore, we did not replicate the pronounceability effect with the original stimuli used by Song and Schwarz.

In summary, the results show that the effect of pronounceability on judgment of riskiness may be much weaker than originally thought or even nonexistent. Although we found the relationship between pronounceability and perceived safety in the final study, the effect seemed to be completely driven by the association between pronounceability and name length. After controlling for name length, the effect of pronounceability disappeared. This is consistent with results of Studies 5 and 6, in which we found no significant effect of pronounceability on perceived riskiness when using the newly constructed names of the same length. Future studies could investigate specific aspects of stimuli responsible for the differences in the results obtained with the newly constructed stimuli and with the original stimuli in Studies 5 and 6, because it is possible that an unknown feature other than fluency caused the apparent association between fluency and perceived safety for the original stimuli. Although the study casts doubt on the effect of pronounceability on perceived risk, we replicated the effect of pronounceability on perceived novelty, and we found the association between word length and judgment of risk. It is possible that word length itself may be associated with processing fluency. Participants were not asked to read the stimuli aloud, so it is possible that pronounceability was a weaker manipulation of reading fluency than word length. Nevertheless, the degree to which the effect of word length is caused by its association with processing fluency and the degree to which it is caused by other factors (e.g., Lewis & Frank, 2016) remains an area of inquiry for future studies.

Our study underscores the importance of using random sampling of stimuli and appropriate analysis methods in both original and replication studies (Fiedler, 2011; Judd, Westfall, & Kenny, 2012; Westfall, Judd, & Kenny, 2015; Westfall et al., 2014). The results of Song and Schwarz (2009) were replicated in three recent studies (Cho, 2015; Dohle & Siegrist, 2014; Topolinski & Strack, 2010). However, two of the studies (Dohle & Siegrist, 2014; Topolinski & Strack, 2010) used the stimuli from Song and Schwarz, and the third (Cho, 2015) used only four different names; in all three studies, stimulus was treated as a fixed factor. Our results show that possible conclusions from these and similar studies are limited, and psychologists should follow the advice of Judd et al. (2012) to treat stimulus as a random factor if they want their results to be generalizable.

Action Editor

D. Stephen Lindsay served as action editor for this article.

Author Contributions

Both the authors designed the study, prepared the materials, collected data, and wrote the manuscript. Š. Bahník wrote the programs used for data collection and analyzed the data.

Acknowledgments

We thank Filip Děchtěrenko and Ira Theresa Maschmann for their help with data collection and Thorsten Michael Erle for his helpful comments.

Declaration of Conflicting Interests

The authors declared that they had no conflicts of interest with respect to their authorship or the publication of this article.

Funding

The work of Š. Bahník was supported by Grant DFG-RTG 1253/2 from the Deutsche Forschungsgemeinschaft and by Grant IP300040 from the Internal Grant Agency of the Faculty of Business Administration, University of Economics. The work of M. A. Vranka was supported by Charles University Grants PRVOUK7 and PRVOUK17.

Open Practices

All data and materials have been made publicly available via the Open Science Framework and can be accessed at https://osf.io/6zdnh. The design and analysis plans for the experiments were preregistered at the Open Science Framework (Study 3: https://osf.io/zua85; Study 4: https://osf.io/ucaby; Study 5: https://osf.io/xv23i; Study 6: https://osf.io/ryhs5; Study 7: https://osf.io/pvzem; Studies 1 and 2 were not preregistered). The complete Open Practices Disclosure for this article can be found at http://journals.sagepub.com/doi/suppl/10.1177/0956797616685770. This article has received badges for Open Data, Open Materials, and Preregistration. More information about the Open Practices badges can be found at http://www.psychologicalscience.org/publications/badges.

Notes

1. In all the studies reported here, a small number of participants judged most items using the same rating. This behavior can be perceived as noncompliance with instructions because these participants probably did not try to read and judge individual items. Therefore, we excluded data from these participants. The exclusion was done ad hoc on the basis of the judgment of an author blinded to participants' condition in the first two studies (1 and 2) and according to preregistered exclusion criteria for the next two studies (3 and 4). The exclusion criteria as well as examination of additional factors not relevant for the present article (e.g., instructions to read the names carefully, order effects) can be found at the Open Science Framework (https://osf.io/fjs56). The number of participants after exclusion is given in Fig. 1.
2. We did not continue data collection after analyzing data, except in the case of Study 2, for which data were pooled from two data sets. An additional study with a small sample size and subpar methods is described at https://osf.io/tswqy.
3. We directly asked a separate sample of participants whether within a given category, they expected exemplars with names that were familiar or easy to pronounce to be more or less dangerous. Although participants expected exemplars of prisoners, cities in a war zone, and poisons to be more dangerous if they had names that were familiar or easy to pronounce, they held the opposite expectation for food additives, tourist destinations, roller coasters, and medicines (the results can be found at the Open Science Framework, https://osf.io/6ytdm).
4. Although we intended to change two letters in the names, an error in code led to this deviation from the preregistered protocol in Studies 5 and 6.
5. Including random slopes for pronounceability in a model, as recommended by Barr, Levy, Scheepers, and Tily (2013), did not change the results of this or the next study. We thus report the results of preregistered analyses without the inclusion of the random slopes.
6. A possible concern may be that the newly constructed names might have retained some similarity to the names of the medicines from which they were derived. To check this possibility, we asked a separate sample of 210 participants to assess the degree to which the names reminded them of the name of an existing substance (1 = *not at all*, 7 = *very*). None of the names reminded the participants strongly of the names of existing substances (all mean ratings were lower than 2.7), and there was no difference between the newly constructed items and the original names, $t(48.0) = 0.17$, $p = .87$, $b = 0.02$, 95% CI = [–0.21, 0.25]. For details of the analysis, see https://osf.io/cswyn.
7. Additional details of the analysis presented can be found at the Open Science Framework (https://osf.io/efh4n).
8. However, we did not include the interaction between name length and scenario in the model, so the interaction between pronounceability and scenario may have been due to the association between name length and pronounceability.

References

Alter, A. L., & Oppenheimer, D. M. (2008). Easy on the mind, easy on the wallet: The roles of familiarity and processing fluency in valuation judgments. *Psychonomic Bulletin & Review*, *15*, 985–990.

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.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*, 255–278.

Cho, H. (2015). The malleable effect of name fluency on pharmaceutical drug perception. *Journal of Health Psychology*, *20*, 1369–1374.

Dohle, S., & Siegrist, M. (2014). Fluency of pharmaceutical drug names predicts perceived hazardousness, assumed side effects and willingness to buy. *Journal of Health Psychology*, *19*, 1241–1249.

Fiedler, K. (2011). Voodoo correlations are everywhere—not only in neuroscience. *Perspectives on Psychological Science*, *6*, 163–171.

Galak, J., & Nelson, L. D. (2011). The virtues of opaque prose: How lay beliefs about fluency influence perceptions of

quality. *Journal of Experimental Social Psychology*, *47*, 250–253.

Judd, C. M., Westfall, J., & Kenny, D. A. (2012). Treating stimuli as a random factor in social psychology: A new and comprehensive solution to a pervasive but largely ignored problem. *Journal of Personality and Social Psychology*, *103*, 54–69.

Lewis, M. L., & Frank, M. C. (2016). The length of words reflects their conceptual complexity. *Cognition*, *153*, 182–195.

McGlone, M. S., & Tofighbakhsh, J. (1999). The Keats heuristic: Rhyme as reason in aphorism interpretation. *Poetics*, *26*, 235–244.

Pocheptsova, A., Labroo, A. A., & Dhar, R. (2010). Making products feel special: When metacognitive difficulty enhances evaluation. *Journal of Marketing Research*, *47*, 1059–1069.

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

Song, H., & Schwarz, N. (2009). If it's difficult to pronounce, it must be risky: Fluency, familiarity, and risk perception. *Psychological Science*, *20*, 135–138.

Topolinski, S., & Strack, F. (2010). False fame prevented: Avoiding fluency effects without judgmental correction. *Journal of Personality and Social Psychology*, *98*, 721–733.

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, *4*, 207–232.

Westfall, J., Judd, C. M., & Kenny, D. A. (2015). Replicating studies in which samples of participants respond to samples of stimuli. *Perspectives on Psychological Science*, *10*, 390–399.

Westfall, J., Kenny, D. A., & Judd, C. M. (2014). Statistical power and optimal design in experiments in which samples of participants respond to samples of stimuli. *Journal of Experimental Psychology: General*, *143*, 2020–2045.