Processing fluency and judgment

Verarbeitungsflüssigkeit und Urteilsbildung

Doctoral thesis for a doctoral degree
at the Graduate School of Life Sciences,
Julius-Maximilians-Universität Würzburg,
Section Neuroscience

submitted by

Štěpán Bahník

from

Prague, Czech Republic

Würzburg, 2016

Submitted on: ……………………………………………………………..……..

Office stamp

Members of the *Promotionskomitee*:

Chairperson: …………………………………………………………………......

Primary Supervisor: Prof. Dr. Fritz Strack

Supervisor (Second): Prof. Dr. Erhard Wischmeyer

Supervisor (Third): Prof. Dr. Karl Mertens

Supervisor (Fourth): ……………………………………………………………….

Date of Public Defence: ……………………………………………….…………

Date of Receipt of Certificates: ……………………………………………….

Content

Publications included in this thesis

Chapter III contains text from a manuscript currently under review in Psychological Science:

Bahník, Š., & Vranka, M. (under review). *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.*

Chapter IV contains text from a manuscript published in Journal of Research in Personality:

Bahník, Š., Vranka, M., & Dlouhá, J. (2015). X good things in life: Processing fluency effects in the "Three good things in life" exercise. *Journal of Research in Personality, 55*, 91-97. http://dx.doi.org/10.1016/j.jrp.2015.02.005

The manuscript is reprinted with permission by Elsevier, which can be found on osf.io/etv2p/.

Zusammenfassung

Um Urteilsprozesse zu vereinfachen, basieren Menschen diese oft auf leicht zugänglichen Informationen. Eine hierfür immer und jederzeit verfügbare Information ist die Verarbeitungsflüssigkeit – ein metakognitives Gefühl von Leichtigkeit bei der kognitiven Verarbeitung. Verarbeitungsflüssigkeit wird als Information bei vielen verschiedenen Arten von Urteilen genutzt, so beispielsweise bei Urteilen hinsichtlich Wahrheit, Vertrauenswürdigkeit und Neuheit. Die vorliegende Arbeit beschreibt die Ergebnisse von drei Studien, die verschiedene Aspekte des Einflusses von Verarbeitungsflüssigkeit auf Urteilsprozesse untersuchen.

Verarbeitungsflüssigkeit wurde in der Literatur manchmal mit der Geschwindigkeit eines kognitiven Prozesses gleichgesetzt. Daher wurden Reaktionszeiten zur Schätzung der Verarbeitungsflüssigkeit herangezogen. In experimentellen Paradigmen umfassen Reaktionszeiten allerdings oft nicht nur die Zeit, die für den Prozess von Interesse benötigt wird, sondern auch die Zeit, die für eine Entscheidung benötigt wird, die auf der aus dem Prozess resultierenden Information basiert. Die Studie, die in Kapitel II beschrieben wird, verwendet daher eine neue experimentelle Methode, die eine Differenzierung zwischen der Zeit, die zum Lesen, und der Zeit, die für die Entscheidung benötigt wurde, ermöglicht. In den Ergebnissen zeigt sich, dass Menschen schneller eine Entscheidung darüber fällen, wie sehr sie bestimmte Pseudowörter mögen, wenn die Pseudowörter schwer aussprechbar sind (also nicht flüssig verarbeitbar sind) als wenn sie moderat aussprechbar sind. Dieser Befund legt nahe, dass Reaktionszeiten nicht zur Schätzung von Verarbeitungsflüssigkeit

herangezogen werden können, da sie auch die Zeit beinhalten, die gebraucht wird, um eine Entscheidung zu treffen.

Eine Studie zum Einfluss von Verarbeitungsflüssigkeit auf Urteilsbildung zeigte, dass Lebensmittelzusatzstoffe mit leicht aussprechbaren Namen im Vergleich zu solchen mit schwer aussprechbaren Namen als weniger gefährlich beurteilt wurden. Während Menschen ungefährlichen Lebensmittelzusatzstoffen öfter begegnen als gefährlichen, könnte dieser Umweltzusammenhang bei anderen Objektkategorien in entgegengesetzter Richtung verlaufen. Zum Beispiel sehen Menschen die Namen von besonders gefährlichen Verbrechern mit einer höheren Wahrscheinlichkeit in den Nachrichten. In Kapitel III wird daher eine Studie beschrieben, die ursprünglich testen sollte, ob der Verarbeitungsflüssigkeit-Sicherheit-Zusammenhang bei manchen Objektkategorien, bei denen man selektiv besonders gefährlichen Exemplaren ausgesetzt ist, in die entgegengesetzte Richtung verläuft. Die Ergebnisse unterstützten diese Hypothese jedoch nicht. Zudem lassen die Ergebnisse weiterer Studien darauf schließen, dass der zuvor gezeigte Zusammenhang zwischen Verarbeitungsflüssigkeit und Sicherheit nur mit den Originalstimuli aus früheren Studien, nicht aber mit neu konstruierten Studien repliziert werden kann.

In Kapitel IV wird eine Studie beschrieben, die einen Befund der Verarbeitungsflüssigkeitsliteratur auf eine Übung aus dem Bereich der Positiven Psychologie anwendet, um deren Effektivität zu steigern. Konkret manipulierte das Experiment die Anzahl an guten Erlebnissen, die die Teilnehmer zwei Wochen lang täglich als Teil der Übung auflisten sollten. Obwohl das Auflisten einer größeren Anzahl an Erlebnissen als schwieriger bewertet wurde, hatte die Anzahl an Erlebnissen, die jeden Tag aufgelistet wurde, keinen Einfluss auf die Effektivität der Übung.

Summary

To simplify a judgment, people often base it on easily accessible information. One cue that is usually readily available is processing fluency – a metacognitive feeling of ease of cognitive processing. Consequently, processing fluency is used as a cue for many different types of judgment, such as judgment of truth, confidence, and novelty. The present work describes results of three studies investigating various aspects of processing fluency effects on judgment.

Processing fluency has been sometimes equated with speed of a cognitive process. Therefore, response times have been used for evaluation of processing fluency. However, response times in experimental tasks often do not encompass only the time needed for a given process, but also the time needed for a decision based on the resulting information. The study described in Chapter II uses a novel experimental method that enables separation of reading and decision times. The results show that people make a decision about liking of pseudowords faster when the pseudowords are hard-to-pronounce (i.e., disfluent) than when they are moderate in pronounceability. This suggests that response times cannot be used as a proxy for processing fluency when they include the time needed to make a decision.

One of the studies of judgmental effects of processing fluency showed that food additives with easier pronounceable names are judged to be less harmful than those with hard-to-pronounce names. While people encounter food additives that are safe more often, this environmental association may be in the opposite direction for some categories of objects. For example, people are more likely to see names of especially

dangerous criminals in the news. Chapter III describes a study which initially tested whether the fluency-safety association may be in the opposite direction for some categories of objects as a consequence of this selective exposure to especially dangerous exemplars. The results did not show support for this hypothesis. Furthermore, subsequent studies suggest that the previously found association between fluency and safety is replicable with the original stimuli used in the previous research, but not with newly constructed stimuli.

Chapter IV describes a study which applied a finding from the processing fluency literature to a positive psychology exercise in order to increase its effectiveness. Namely, the experiment manipulated the number of good things that participants listed daily for two weeks as part of the exercise. While listing more things was considered harder, the number of things listed each day had no effect on effectiveness of the exercise.

Acknowledgements

I would like to thank my supervisor Fritz Strack for taking me as his PhD student, giving me independence to work on my own projects, and working with me on our joint projects. I could not have found a better place for my PhD studies than Lehrstuhl II.

Research is often a lonely endeavor, but I have been lucky to find a collaborator working with whom has been both fun and productive. I am greatly thankful to Marek Vranka for all the work we have done together so far, and I am looking forward to all the work we will still do together in the future.

I would like to thank all my colleagues from the department, who were always willing to help me with whatever language or other issue I had encountered. I am especially indebted to Felix Götz in this respect.

I am also grateful to all the research assistants and other people who helped me with my research in various ways. I am also thankful for the generous support by Deutsche Forschungsgemeinschaft (DFG-RTG 1253/2) that funded the bigger part of my PhD studies.

Last but not least, I would like to thank my family, friends, and my girlfriend for making me look forward to my returns to Prague and for all the support during the last three years.

CHAPTER I – Introduction

Judgment

More than 40 years ago, Tversky and Kahneman (1974) described mental shortcuts that people sometimes use when making a judgment. These so-called “heuristics” stood in contrast to the view that saw people as rational reasoners and decision makers, which had been widely held and still survives until this day, for example, in economics. Heuristics have been traditionally identified by demonstrating associated biases, that is, systematic mistakes and fallacies to which they lead. This process of researching heuristics was akin to identifying perceptual processes using visual illusions (Kahneman, 1991). Due to the process used to study heuristics, the “heuristics and biases” tradition of judgment and decision research focused on mistakes associated with intuitive judgment. However, as both the proponents and critics of the heuristics and biases program noted, judgment based on heuristics is often efficient and well-suited for answering the question hand (Gigerenzer, 2008; Gigerenzer, Czerlinski, & Martignon, 2002; Gilovich & Griffin, 2002).

By the virtue of simplification of judgment, heuristics may make the process much more efficient than if a more deliberate, rule-based judgmental process was used. The simplification has many potential advantages – saving of time needed for making a judgment being probably the biggest one. There is a variety of ways how judgment can be simplified (Shah & Oppenheimer, 2008), one of which is the use of easily accessible information. The use of easily accessible information can be prominently seen in the anchoring effect – that is, the assimilation of a judgment towards a previously considered standard (Bahník, Englich, & Strack, in press). In the case of anchoring, the previously considered standard activates information that is associated with it and this information is then more easily accessible, and therefore more likely to be used in the subsequent judgment (Strack, Bahník, & Mussweiler, in press). Apart from the use of already activated information as in anchoring, the use of easily accessible information is seen in the use of cues that are easy to process (Shah & Oppenheimer, 2007, 2009) or that are readily available. Such cues may result from automatic processes associated with perceptual and basic cognitive processing – for example, from categorization in the case of similarity (Tversky & Kahneman, 1982, 1983) – other may be easily accessible as a side-effect of cognitive processing itself. For example, all cognitive processes are associated with the duration that was needed for their completion, and from perceived duration of a cognitive process a feeling of ease of processing – also called "processing fluency" – may result. This metacognitive information can then be used itself as a cue for judgment.

Processing fluency

Since the feeling of ease of processing is associated with a large range of cognitive processes and is thus readily available, it is used as a cue in many different types of judgment. For example, processing fluency was found to be associated with truth (McGlone & Tofighbakhsh, 1999), liking (Reber, Winkielman, & Schwarz, 1998), confidence (Simmons & Nelson, 2006), familiarity (Tversky & Kahneman, 1973), and risk (Song & Schwarz, 2009; but see Chapter III). The variety of processes that have been studied in relation to processing fluency is also large. Depending on the cognitive process, people may feel, for example, perceptual fluency as a result of high figureground contrast (Reber et al., 1998), phonological fluency as a result of good pronounceability (Song & Schwarz, 2009), orthographic fluency as a result of easy reading of a written text (Alter, Oppenheimer, Epley, & Eyre, 2007), or retrieval fluency as a result of fast recollection of memories (Winkielman, Schwarz, & Belli, 1998). It has been argued that from a combination of these various forms of fluency, a common feeling of fluency arises (Alter & Oppenheimer, 2009; Reber, Wurtz, & Zimmermann, 2004; Wurtz, Reber, & Zimmermann, 2008), which can then in turn be used as a cue in judgment. Consequently, the various forms of fluency should lead to the same judgmental effects, some of which were mentioned above.

Processing fluency as a cue

In many situations, processing fluency serves as a valid cue (Unkelbach & Greifeneder, 2013). For example, people may judge a fluent statement as more likely to be true, because they were more often exposed to true statements than false statements and previous exposure causes fluent processing of the statements (Reber & Unkelbach, 2010). In such cases, fluency is used a cue due to its association with the target attribute that exists as a feature of the environment. Fluency will then be an invalid cue when the association does not hold. For example, fluency of a statement may be a result of better font rather than previous exposure. As long as the font is just an incidental feature of the evaluated statement, it will not serve as a valid cue. In such a situation, it is possible that the person evaluating the statement will recognize the source of fluency and correct for its lack of validity. This is possible to see in experiments where participants are told a plausible source of fluency to which they may attribute the experienced ease of processing. The misattribution, that would otherwise lead to the use of fluency in judgment, then disappears. For example, in one study participants were asked to list 6 or 12 examples when they behaved assertively or unassertively. Those who recalled 12 examples of assertive or 6 examples of unassertive behavior judged themselves to be more assertive than those who recalled 12 examples of unassertive or 6 examples of assertive behavior. This is in line with the fluency account of the use of recollection in judgment since participants used their ease of recollection rather than recollected content in their judgment. However, this pattern of results occurred only when participants were able to attribute the ease of recollection to themselves rather than to some outside influence. For example, when participants could have assigned the ease or difficulty of recollection to the feature of the task or to an outside environmental influence, the effect disappeared (Schwarz et al., 1991). Similarly, participants may spontaneously discount fluency as a cue if they are able to attribute it to a specific feature of the stimulus. For example, people discount fluency of a name in the judgment of frequency of the name if they know that the fluency is caused by an association of the name with a well-known person (e.g., Bush) of by its semantic meaning (e.g., Stone). In this case, people know that they cannot use fluency as a valid cue for the frequency judgment and consequently no fluency effect is observed (Oppenheimer, 2004; Oppenheimer & Monin, 2009).

It seems that processing fluency is used as a cue in judgment only when participants have a standard by which they can judge how fluent the cognitive process is. Some studies thus show that fluency has no effect when it varies between subjects, but not within the experimental task (Dechêne, Stahl, Hansen, & Wänke, 2009; Hansen, Dechêne, & Wänke, 2008). Another study found that visual clarity had an effect on perceived fluency and judgment of certainty, but not liking using a manipulation between subjects. However, the same manipulation influenced all three measures when it varied within subjects (Forster, Rerger, & Leder, 2015). Other, already described, studies manipulated fluency between participants, but still found its effect on judgment. It is possible that in some situations people have a standard based on previous experience which they can use in the experimental situation. In such a case, a stimulus can still be perceived as relatively fluent or disfluent in comparison to this internal standard.

Mechanism of processing fluency

Given the number of different types of processing fluency, the mechanism that connects these processes is of interest. According to one view, some of the fluency effects might have a mental simulation as an underlying mechanism. For example, reading fluency may rest at least partly on the ease of automatic simulation of pronunciation of the read words. Consequently, it is possible to prevent the usual fluency effects by blocking the simulation. For example, chewing a gum may compete with the automatic simulation of pronunciation. Chewing a gum thus interferes with the effect of previous presentation on liking of words, but not of visual characters (Topolinski & Strack, 2009a; but see also Westerman, Klin, & Lanska, 2015). In another study (Beilock & Holt, 2007), skilled typists preferred letter dyads that were easy to type; even though, they did not type them during the experimental task. The effect was attenuated when participants conducted a motor task while making the preference judgments, presumably blocking mental simulation of typing the letter dyads which caused the effect. Similarly, it has been found that people associate positive concepts with the side of their dominant hand since the manipulation of objects on that side of the body is generally easier.

Furthermore, this association reverses when people are in a situation where their dominant hand is no longer more dexterous than their non-dominant hand, such as for patients with hemiparesis or participants in a laboratory experiment who are asked to work with their dominant hand wearing a clumsy mitten (Casasanto & Chrysikou, 2011).

Processing fluency has been often associated with positive affect (Forster, Leder, & Ansorge, 2013; Winkielman & Cacioppo, 2001). According to the hedonic marking hypothesis (Winkielman, Schwarz, Fazendeiro, & Reber, 2003; see also Topolinski & Strack, 2009b), the judgmental effects of processing fluency are mediated by positive affect that arises from fluent processing. While the association of fluency with positive affect was observed in many studies, a recent study (Albrecht & Carbon, 2014) suggests that fluency may amplify affect rather than elicit positive affect. The authors presented participants pictures that varied in their valence and manipulated fluency using perceptual priming. As in previous studies, participants rated pictures with positive valence more positively when they were primed. However, the priming also made them rate pictures with negative valence more negatively. The usually found association of fluency with positive affect might have been therefore caused by predominant use of neutral and positive stimuli in previous experiments. Furthermore, Westerman, Lanska, and Olds (2015) argue that if positive affect mediates the effect of fluency on judgment, as suggested by the hedonic marking hypothesis, the fluency effect on liking should be generally stronger than on other judgments. Contrary to this prediction, they found that the same fluency manipulation may influence familiarity but not liking.

Apart from the account that argues that fluency effects on judgment are mediated by positive affect, another account that may unify the various processing fluency effects is based on the speed of the process at hand. That is, felt fluency may be a result of a relatively short duration that it took to complete the fluent cognitive process. For example, Hertwig, Herzog, Schooler, and Reimer (2008) showed that people use speed of retrieval as a cue in various types of judgment and demonstrated how this use of fluency may make the judgments better.

It is noteworthy that the three described mechanisms – mental simulation, positive affect, and the speed of a cognitive process – are not necessarily exclusive since they are all concerned with different aspect of the judgmental effects of fluency. For example, it is possible that people perceive some process as fluent when the associated mental simulation is relatively fast and that the resulting feeling of fluency elicits positive affect, which is then used as a basis of subsequent judgment.

Association of fluency and a target attribute

Alter and Oppenheimer (2009) argue that the association between fluency and a target attribute of a judgment results from the use of naïve theories that connect a common feeling of fluency and the target attribute. This can be seen, for example, in a study by Winkielman and Schwarz (2001) where participants were asked to recall either 4 or 12 childhood events. Directly manipulating naïve theories, participants were either told that memories from unpleasant or pleasant periods of life are difficult to recall. Participants who recalled 12 events experienced difficulty in their recollection. Those who were told that memories from unpleasant periods are difficult to recall judged their childhood more negatively than those who were told that memories from pleasant periods are difficult to recall. This shows that participants used the manipulated naïve theories in their interpretation of the felt disfluency. On the other hand, there was no significant effect of the manipulation for participants recalling 4 events. Similarly, in a study by Briñol, Petty, and Tormala (2006), participants were asked to list 2 or 10 arguments for implementation of a new academic policy. In order to manipulate naïve theories about positivity of experienced ease and difficulty of processing, participants were told that it is usual either for unintelligent or for intelligent people to experience difficulties while thinking. Consequently, participants who were told that intelligent people experience difficulties when thinking were more likely to be in favor of the new policy if they listed 10 arguments (i.e., if they experienced difficulty in coming up with the arguments) than if they listed 2 arguments. The opposite association held for participants who were told that unintelligent people experience difficulties when thinking.

The influence of the interpretation of processing fluency on its effect on judgment can be seen also by measuring pre-existing beliefs. For example, people with the belief that intelligence is fixed interpreted experienced difficulty in learning, manipulated by a font of learned words, as a failure. Consequently, they judged that their future performance will be worse than if they learned words written in easily legible font. On the other hand, people who held the belief that intelligence is malleable interpreted the difficulty in the opposite way – as an evidence that they are learning. Accordingly, they judged that their future performance will be better if they were given words in an easy-to-read font than if the font was less legible (Miele, Finn, & Molden, 2011). On a similar note, people judge the same short story as better if they read it in an easily readable font if it is described as a historical study. However, the same story is considered better if read in a difficult-to-read font if it is introduced as a short story. People assume that a historical study should be harder to read than a short story and this expectation influences their evaluation of the text. When their expectations fit their experience, they evaluate the text more positively (Galak & Nelson, 2011). While Galak and Nelson manipulated the description of the same evaluated object, Labroo, Lambotte, and Zhang (2009) influenced participants' focus. In particular, they focused participants either on memorability or comprehensibility of a theory by asking them a question about one of these characteristics of the theory. The association with importance of the theory differs between these two characteristics. A more important theory is likely to be harder to comprehend but easier to remember. Therefore, ease of processing led to higher perceived importance of the theory when participants focused on memorability, but lower perceived importance if they focused on comprehensibility.

Given that fluency is often used as a valid cue, the association between fluency and a target attribute of a judgment can be also learned. This has been demonstrated in studies where participants were taught the association opposite to the usually observed by exposure to instances where the target attribute was correlated with fluency in the opposite than the usual direction. For example, the usually observed association between fluency and truth was reversed by exposure to statements for which the true statements were presented in a disfluent font and the false statements in a fluent font. After this learning of the association between disfluency and truth, participants were more likely to judge fluent statements as false contrary to the usually observed effect. Furthermore, this effect occurred even when fluency of the evaluated statements was manipulated in a different manner from the learning phase (Unkelbach, 2007). A similar reversal was shown for recognition judgment when previously seen stimuli were associated with disfluency in the learning phase (Unkelbach, 2006). Other studies also used learning of the reversed association between fluency and the target attribute of a judgment; however, they did not find the reversal of the fluency effect. It is possible that the learning has to be sufficiently long to reverse, and not only decrease, the effect of fluency on judgment (Olds & Westerman, 2012; Thomas & Morwitz, 2009).

Applications of processing fluency research

In the laboratory experiments, it is possible to manipulate the amount of information available for judgment. When fluency is one of a few cues that can be used for judgment, it is likely that it will be used. While laboratory experiments may show that it is possible that people use fluency as a cue for a given type of judgment and what the direction of the fluency effect likely is, studies outside of the laboratory are important for understanding the size of the fluency effect and its potential impact in the real-world situations.

One applied study tested whether disfluency may lead to better retainment of material learned during high school classes (Diemand-Yauman, Oppenheimer, & Vaughan, 2011). In particular, the authors followed-up on studies showing that disfluency may lead to deeper processing (Alter et al., 2007) and changed the font used in slides presented during several classes. They found that materials presented in a difficult-to-read font led to better performance in exams than materials presented in an easy-to-read font. As Greifeneder et al. (2010) showed, fluency may play a role not only in the learning, but also in the evaluation by teachers. In their study, essays written in less legible handwriting were evaluated more negatively than essays written in more legible handwriting. In another study from the educational setting, Fox (2006) showed that students asked to list 10 way how a university course can be improved rated the course as better than students asked to list 2 ways of possible improvement. Presumably, the students listing 10 ways experienced disfluency when coming up with the ideas for improvement. Therefore, they concluded that there are not that many ways how the course could be improved and thus rated it more positively.

The effect of ease of recollection was also shown in studies exploring the attribution of the amount of work done by oneself and others. Since it is easier to recall the work one has done, people usually overestimate the amount of work they have done in comparison to the work done by others. This has been shown in estimates of work done in couples (Ross & Sicoly, 1979; Thompson & Kelley, 1981) as well as teams (Ross & Sicoly, 1979; Savitsky, Van Boven, Epley, & Wight, 2005).

Overview of the studies

Since processing fluency is associated with easy and fast processing, response times are sometimes used for evaluation of fluency of a cognitive process; that is, shorter response times are said to indicate fluency. **Chapter II** describes a study which tested whether disfluency may lead to faster decision times when it serves as a strong cue in judgment. The experiment manipulated retrieval fluency using previous presentation and phonological fluency by varying pronounceability of pseudowords. The results showed the effects of both retrieval and phonological fluency on liking. Furthermore, a predicted inverted-U shaped relationship between pronounceability and decision times was found. Decisions were faster for disfluent and fluent pseudowords than for pseudowords moderate in fluency. The study thus demonstrates the importance of separating different processes comprising judgment when using response times as a proxy for processing fluency.

Previous research has shown that people judge food additives with more difficult-to-pronounce (i.e., disfluent) names as more harmful (Song & Schwarz, 2009). The study described in **Chapter III** originally explored the possibility that the association between disfluency and harmfulness might be in the opposite direction for some categories of objects. While initial studies seemed to support the hypothesis, further experiments indicated that the fluency effect is strongly dependent on the stimuli used. The next two experiments used stimuli sampling and showed that the original fluency–safety association was not replicable with newly constructed stimuli. Another study showed the fluency–safety association with newly constructed stimuli, however, only when pronounceability was confounded with word length. The results cast doubt on generalizability of the association of pronounceability and safety and underscore the importance of treating stimuli as a random factor.

The study described in **Chapter IV** applied the findings from the processing fluency literature to a positive psychology exercise that is supposed to increase happiness. In the "Three good things in life" exercise, people are asked to list each day three good things that happened to them during that day. We manipulated the number of things people were asked to write down each day. The participants thus daily wrote 1 to 10 good things for two weeks. We did not find any effect of the number of things people listed daily on life satisfaction, positive or negative affect. Furthermore, participants' life satisfaction or positive affect did not change after the two weeks, but their negative affect somewhat decreased. Various other aspects of the exercise were investigated in exploratory analyses.

CHAPTER II – Disfluent, but fast

Introduction

Processing fluency has often been associated with speed of a cognitive process. For example, easier pronounceable words might be felt as fluent because they are pronounced faster. Consequently, response times have been sometimes used as a proxy for processing fluency (e.g., Thomas & Morwitz, 2009; Unkelbach, 2007). Another possible view of the relationship between processing fluency and response time was proposed by Reber, Wurtz, and Zimmermann (2004). Reber and colleagues argue that response latencies can be used to measure *objective* fluency, which contributes to *subjective* fluency, denoting the subjective feeling of ease of a cognitive process.

While response times may often be useful for assessment of processing fluency, it is necessary to take into account that fluency may differ between processes that comprise the response. For example, a judgment of liking of a word does not consist of a single process. An easily pronounceable word may be read easily, but the decision between possible answers might be hard because fluency serves just as a single cue that can be opposed by other cues. If the decision process is not separated from reading, response times may paint a misleading picture about fluency of the underlying processes. In the present experiment, we tried to separate the perceptual process (in our case – reading) from the decision process. By separating the two processes, it was possible to assess the effect of fluency on response times for the two processes individually. We expected that processing fluency manipulated by both a previous presentation and varying pronounceability of words would lead to faster response times during reading. However, we expected that perceptual fluency will have an inverted-U shaped relationship with decision times.

The expected relationship of fluency and decision times can be understood from the perspective of the diffusion decision model (Ratcliff, Smith, Brown, & McKoon, 2016; Voss, Nagler, & Lerche, 2013). According to the diffusion model, decision making can be viewed as a process of noisy accumulation of evidence. While the details of the model are not important for the present study, the diffusion model simply states that a decision is made when a decision maker accumulates enough evidence to pass a certain threshold. The average rate of accumulation is determined by the so-called drift rate. For evaluation of a stimulus in terms of liking, the drift rate is influenced by various features of the stimulus. The feature that is of interest for the present study is processing fluency. Importantly, holding other variables in the diffusion model constant, the drift rate influences not only the probability of the “like” and “dislike” responses, but also the time needed to reach the decision. If the same amount of evidence is required for both answers and the decision maker is not strongly biased toward any of the answers by default, the higher the slope of the drift rate, the higher the probability of a given response is and the faster it will be made. If fluency generally influences the drift rate toward positive values (i.e., toward the option “like”) and disfluency toward negative values, the diffusion model predicts the inverted-U shaped relationship between fluency and decision times.

In the present experiment, we used two manipulations of fluency. We presented participants with pseudowords that varied in their pronounceability and that were either shown previously, or not. Furthermore, we used two types of judgments. Participants had to decide whether they had seen the pseudoword before or whether they like it. Importantly, they did not know during the reading which question they would answer and which answer will be assigned to which key. Furthermore, the word disappeared once participants indicated that they had read it. Consequently, they could not have easily prepared the answer during the reading and the reading and decision phases were thus separated in the experiment.

Method

The materials, analysis scripts, and data can be found on osf.io/9fxeh/.

Participants

Two hundred participants (84.5 % students; 89 % right handed, $Mdn_{\text{age}} = 23$) participated in the study which was administered as the first experiment in a larger set of studies.

Procedure

The experiment consisted of two parts, both administered on a computer using a custom written Python program (see Figure 1 for schema of the experiment). First, participants read sequentially presented 40 pseudowords (e.g., inptagzakr, aktenmiatz, deseizurrz; henceforth “words” for simplicity), each for one second (with an interstimulus interval of 500 ms). The second part consisted of 80 trials. During each trial, participants first saw a word and replied by pressing the spacebar once they read it. A half of the words had been shown in the first part of the experiment. Afterwards, the word disappeared and participants were offered one of two possible pairs of answers – either “like” and “dislike”, or “seen” and “not seen”. The answers determined whether they should reply if they like the word or if they saw the word during the first part of the experiment. The answers were shown on the two sides of the screen and corresponding keys (“S” on left and “K” on right) were shown below them. The side of both answers as well as the pair of answers offered was randomly determined for each trial. Therefore, participants could not easily prepare the answer when they read the word because they did not know the question that they would be asked and which key would correspond to their prefered answer. After each trial, an intertrial interval of 250-750 ms (randomly determined) followed before presentation of the next word. Participants were instructed to respond as fast as possible both after reading the word and while answering the question related to the word. They were also instructed to try to use both “like” and “dislike” options and told that they had seen half of the words during the first part of the experiment. Before the experiment, participants practiced the procedure on 10 trials with 5 previously presented words.

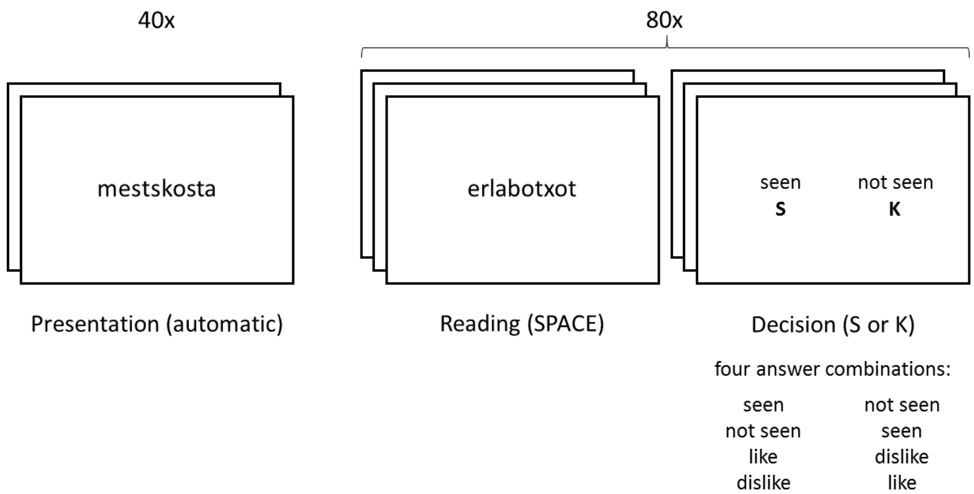

Figure 1: Schema of the experimental procedure. Participants were first sequentially presented 40 words. The next part of the experiment consisted of 80 trials. As a part of each trial, participants were first shown a word (half of the words had been presented before). After indicating that they read the word by pressing SPACE, participants were asked to decide whether they had seen the word or whether they liked the word. The answers were randomly assigned to “S” and “K” keys on each trial.

Stimuli

We used a list of 130 ten-letters-long pseudowords as stimuli. The words were randomly constructed such that they were not familiar to the participants and that they varied in pronounceability. We first took 12-letter long Basque words and then randomly took first 3 letters from a random word from the sample of the Basque words, appended letters 4-6 from a random word from the sample, and similarly appended letters 7-9 and 10-12. To obtain 10-letter words, a random consecutive 10-letter string was taken from the created 12-letter words. The same ten words were always used for the practice session. Out of the remaining 120 words, each participant received 80 randomly selected words during the experimental part of the study. After completing unrelated studies, each participant rated additional 25 words in terms of their pronounceability on a scale from 1 (hard pronounceable) to 7 (easily pronounceable) and 15 remaining words in terms of whether they believe that they exist in some world language on a scale from 1 (surely does not exist) to 7 (surely exists). The average ratings of words in these two measures correlated highly, $r(118) = .83$, 95% CI = [.77, .88], $p < .001$. We therefore used only pronounceability in subsequent analyses, which has been often equated with fluency, that was the primary topic of the study.

Results

All analyses were done with mixed-effect models and generalized mixed-effect models (Baayen, Davidson, & Bates, 2008; Gelman & Hill, 2007) using R library lme4 (Bates, Maechler, Bolker, & Walker, 2014). We used in total five dependent variables – reading times, liking and recognition answers, and decision times for the liking and recognition answers. For the reaction times dependent variables, we used their logarithms in all analyses. The log-transformation led to distributions close to normal.[1] We also removed outlying trials that were three or more standard deviations from the mean. For all dependent variables, we further excluded trials on which the reading time was three or more standard deviations below the mean (less than 217 ms) because we assumed that participants could not have read the word properly on these trials.

We included centered pronounceability, its squared value, previous presentation, and presentation order as predictors in all analyses. The presentation order and previous presentation were recoded to a -0.5 to 0.5 scale. We also computed partial

[1] Descriptive statistics were computed using non-transformed response times.

autocorrelations for each participant for both reading and decision times. The average partial autocorrelation was significantly higher than zero up to the lag of five trials for both reading and decision times. We thus included the reaction times on five previous trials as predictors in all reaction times analyses to account for autocorrelation in the data (Baayen & Milin, 2015). They are not reported in the results because they were not of interest for our hypotheses. For the decision times, we also included the answer as a predictor. While not of primary interest for the present study, we were also able to test a previously studied effects of the dominant hand use on answer and reaction times (Casasanto, 2014). We thus included dominant hand use and answer on dominant hand side as predictors in analyses where they were relevant.

We built the final models from an initial model which included all fixed factors and random intercepts for participants and words. We then added random intercepts to this model and checked whether the new random intercepts significantly improve the model. Those that significantly improved the model were then sequentially added to the analysis alongside with their correlations with the other random factors and we checked at each step whether the inclusion improves the model (Bates, Kliegl, Vasishth, & Baayen, 2015). Only the results of the final model are reported. However, the other models yielded mostly similar results.

Reading time

The final model of log-transformed reading times included presentation order, previous presentation, and pronounceability random slopes for participants. The median reading time was 1162 ms (*IQR* = 809 ms). Previously presented items were read faster than those that had not been presented before, $t(198.1) = -2.59$, $p = .01$, $b = -0.015$, 95% CI = [-0.027, -0.004], $Mdn_{old} = 1159$ ms, $Mdn_{new} = 1165$ ms. The time needed to read a word was longer in later trials, $t(176.3) = 3.87$, $p < .001$, $b = 0.081$, 95% CI = [0.040, 0.122]. No effect on reading time was found for either pronounceability, $t(127.2) = 0.14$, $p = .89$, $b = 0.000$, 95% CI = [-0.006, 0.007], or squared pronounceability, $t(118.6) = -0.77$, $p = .44$, $b = -0.002$, 95% CI = [-0.007, 0.003]. While previous exposure shortened reading times, pronounceability did not play a role.

Liking answers

To analyze the effect of processing fluency on liking, we built a generalized mixed-effect model with a logit link function. The liking answer served as a binomial dependent variable. Apart from the above mentioned predictors common to all analyses, we included in the model whether the "like" answer corresponded to a key on participant's dominant hand side. The final model included pronounceability and presentation order random slopes for participants. The model showed that previously presented, $z = 5.98$, $p < .001$, $OR = 1.34$, 95% CI = [1.22, 1.48], $P_{(like|old)} = 0.518$, $P_{(like|new)} = 0.454$, as well as easier pronounceable words, $z = 10.29$, $p < .001$, $OR = 1.51$, 95% CI = [1.40, 1.63], were liked more. Both forms of fluency therefore positively influenced liking of words. Negative effect of pronounceability squared suggested that the effect of pronounceability was somewhat stronger for harder-to-pronounce words, $z = -1.75$, $p = .08$, $OR = 0.96$, 95% CI = [0.91, 1.01]; however, the effect was not significant. The effect also did not suggest inverted U-shaped relationship of pronounceability and liking since the estimated effect of pronounceability was monotonically increasing on the range of possible values. The association between pronounceability and liking on an item level is displayed in Figure 2. Neither the effect of order, $z = -0.55$, $p = .58$, $OR = 0.95$, 95% CI = [0.78, 1.15], nor of the answer on participant's dominant hand side, $z = 0.01$, $p = .99$, $OR = 1.00$, 95% CI = [0.91, 1.10], was significant.

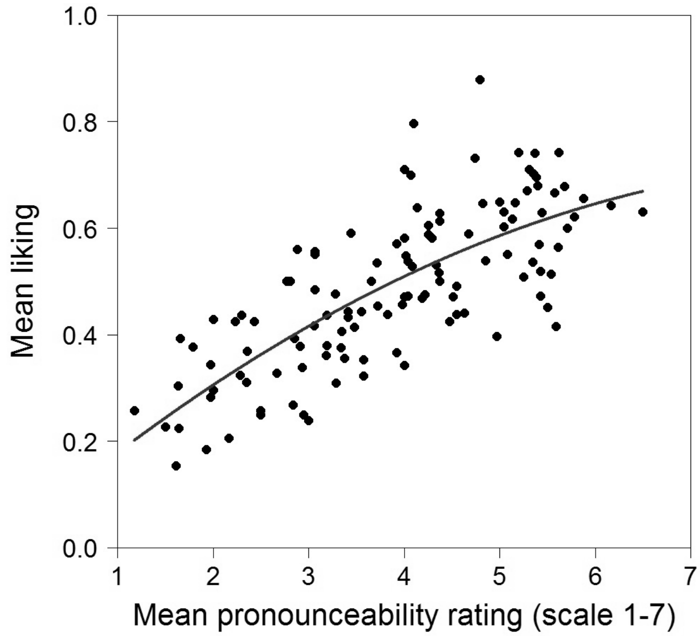

Figure 2: The effect of pronounceability on liking answers. The figure depicts the association between pronounceability and liking answers on an item level. The regression curve was computed using linear regression without any other predictors and with the proportion of "like" answers as a dependent variable.

We did not add reading times as a predictor in the described model because they can be influenced by the same factors as liking answers. However, we built a new model including the logarithmized reading times to check whether reading times may still predict liking since they are sometimes taken as a proxy of objective fluency. They did not, $z = 0.73$, $p = .46$, $OR = 1.05$, 95% CI = [0.93, 1.18].

Recognition answers

Recognition answers were analyzed similarly as liking answers. The final model included random slopes only for the presentation order. Participants were more likely to say that they had seen words which were previously presented, $z = 17.45$, $p < .001$, $OR = 2.38$, 95% CI = [2.16, 2.62], $P_{(seen|old)} = 0.524$, $P_{(seen|new)} = 0.335$, showing the ability to correctly recognize the previously presented words corresponding to the 59.4% overall correct response rate. Pronounceability did not predict recognition answers, $z = 1.16$, $p = .24$, $OR = 1.05$, 95% CI = [0.97, 1.13], but squared pronounceability did, $z = -1.98$, $p = .05$, $OR = 0.94$, 95% CI = [0.89, 1.00]. The effect suggests an inverted-U shaped relationship between pronounceability and probability of the “seen” answer, which is depicted in Figure 3. The model estimated the highest probability of answering that the word was previously presented for the pronounceability rating 4.27 on the 1-7 scale. Participants were more likely to say that they did not see the presented word in later trials, $z = -3.10$, $p = .002$, $OR = 0.75$, 95% CI = [0.62, 0.90], which may be a result of memory decay, and less likely to answer that they had seen the word if this answer was on their dominant hand side, $z = -1.92$, $p = .06$, $OR = 0.91$, 95% CI = [0.83, 1.00].

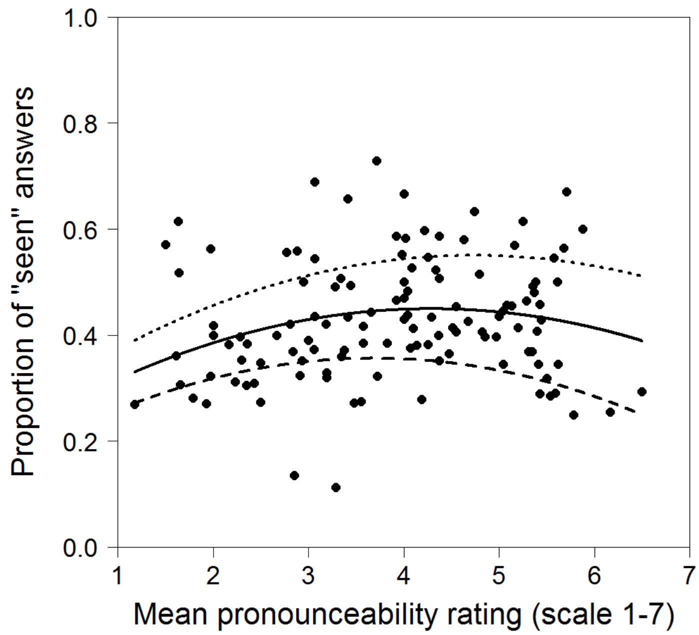

Figure 3: The effect of pronounceability on recognition answers. The figure depicts the association between pronounceability and recognition answers on an item level. The regression curves were computed using linear regression without any other predictors and with the proportion of “seen” answers as a dependent variable. The solid curve corresponds to regression computed with all trials, the dotted curve presents regression only for trials where the word had been previously presented, and the dashed curve only for trials where the word had not been previously presented.

We again estimated the effect of reading times by adding logarithmized reading times as a predictor in the final model. The results showed that participants were more likely to say that they had previously seen the word on trials with shorter reading times, $z = -5.41$, $p < .001$, $OR = 0.72$, 95% CI = [0.64, 0.81]. Adding reading times also improved fit of the simpler model without them, $X^2(1) = 30.09$, $p < .001$.

Next, we analysed whether pronounceability can influence not only the tendency to answer that the presented word had been previously seen, but also correctness of the answers. Therefore, we included the interactions between previous presentation and both pronounceability factors in the model. The results suggested that pronounceability influenced correct recognition; however, we found only linear effect of pronounceability, not quadratic. In particular, the effect of previous presentation was larger for easier pronounceable words as indicated by the interaction between previous presentation and pronounceability, $z = 2.83$, $p = .005$, *ratio of OR* = 1.12, 95% CI = [1.04, 1.22]. This can be seen as the increasing spread between the blue and green regression curves in Figure 3. The interaction between previous presentation and pronounceability squared was not significant, $z = 0.26$, $p = .79$, *ratio of OR* = 1.01, 95% CI = [0.95, 1.07]. The lack of the interaction suggests that the quadratic effect of pronounceability does not influence correctness of the answer. Rather, it only biases participants to say less often that they had seen the most easy- and hard-to-pronounce words. On the other hand, easier pronounceable words are better recognized, but the linear effect of pronounceability does not influence the tendency to answer that the word had been shown before. The estimates of the other effects did not change considerably and are thus not reported for simplicity.

Liking decision time

The final model of log-transformed liking decision times included for participants random slopes for presentation order, previous presentation, dominant hand use during answering the question, and answer. Liking decision times did not depend on previous presentation, $t(196.4) = -1.01$, $p = .31$, $b = -0.007$, 95% CI = [-0.020, 0.006], but they were faster in later trials, $t(161.1) = -4.01$, $p < .001$, $b = -0.057$, 95% CI = [-0.084, -0.029]. While the effect of dominant hand use answering the question seemed to differ between participants, it did not influence liking decision times in general, $t(194.1) = -$

0.51, p = .61, b = -0.003, 95% CI = [-0.016, 0.009].[2] Decision times did not differ between "like" and "dislike" answers, $t(205.9)$ = -0.64, p = .52, b = -0.005, 95% CI = [-0.018, 0.009]. Better pronounceability was associated with somewhat slower reaction times, but the effect was not significant, $t(124.3)$ = 1.81, p = .07, b = 0.005, 95% CI = [-0.000, 0.011]. Most importantly, we found an inverted U-shaped effect of pronounceability on liking decision times as indicated by the negative squared pronounceability effect, $t(112.3)$ = -2.72, p = .008, b = -0.006, 95% CI = [-0.010, -0.002]. The model suggested that the slowest decision times were for pronounceability of 4.34 (i.e., close to the midpoint of the scale).[3] The association between pronounceability and liking decision time on an item level can be seen in Figure 4.

[2] The effect of the answer on the side of the dominant hand was not significant for a simpler model and was removed from the analysis due to convergence issues occurring in model estimation.

[3] We also built a model including interaction of previous presentation and both pronounceability factors. Neither interaction was significant and the pronounceability effect was virtually the same as in the simpler model.

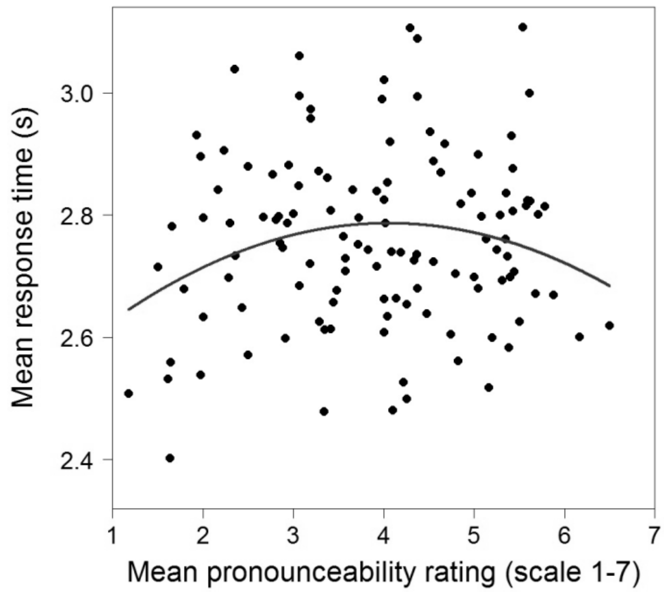

Figure 4: The effect of pronounceability on liking decision time. The figure depicts the association between pronounceability and liking decision times on an item level. The regression curve was computed using linear regression without any other predictors and with non-transformed decision times as a dependent variable.

Recognition decision time

The final model of log-transformed recognition decision times included the presentation order and chosen answer as random slopes for participants. Previous presentation did not influence recognition decision times, $t(6863.0) = -1.14$, $p = .26$, $b = -0.008$, 95% CI = [-0.021, 0.005]. Contrary to decision times for liking, later trials were associated with slower decision times for recognition, $t(138.1) = 2.95$, $p = .004$, $b = 0.040$, 95% CI = [0.013, 0.067]. Decision times were influenced neither by use of dominant hand in answering, $t(6872.8) = 1.17$, $p = .24$, $b = 0.008$, 95% CI = [-0.005, 0.021], nor by the key associated with the “seen” answer, $t(6874.5) = 0.89$, $p = .37$, $b = 0.006$, 95% CI = [-0.007, 0.019]. The responses were faster when participants answered that they had previously seen the word, $t(214.6) = -11.54$, $p < .001$, $b = -0.094$, 95% CI = [-0.110, -

0.078], Mdn_{seen} = 2251 ms, $Mdn_{not\ seen}$ = 2520 ms. Finally, better pronounceability led to higher decision times, $t(121.6) = 2.37$, $p = .02$, $b = 0.007$, 95% CI = [0.001, 0.013], but squared pronounceability had no effect, $t(121.5) = -0.41$, $p = .69$, $b = -0.001$, 95% CI = [-0.005, 0.003]. The association between pronounceability and recognition decision times is depicted in Figure 5. The linear effect of pronounceability suggests a possible speed-accuracy tradeoff since participants were more accurate, but slower when deciding about easier-to-pronounce words.

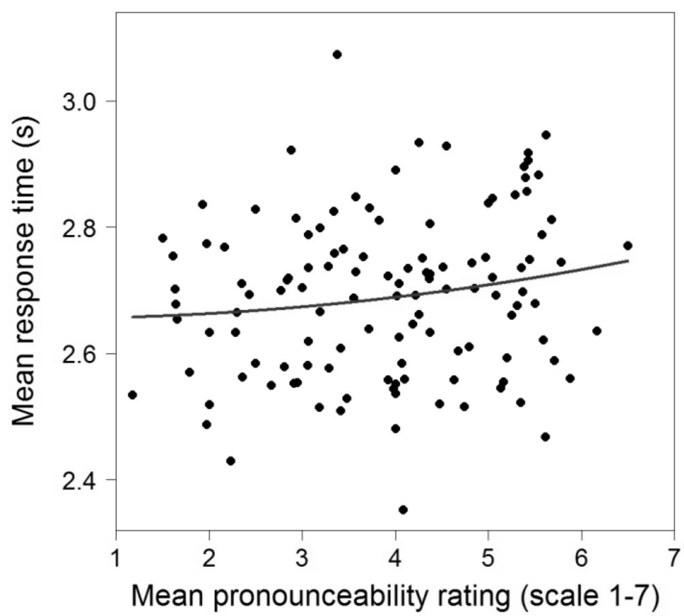

Figure 5: The effect of pronounceability on recognition decision time. The figure depicts the association between pronounceability and recognition decision times on an item level. The regression curve was computed using linear regression without any other predictors and with non-transformed decision times as a dependent variable.

Discussion

The procedure used in the present study allowed us to separate reading and decision times, which are usually confounded in studies of processing fluency. We expected that harder pronounceability would lead to slower reading times, but that pronounceability may serve as a strong cue when making decisions about difficult-to-pronounce words, resulting in faster decision times for difficult-to-pronounce words than for more easily pronounceable words. While we did not find the effect of pronounceability on reading times, we showed the predicted inverted-U shaped effect of pronounceability on liking decision times. Pronounceability influenced liking of words and words that differed in pronounceability from others were thus easily judged as either liked in the case of easy-to-pronounce words, or disliked in the case of difficult-to-pronounce words. The effect shows that it is important to consider in which part of the judgmental process fluency plays a role. It is also not possible to simply equate fluency with shorter response times unless the response times relate only to a single cognitive process.

Similarly as in previous research (for a review, see Winkielman, Schwarz, Fazendeiro, & Reber, 2003), we found the effect of fluency on liking. Moreover, the effect was present for both types of fluency. As in previous studies, we found the effect of previous presentation on liking, also known as the mere-exposure effect (Zajonc, 1968), and the effect of pronounceability on liking (Laham, Koval, & Alter, 2012). Our experiment thus replicated the previously found effects using pseudowords as stimuli. A recent study suggested that fluency does not generally lead to positive evaluation, but merely amplifies emotions associated with given stimuli (Albrecht & Carbon, 2014). However, the present study used neutral pseudowords without any meaning and both “like” and “dislike” options. We still found the positive effect of fluency on liking which might suggest that apart from the amplification of affect, fluency also has some general positive effect on liking.

Given the effect of pronounceability on liking, it is not straightforward how to explain the lack of the effect of pronounceability on reading times. One possibility is that pronounceability is related just to one process comprising reading and that fluency of the other processes is not similarly influenced as pronounceability. Furthermore, participants did not read out loud the words in the present experiment, so reading speed might not have be so strongly related to pronounceability as it would have been if participants had to actually pronounce the words. Still, the results speak against a simple association between pronounceability, fluency, and response times. If we assume that pronounceability is a form of fluency, then the absence of the effect on reading times shows that processing fluency cannot be just equated with the speed of the given process. That is further shown by the finding that the trials in which words were read faster were not associated with higher liking of the words. Unlike pronounceability, previous presentation influenced both reading times and reading times were associated with recognition answers. Participants were more likely to indicate that they had previously seen the word if they read it faster. However, this association should be interpreted with caution because both reading times and recognition answers can be just influenced by a same factor and there does not have to be necessarily a causal connection between them. It is also noteworthy, that while previous presentation shortened reading times, the effect was just 6 ms with the median reading time 1162 ms. The present study was not designed to test the association of processing fluency and speed of the process and their downward influence on judgment, but it suggests that the topic deserves attention in future research.

The recognition answers were affected by previous presentation. However, it is not possible to separate the effect of fluency and recollection in our design. The effect of previous presentation on recognition answers could have been caused entirely by correct recollection of the previously presented stimuli. While previous presentation influenced the recognition answers as expected, the relationship between pronounceability and recognition was more complex. We did not find the predicted linear effect of pronounceability on recognition answers which would suggest misattribution of fluency associated with easily pronounceable words to previous presentation. However, we found an inverted-U shaped relationship of pronounceability and recognition answers. That is, easy and hard pronounceable words were more likely to be considered as new in comparison to words moderate in pronounceability. As a speculation, it is possible that participants used a metamemory strategy inferring that if they cannot recollect these salient words, they had not seen them. Apart from the quadratic effect of pronounceability on recognition answers, we also found an interaction between pronounceability and previous presentation. Easily pronounceable words were associated with a higher probability of correct recognition answers. Some previous research suggested that disfluency is associated with deeper processing (Alter, Oppenheimer, Epley, & Eyre, 2007; Song & Schwarz, 2008; but see also Meyer et al. 2015; Thompson et al. 2013, for related null effects) which might have implied that participants could have remembered the disfluent words better. However, we found the opposite effect.

Most importantly, the inverted-U shaped relationship of pronounceability and liking decision times supported our prediction that disfluency may be used as a strong cue for disliking and thus lead to faster decisions. It is noteworthy that given the lack of the effect of pronounceability on reading times, the total duration of reading and deciding was lower for disfluent words than for words moderate in fluency. This shows that response times cannot be simply equated with fluency if they do not pertain just to a single cognitive process. While the result has important ramifications for future fluency research, it is not clear how it affects previous studies using response times as a proxy for fluency. Under certain conditions, the relationship of fluency and response times will be monotonically decreasing. The conditions can be inferred from the diffusion decision model. For example, when the drift rate does not change its sign depending on fluency, the relationship of fluency and decision times will not have the inverted-U shape observed in the present study. Similarly, when participants have only one possible answer available, the relationship will be monotonic. It is also possible that in circumstances when fluency is not continuous, but has only a limited number of discrete levels, the non-monotonic relationship might be obscured.

While we did not expect that pronounceability will have the inverted-U shaped effect on decision times only for liking and not for reading, the result is consistent with the lack of a linear effect of fluency on recognition answers. The association of the extremes of pronounceability with shorter decision times requires that participants use pronounceability as a cue in judgment. Since recognition answers were not significantly influenced by pronounceability, the lack of the inverted-U shaped effect in recognition decision times is not surprising.

While not of primary interest in the present study, we had available data about participants' dominant hand and we were thus able to assess effects of the assignment of an answer to the dominant hand side. We did not find that participants were faster responding using their dominant hand. While a previous study (de la Vega, de Filippis, Lachmair, Dudschig, & Kaup, 2012) showed that people are faster to classify words as positive with dominant hand and as negative with non-dominant hand, we found no effect of the answer on participant's dominant hand side on speed of an answer. It is possible that the different results were due to differences in the tasks. Whereas in the study by de la Vega et al. (2012), participants only classified words as positive or negative, in our study participants had to judge their liking of words. Furthermore, de la Vega et al. (2012) argue that the effect of response side on reaction times occurs only if the mapping of valence to a side is made salient, which was not the case in the present experiment.

In summary, the present experiment showed that retrieval and phonological fluency had different effects on reading times and judgment. While previous presentation shortened the reading times, pronounceability had no effect. Both types of fluency led to higher liking, but only the previous presentation influenced the recognition answers. Finally, we found an inverted-U shaped effect of pronounceability on liking decision times which shows that disfluency may lead to faster decisions.

CHAPTER III – If it's difficult to pronounce, it might not be risky

Introduction

One of the studies of judgmental effects of processing fluency showed that people perceive food additives and amusement-park rides with harder-to-pronounce names as riskier (Song & Schwarz, 2009; hereafter referred to as S&S). People tend to avoid risk; therefore, a possible explanation for the association between processing fluency and safety is that people encounter safe objects more often which increases their familiarity and leads to more fluent processing. People thus learn a naive theory (Alter & Oppenheimer, 2009) that fluency is associated with safety, which they use in judgment. Consistently, S&S found that hard-to-pronounce names are also judged to be more novel and that novelty mediates the effect of pronounceability on judgment of risk.

Some studies show that fluency can have different effects on judgment depending on context (Galak & Nelson, 2011; Pocheptsova, Labroo, & Dhar, 2010). This suggests that naive theories about the association of fluency and a judged attribute may differ for specific categories of objects. Initially, we attempted to build on these findings and explored a hypothesis about context dependence of the fluency–safety association (studies 1-5). That is, we tested whether people may associate fluency with risk under some circumstances. We expected that this may happen for categories of objects where people encounter riskier exemplars more often.

However, following the methods used by S&S, the initial studies (1-4) treated stimuli as a fixed factor. Consequently, we found some initial support for the context dependence of the fluency-safety association, 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 the findings by S&S to newly generated stimuli in three further studies (5-7). In these studies, we used randomly constructed and sampled stimuli, which ensured that there was no possibility of bias in their selection. In studies 5 and 6, we replicated the original fluency–safety association using the original stimuli by S&S. However, we were unable to replicate the effect using new stimuli. Using materials from a different experiment in S&S, Study 7 showed a different pattern of results – we found the association between pronounceability and safety with newly sampled stimuli, but not with the original stimuli by S&S. However, the association of pronounceability and safety was confounded with word length and the effect disappeared after controlling for word length. Therefore, in line with studies 5 and 6, even the last study casts doubt on the existence of a generalizable association between pronounceability and perceived risk and illustrates the importance of stimuli sampling.

Studies 1-4

In the first four studies, we examined the hypothesis that the association between processing fluency and safety may be dependent on the category of an evaluated object. While 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 dangerous criminals in the news more often than names of less dangerous criminals. Similarly, cities in a war zone are more likely to be mentioned in the news if fighting has occurred there. People may therefore learn the opposite association between fluency and risk for these categories of objects. They might use it then in judgment similarly as they use the more common association between fluency and safety for other categories of objects.

Method[4]

All four studies shared the same general procedure adopted from S&S. Participants were given one or two scenarios describing a hypothetical situation in which they encountered ten exemplars of a certain category. Then, they judged dangerousness of the exemplars on a 7-point scale (very safe - 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 a half was relatively easy to pronounce (e.g., Allotoneline, Magnalroxate) and the other half was hard to pronounce (e.g., Ribozoxtlitp, Nxungzictrop).

We used different participant populations for each study (see Figure 1). We did not continue data collection after analyzing data, with the exception of Study 2 which is pooled from two data sets.

We used in total four hypothetical situations and categories of stimuli. In the *food additives* scenario (adopted from S&S), participants were told to imagine reading names of food additives on a food label. The *cities in a war zone* and *criminals* scenarios were constructed such that we expected that people would judge easier-to-pronounce names as more dangerous. In the cities in a war zone scenario, participants were told to imagine that they travel through war-stricken Syria and judge dangerousness of cities

[4] Additional details of the procedure, exact wording of materials, data, R scripts for analysis, as well as pre-registrations of studies 3-7 can be found on osf.io/fjs56/.

they travel through. In the criminals scenario, participants were told to rate 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, only "Syria" was replaced by "Turkish Riviera" and beach resorts were rated instead of cities.

We directly asked a separate sample of participants whether they expect exemplars of given categories with easy-to-pronounce and familiar names to be more or less dangerous. While participants expected exemplars of prisoners, cities in a war zone, and poisons to be more dangerous if they had easy-to-pronounce or familiar names, they held the opposite expectation for food additives, tourist destinations, roller coasters, and medicines (see Table 1).

Category	Easily pronounceable –dangerous [scale 1 (less dangerous) - 6 (more dangerous)]	Familiar – dangerous [scale 1 (less dangerous) - 6 (more dangerous)]
American airlines	2.52	2.15
Tourist destination	2.59	2.00
Plant	2.68	2.32
Human	2.71	2.09
Factory of a firm	2.71	2.58
Scientist	2.78	1.89
Medicine	2.83	2.31
Animal	2.86	2.31
Roller coaster	2.97	2.59
Food additive	3.00	3.28
Disease	3.11	4.00
Indian airlines	3.14	3.11
Nuclear power plant	3.19	3.51
Parasite	3.37	4.31
Politician	3.43	3.30
Tropical disease	3.44	4.24
Poison	3.62	4.64
Prisoner	3.68	4.64
City in a war zone	3.75	4.58
Serial killer	3.89	5.11
Poison gas	3.97	4.65

Table 1: Naïve theories about assocciations between pronounceability and familiarity and dangerousness. The table contains average ratings of naïve theories

about dangerousness of exemplars of various categories that have names that are easily pronounceable or familiar.

Results and discussion

In all studies, a small part of participants judged most items using the same rating. This behavior can be perceived as non-compliance 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 based on judgment of an author blinded to participants' condition in the first two studies (1 and 2) and according to pre-registered exclusion criteria for the later studies (3 and 4). Given the focus of the remainder of the chapter, the exclusion criteria as well as examination of additional factors (e.g., instructions to read the names carefully, order effects, etc.) are not highly relevant and were thefore not included. However, further information can be found on osf.io/fjs56/.

The results of the four studies are shown in Figure 5. We were able to replicate the results of S&S using their original materials in two studies (2 and 3). While Study 2 suggested that the effect of pronounceability on judgment of riskiness might be reversed for some categories of objects, we were not able to obtain the same effect in further studies (3 and 4). In fact, we obtained the effect in the original direction (easy-to-pronounce – safe; Study 3) even when using the same scenario where 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 be dependent on the particular items used. Furthermore, similarly as S&S, studies 1-4 incorrectly treated stimuli as a fixed factor which precludes the possibility to generalize their results. This shortcoming was addressed in studies 5-7.

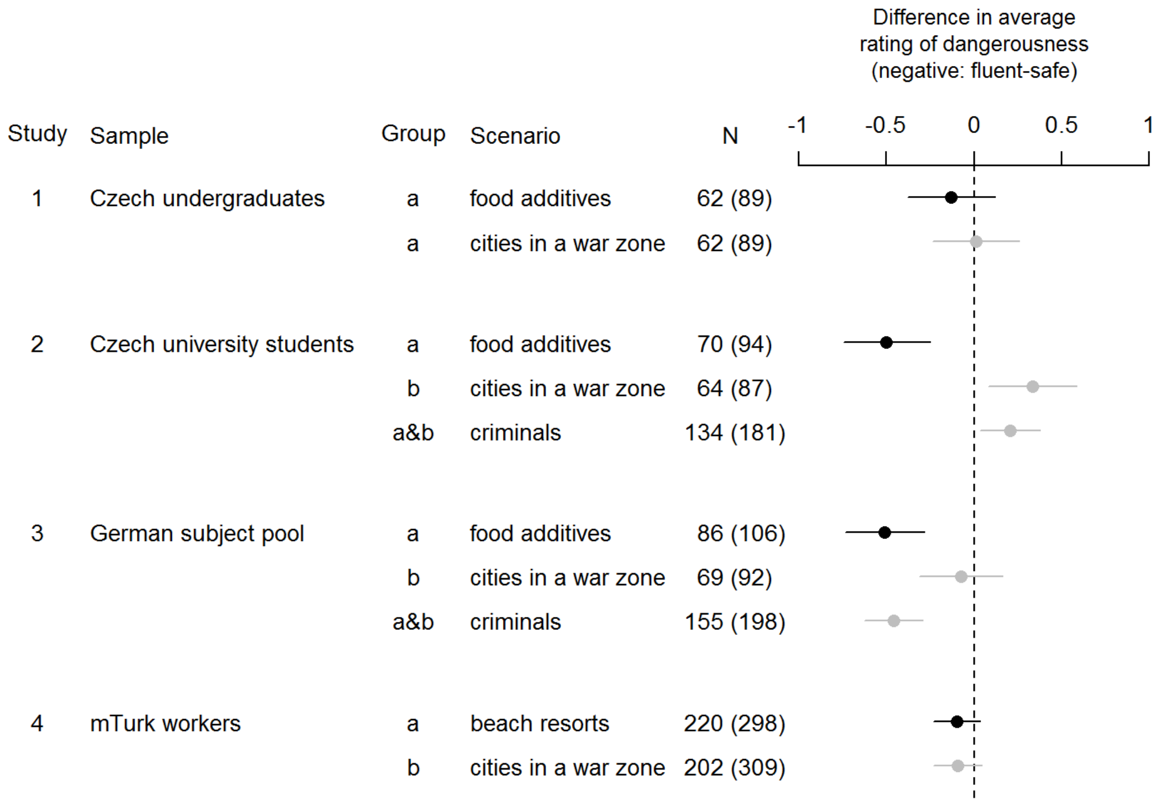

Figure 5. The figure shows results of the first four studies. In some studies, participants were divided in two separate groups (labeled *a* and *b*) and some scenarios were given to both groups (indicated as *a&b*). Points represent Cohen's d_zs for the average difference in ratings of easy- and hard-to-pronounce names. Lines represent 95% confidence intervals for Cohen's d_z. Negative values of Cohen's d_z mean that easy-to-pronounce names were judged to be less dangerous (i.e., the original association observed by S&S). Effect sizes are displayed in black for scenarios where we expected the original association and in grey for scenarios where we expected the opposite association (i.e., easy-to-pronounce names to be judged as more dangerous). The number of participants before exclusion is provided in parentheses. Descriptions of scenarios can be found in the Method section.

Study 5

The results of the first four studies indicated that fluency effects can strongly depend on particular items used. Therefore, in addition to studying the possibility of reversal of the fluency–safety association, in Study 5 we also directly compared the items originally used by S&S with similar newly constructed items.

Method

We recruited 616 mTurk workers to participate in the study. Following pre-registered exclusion criteria, 44 participants with more than one missing data point or using one rating more than 7 times were excluded.

To explore the influence of particular items, we used the names from the original food additives scenario and added 50 new items (e.g., Enzalutmmide, Griseofplvin). Each participant was given 10 randomly selected names out of the 60. The new items were constructed by taking 12 letters long names of existing medications, randomly changing two letters in the names, and removing names that sounded too similar to well-known substances (e.g., Tedtosterone). We used a new scenario where participants imagined that they were members of a team of scientists searching through archives of a laboratory that had researched either poisons, or medicines (depending on participant's condition), and judged their harmfulness on a scale from 1 (harmless) to 7 (very harmful) based on 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). For easier comparability with the other studies, we reversed the ratings such that the pronounceability variable was higher for easier-to-pronounce names. A negative slope for pronounceability therefore indicates the association between disfluency and harm. We also centered the variable 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 names were not judged to be significantly more harmful, $t(65.4) = -0.91$, $p = .37$, $b = -0.12$, 95% CI = [-0.38, 0.14]. 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 stronger for the names used in the original study by S&S than for the newly constructed names, $t(56.2) = -1.45$, $p = .15$, $b = -0.35$, 95% CI = [-0.81, 0.12]. However, neither of these effects was significant. 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 and newly constructed names, $t(56.0) = -1.20$, $p = .24$, $b = -0.18$, 95% CI = [-0.47, 0.11]. Including random slopes for pronounceability in a model as recommended by Barr, Levy, Scheepers, and Tily (2013) does not change the results of this or the next study.

While the interaction between the source of names and pronounceability was not significant, we were interested whether the effect of pronounceability on judgment of harmfulness may differ between the original and newly constructed names. Therefore, we conducted the analysis for the original and newly constructed names separately. The results showed a significant pronounceability effect for the original names, $t(12.8) =$

-3.02, *p* = .01, *b* = -0.48, 95% CI = [-0.79, -0.17], but not for the newly constructed names, *t*(58.8) = -0.79, *p* = .43, *b* = -0.11, 95% CI = [-0.39, 0.16]. The interaction between the category of judgment and pronounceability was not significant for either of the two sources of names. Figure 6 shows the effect of pronounceability on harmfulness ratings on the item level.

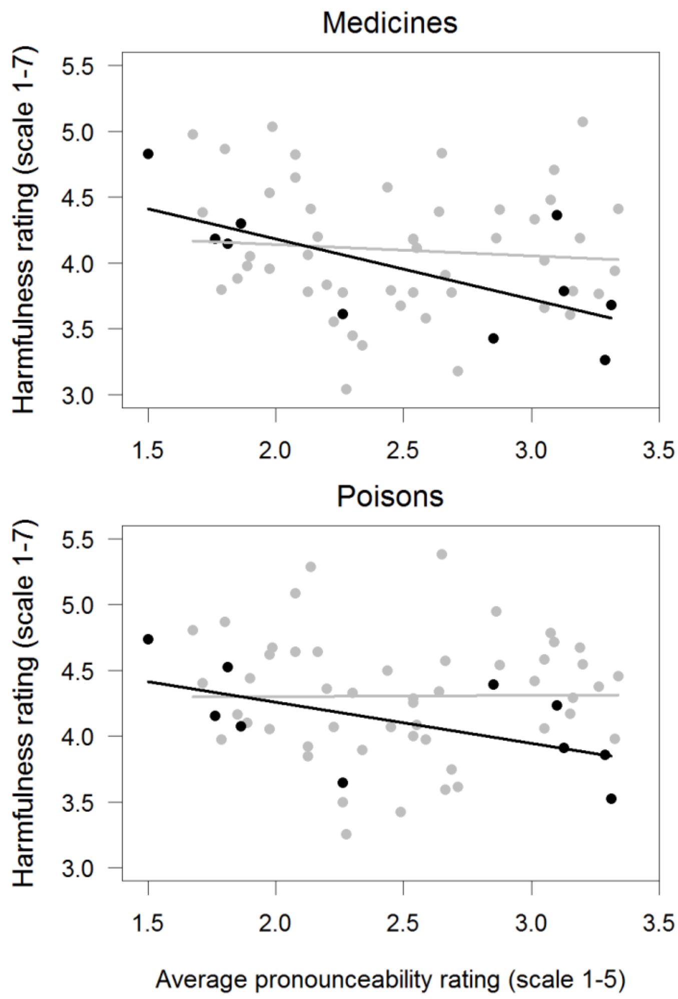

Figure 6. The figure shows the association between pronounceability and harmfulness ratings on the item level. While there is almost no association for the newly constructed names (grey points and grey line), it is possible to see the disfluency–harmfulness association for the original names (black points and black line). Note that the axes are truncated and higher pronounceability ratings correspond to easier pronunciation.

Discussion

While we found that the effect of disfluency on judgment of harmfulness was somewhat higher for medicines than for poisons, the interaction was not significant and the effect was not in the opposite direction for poisons. The result of Study 2 suggesting that the effect of fluency on judgment of harmfulness may be reversed for some categories of stimuli was therefore not corroborated.

Interestingly, the results suggested that the effect of fluency on judgment of harmfulness might be limited only to the original names used by S&S. As in studies 2 and 3, we were again able to replicate the effect for the original names, even when using different scenarios. However, we found no effect for the newly constructed stimuli. Since the interaction of the source of names 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 to compare results of a direct replication of S&S (with original items) and a replication using the same materials, but newly constructed items.

Method

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

Participants were given the food additive scenario adopted from S&S and judged harmfulness of 15 additives on a 7-point harmfulness scale. Similarly as in Study 5, we used 10 names from the original study and additionally constructed 40 new names. The new names were constructed using the list of 12 letters long names of medicines from Study 5, randomly substituting 2 letters in each word, randomly selecting 10 letter continuous string from this newly constructed name, and appending a 2 letter suffix from a list of suffixes that we based on a list of Czech food additive names. The constructed names varied in pronounceability and were similar to the original names used in Song & Schwarz (2009).

All participants rated harmfulness of randomly selected 10 newly constructed names and 5 original names of additives. Next, participants were divided into two groups. The first group (*N* = 105; 100 after exclusion) rated novelty of 10 randomly selected newly constructed names and 5 remaining original names of additives. The second group (*N* = 95, 86 after exclusion) rated pronounceability of 20 newly constructed names and 5 remaining original names.

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]. Similarly as in Study 5, the effect of pronounceability on judgment of harmfulness was stronger for the original names than for the newly constructed names, *t*(39.9) = -1.70, *p* = .10, *b* = -0.12, 95% CI = [-0.26, 0.02], even though not significantly. We again conducted the analysis separately for the original and newly constructed names: While participants judged harder-to-pronounce names as more harmful when judging the original names, $t(7.9) = -2.54$, $p = .03$, $b = -0.15$, 95% CI = [-0.27, -0.04], there was no effect for the newly constructed names, $t(39.5) = -0.59$, $p = .56$, $b = -0.03$, 95% CI = [-0.11, 0.06].

Similar analysis for novelty ratings showed no interaction between the source of names and pronounceability, $t(41.7) = -0.30$, $p = .77$, $b = -0.03$, 95% CI = [-0.20, 0.14]. However, the analysis showed that easier-to-pronounce names were judged as less novel, $t(46.9) = -3.81$, $p < .001$, $b = -0.21$, 95% CI = [-0.31, -0.10]. When the analysis was conducted separately for the original and newly constructed names, we found that easier-to-pronounce names were judged as less novel for both original names, $t(7.3) = -4.14$, $p = .004$, $b = -0.23$, 95% CI = [-0.34, -0.12], and newly constructed names, $t(38.2) = -3.65$, $p < .001$, $b = -0.20$, 95% CI = [-0.31, -0.09]. Figure 7 shows the effect of pronounceability on harmfulness and novelty ratings on the item level.

A possible concern may be that the newly constructed names might have retained some similarity to the names of medicines from which they were derived. To check this possibility, we asked a separate sample of 210 participants to assess to what degree the names reminded them of an existing substance on a 7-point scale (not at all - very). None of the names reminded the participants strongly of existing substances (all mean ratings were lower than 2.7) and there was no difference between the newly constructed and original names, $t(48.0) = 0.17$, $p = .87$, $b = 0.02$, 95% CI = [-0.21, 0.25]. We also checked the robustness of the conducted analyses of the pronounceability effect by excluding items with highest similarities from analysis. Excluding the items highest in similarity has no appreciable effect on the pronounceability effect either for the new or for old items. The result obtained in the study were therefore not caused by similarity of the newly constructed items to existing substances.

Discussion

The results together with the results of Study 5 suggest that the effect of pronounceability on judgment of harmfulness may be limited only to the names used by S&S. There was no evidence for the effect in newly constructed names. While the interaction between the source of names 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 names (Westfall, Kenny, & Judd, 2014). Furthermore, when the interaction effect was meta-analytically combined from both studies, it was significant, $z = 2.14$, $p = .03$, $r = .22$, 95% CI = [.02, .40]. On the other hand, the effect of pronounceability on perceived novelty is evident even for the newly constructed names, which suggests that a robust different fluency effect can be replicated even with the newly constructed stimuli.

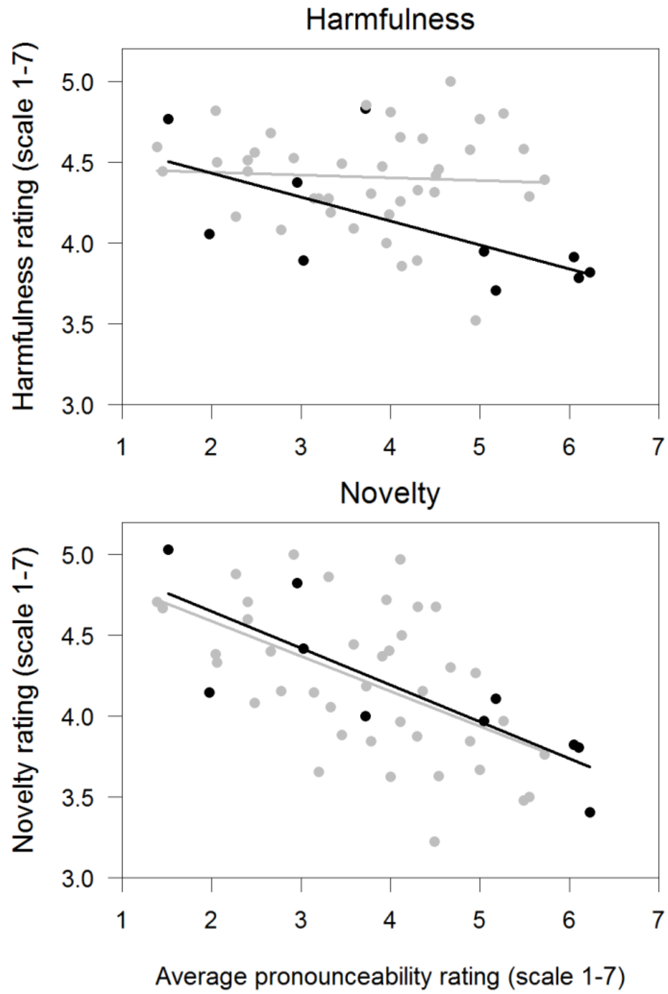

Figure 7. The figure shows the effect of pronounceability on harmfulness and novelty ratings on the item level. For harmfulness ratings, there is almost no effect for the newly constructed names (grey points and grey line), but it is possible to see the disfluency-harmfulness association for the original names (black points and black line). On the other hand, there is a clear effect of pronounceability on judgment of novelty for both original and newly constructed names. Note that the y-axis is truncated and higher pronounceability ratings correspond to easier pronunciation.

Study 7

Apart from the food additives scenario used in our studies 1-3 and 6, S&S also conducted an experiment using two scenarios where people had to judge how adventurous or risky they view amusement-park rides. Study 7 replicated this experiment using again the original items and newly constructed stimuli.

Method

We recruited 1042 mTurk workers to participate in the study. Following a pre-registered exclusion criterion, 60 participants using the same rating for all items in a scenario were excluded. We also excluded 32 participants who did not complete the whole study resulting in the final sample of 950 participants.

The study used two scenarios adopted from Study 3 in S&S. 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 too adventurous rides, 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). Each participant evaluated 11 items in both scenarios, which were presented in a random order.

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

An independent sample of 303 mTurk workers was given a random sample of 50 names out of the 206 for rating of pronounceability on a scale from 1 (easily pronounceable) to 7 (hard pronounceable). For easier comparability with the previous studies, we reversed the ratings such that the pronounceability variable was higher for easier-to-pronounce names. We also centered the variable by subtracting the mean of pronounceability ratings from all values. The name length was recoded to a scale from -0.5 to 0.5.

Results

A pre-registered analysis of ratings of rides showed that longer names were perceived as riskier, $t(230.3) = 5.18$, $p < .001$, $b = 0.71$, 95% CI = [0.44, 0.98]. There was no effect of pronounceability on riskiness ratings, $t(198.1) = 0.28$, $p = .78$, $b = 0.01$, 95% CI = [-0.07, 0.09]. Participants gave higher ratings in the scenario presented as second, $t(947.7) = 3.82$, $p < .001$, $b = 0.16$, 95% CI = [0.08, 0.24], and somewhat lower ratings in the undesirable risk scenario, $t(949.4) = -2.02$, $p = .04$, $b = -0.05$, 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 first presented scenario. The effect of pronounceability did not differ based on name length, $t(197.9) = -0.42$, $p = .67$, $b = -0.04$, 95% CI = [-0.20, 0.13], but it depended on the scenario, $t(18749.3) = 5.53$, $p < .001$, $b = 0.11$, 95% CI = [0.07, 0.14]; that is, it was stronger for the undesirable risk scenario. However, we did not include the interaction between name length and scenario in the model, so the interaction between pronounceability and scenario may be due to the association of name length and pronounceability. When the interaction between name length and scenario is included in the model, the interaction between pronounceability and scenario disappears, $t(19053.4) = 0.70$, $p = .48$, $b = 0.02$, 95% CI = [-0.04, 0.08].

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 Figure 8), we next analyzed the data only from the first presented scenario and for both scenarios separately. There was no association between pronounceability and risk for either the desirable risk scenario, $t(213.2) = 0.38$, $p = .70$, $b = 0.02$, 95% CI = [-0.09, 0.13], or the undesirable risk scenario, $t(180.4) = 0.29$, $p = .77$, $b = 0.01$, 95% CI = [-0.08, 0.11], when name length was taken into account, while name length still significantly predicted the ratings for both the desirable risk scenario, $t(240.1) = 4.98$, $p < .001$, $b = 0.89$, 95% CI = [0.54, 1.25], and undesirable risk scenario, $t(222.3) = 5.76$, $p < .001$, $b = 0.90$, 95% CI = [0.59, 1.21]. Furthermore, adding pronounceability as a predictor to a model with name length did not improve the model for either adventurousness, $X^2(1) = 0.74$, $p = .39$, or riskiness ratings, $X^2(1) = 0.15$, $p = .70$, while adding name length to a model with pronounceability resulted in a significantly better fit for both adventurousness, $X^2(1) = 28.47$, $p < .001$, and riskiness, $X^2(1) = 39.93$, $p < .001$. Therefore, the pronounceability effect that is seen without inclusion of name length as a predictor for both riskiness, $t(275.0) = 7.03$, $p < .001$, $b = 0.23$, 95% CI = [0.17, 0.29], and adventurousness ratings, $t(286.0) = 6.73$, $p < .001$, $b = 0.25$, 95% CI = [0.17, 0.32], can be entirely driven by the association of pronounceability with name length. Visual inspection of the data (see left graphs in Figure 8) suggests that the relationship of pronounceability and risk may be present for names 7 and 13 letters long. However, analysis including the pronounceability effect separately for all name lengths does not yield any significant pronounceability effect. In summary, 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 S&S used only 3 fluent and 3 disfluent names and all disfluent names were longer than the fluent names. It was therefore not reasonably possible to evaluate the difference between the original and newly constructed names. However, it is noteworthy that we found no association between pronounceability and riskiness ratings for the original names for either of the scenarios. Mean riskiness ratings averaged across both scenarios were 4.98, 4.19, and 4.35 for the fluent names and 4.04, 4.69, and 4.57 for the disfluent names. Curiously, while we were able to replicate the results of S&S in studies 5 and 6 only using the original names from their study, Study 7 obtained the opposite pattern of results – we found the association of pronounceability and perceived risk with newly sampled items, but not with the original items used by S&S.

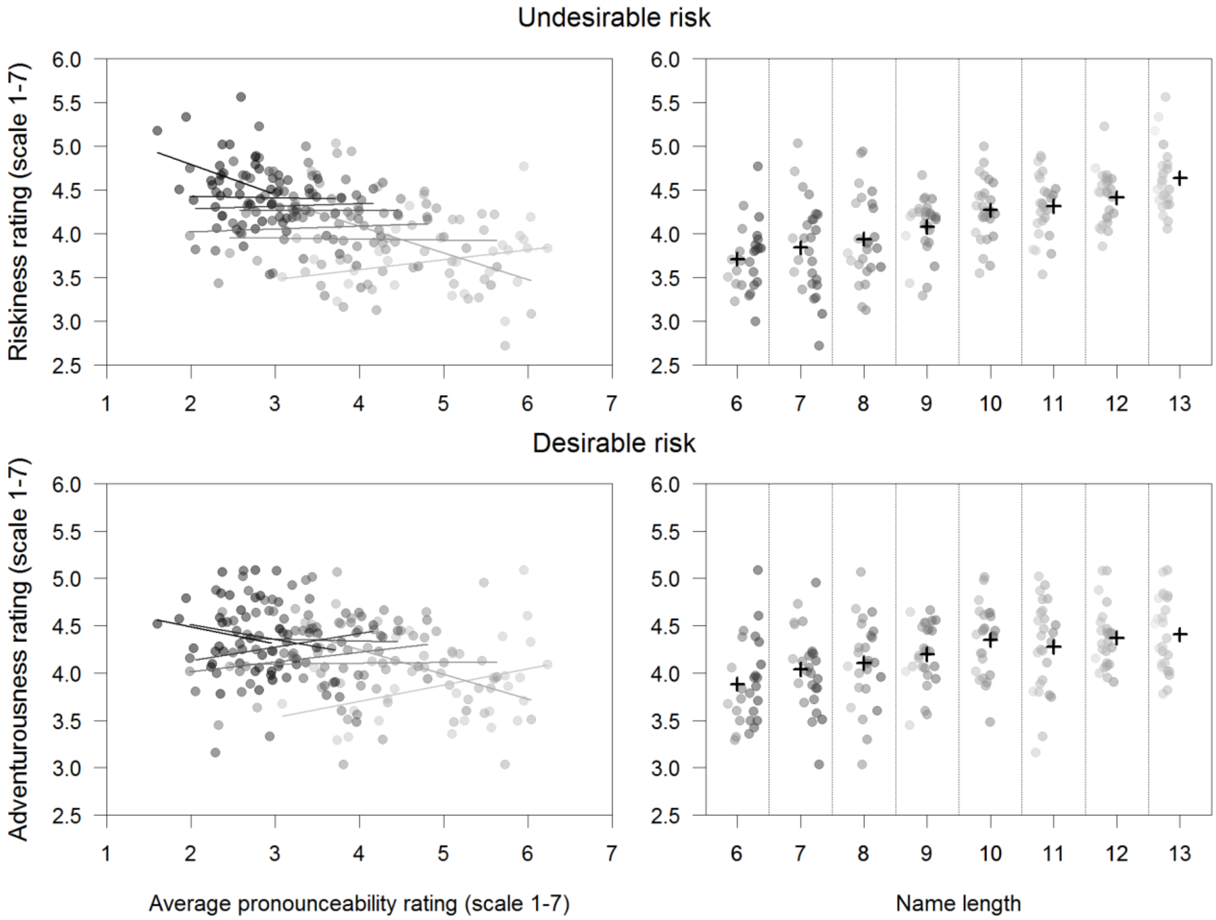

Figure 8. The figure shows the effect of pronounceability (*left*) and name length (*right*) on riskiness (*top*) and adventurousness (*bottom*) ratings on the item level. It is possible to see that pronounceability was associated with both riskiness and adventurousness ratings. However, looking at the associations for names of different lengths (6 letters in light grey - 13 letters in black) separately – as displayed by the regression lines – shows that there is no reliable association of pronounceability with risk when name length is taken into account. The graphs on right show the association of name length with undesirable and desirable risk. Average riskiness ratings for names of given lengths are displayed by crosses. The x-coordinates for name lengths are shifted by -0.5 to 0.5 according to the average pronounceability of the names (also shown by color which ranges from black [1] to white [7]). The confounding association of pronounceability with name length can be seen in all graphs by change of color from left to right. Note that the y-axis is truncated and higher pronounceability ratings correspond to easier pronunciation.

General discussion

The present research originally tested whether the association between processing fluency and judgment of risk differs depending on the category of evaluated objects (studies 1-4). While 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 stimuli as a fixed factor and using different stimuli for each scenario. Consistently, when we gave participants the same stimuli for two scenarios where we expected opposite associations, we found no effect of the scenario (Study 4). We then randomly constructed new stimuli to eliminate any possible bias in the stimuli selection. Using these new stimuli, we found no effect of pronounceability on judgment of harmfulness. On the other hand, when analyzing only the original stimuli used by S&S, we were able to find the original effect even for new scenarios (Study 5). This pattern of results was replicated even when using the original food additives scenario by S&S. While we found no overall effect of pronounceability on judgment of harmfulness, we were able to find the effect of pronounceability on judgment of novelty for both original and newly constructed stimuli (Study 6). In the final study, using another scenario employed by S&S, we found the association between pronounceability and safety for newly sampled stimuli. However, the effect might have been driven completely by word length, which was confounded with pronounceability in S&S. Furthermore, we were not able to replicate the effect with the original stimuli used by S&S (Study 7).

In summary, the results show that the effect of fluency on judgment of riskiness may be much weaker than originally thought or even non-existent. While we found the relationship between pronounceability and safety in the final study, the effect seemed to be completely driven by the association of 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 effect of pronounceability on perceived riskiness when using 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 new 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 safety in the original study. While the study casts doubt on the effect of fluency on perceived risk, we were able to replicate the effect of pronounceability on perceived novelty and we found the association between word length and judgment of risk.

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 et al., 2014; Westfall, Judd, & Kenny, 2015). The result of S&S was 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 S&S and the last study (Cho, 2015) used only four different names and treated them 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) and treat stimuli as a random factor if they want their results to be generalizable.

CHAPTER IV – X good things in life[5]

Introduction

Since the pioneering publication of Seligman and Csikszentmihalyi (2000), the scope of therapeutic interventions has expanded greatly. Nowadays it comprises not only traditional treatments focused on negative symptoms, but also positive interventions aimed at raising well-being and life satisfaction. Some of them have undergone randomized controlled trials that suggest that the interventions are effective in improving life satisfaction and well-being (Bolier et al., 2013; Sin & Lyubomirsky, 2009). Among these exercises is the "Three good things in life" exercise (TGT exercise for short) that requires people to write down each night three things that went well that day. They should also state a cause of each good thing and provide an explanation for why it happened (Seligman, Steen, Park, & Peterson, 2005). Despite evidence suggesting that the TGT exercise can increase happiness and lower depressive symptoms, and the fact that it is routinely presented as a proven positive psychology exercise (Vella-Brodrick, 2013), its working mechanism remains unclear (Mongrain & Anselmo-Matthew, 2012). Based on a different series of studies showing that the number of recalled good memories can influence reported life satisfaction (O'Brien, 2013), we hypothesized that the number of things and the subjective ease with which they can be recalled could play an important role in the TGT exercise. An examination of our hypothesis might not only have significant consequences for optimization of the TGT exercise (e.g. making it

[5] The text of this chapter is reprinted from: Bahník, Š., Vranka, M., & Dlouhá, J. (2015). X good things in life: Processing fluency effects in the "Three good things in life" exercise. *Journal of Research in Personality, 55*, 91-97. http://dx.doi.org/10.1016/j.jrp.2015.02.005

easier and therefore more effective by asking people to recall fewer things) but also shed some light on its working mechanism.

Previous studies

In the study introducing the TGT exercise, Seligman and colleagues randomly assigned the exercise to a group of 59 volunteers who signed up via link advertising “Happiness Exercises” on a website accompanying one of Seligman’s popular books about happiness (Seligman et al., 2005). Participants’ happiness (measured by the Steen Happiness Index, SHI) kept increasing on each measurement after completion of the exercise (i.e. immediately after completion, one week, and one, three, and six months afterwards). The increase from the baseline happiness was statistically significant after one month. Participants’ depressive symptoms (measured by the Center for Epidemiological Studies–Depression Scale, CES-D [Radloff, 1977]) already significantly decreased on the first measurement after the completion and stayed on a lower level in all subsequent measurements. A comparison group of 70 volunteers was given a “placebo” exercise consisting of writing about early memories each day for a week. Although their happiness also increased on the first measurement after finishing the exercise, it returned to the baseline level on all subsequent measurements. Similarly, their depressive symptoms lowered only by a negligible amount.

Mongrain and Anselmo-Matthews (2012) replicated the results of Seligman and colleagues. The only substantial change they made consisted of an addition of a new condition based on the placebo exercise in which they asked the participants to recall *positive* childhood memories each day for a week. Participants in the original placebo condition improved only initially and then returned to the pre-intervention level of well-being, while participants in the TGT group remained happier on all subsequent measurements. However, participants in the newly added “positive childhood memories” group improved to the similar extent as those in the TGT group. This raises the possibility that stating the causes of good things and thinking about why they happened are not important elements of the intervention. Solely thinking about positive things each day may be sufficient as an explanation of the observed effect (Mongrain & Anselmo-Matthews, 2012). In contrast with the original study, Mongrain and Anselmo-Matthews did not observe larger decrease of depressive symptoms (measured by the CES-D) in the treatment condition than in the placebo group.

Other studies also examined the effects of the TGT exercise on well-being. Gander, Proyer, Ruch, and Wyss (2013) followed closely the design of the original study (Seligman et al., 2005) and investigated one- and two-weeks long versions of the TGT exercise. The authors used the same measurements, means of administration, and times of data collection as the original study. However, there was no effect of the exercise on depressive symptoms and happiness increased only in the group with the shorter period of the exercise. Moreover, even this increase was not distinguishable from a change of happiness in the placebo group.

Similarly, in a large recent study Sekizawa and Yoshitake (2015) found only partial support for the effectiveness of a four weeks version of the TGT exercise. Participants in the TGT and placebo group did not differ on the CES-D measure, as well as on measures of life satisfaction, optimism, pessimism, and belief in trustworthiness of others. The sole observed difference between the two groups was in positive affect.

As many authors from this area of research admit (e.g. Burton & King, 2004; Pennebaker & Beall, 1986; Seligman et al., 2005), their studies are focused primarily on testing whether their proposed interventions can increase well-being more than a placebo exercise. Clear understanding of the underlying working mechanism is therefore missing. Our study attempts to examine a possible moderating factor of the success of the exercises, namely the number of good things that participants should recall.

The present study

Several authors suggested that the ease with which people can complete an assigned exercise (or the lack of it) influence induced positive affect and therefore the effect of the exercise. For example participants who are more imaginative (Odou & Vella-Brodrick, 2013), in touch with their emotions (Greenberg et al., 1996), or mindful (Seear & Vella-Brodrick, 2013) are supposed to complete exercises like the TGT more easily and feel more competent while doing them. On the other hand, when an exercise evokes negative affect, presumably because it is too difficult or too boring (Lyubomirsky, Sheldon, & Schkade, 2005), its positive effect can be limited or canceled out completely. Even Seligman and colleagues (2005) mentioned that one week may not be enough time for participants doing the TGT exercise to develop sufficient skill and therefore it could be difficult for at least some of them. Additionally, they suggested that the long-term benefits had been mediated by willingness to continue the exercise voluntarily after the end of the prescribed one-week period. This also points to a possible connection between difficulty of the exercise, willingness to continue doing it, and resulting benefits. Hence, we believe that it is possible that some of the previously found effects can be explained if we take into account the difficulty of the exercise and its subsequent change.

This line of reasoning is supported by a study of O’Brien (2013), who showed that people report lower life satisfaction when they perceive recollection of good memories as hard. According to O’Brien, the metacognitive ease (i.e. fluency) with which one can recall positive or negative memories influences judgment of overall “goodness” or “badness” of a given period. This can be explained in the framework of attribute-substitution theory of heuristics (Kahneman & Frederick, 2002) with fluency of recollection of good memories functioning as a sign of how many good events happened in the past. For example, if it is hard to recall good memories from the last day, people may use such difficulty as a heuristic cue resulting in a judgment that the day was probably not so good. The ease of recollection is usually a valid cue for such judgment. However it may be misleading when there is another source of ease or difficulty. For example, when people are asked to recall many instances of good events from the last day, it may be difficult because of the task at hand and not because of any property of the day to which the experienced difficulty is nevertheless misattributed. This misattribution of difficulty of recollection is well supported by past research (O’Brien, 2013; Schwarz et al., 1991). Consequently, people who are asked to recall only a few good things from the last day can do so more easily than people who are asked to recall many positive things. This may in turn influence perceived quality of the day or result in higher positive affect. A similar reasoning may be applicable in case of the TGT exercise. The lower effectiveness at the beginning of the exercise may be a result of initial difficulties in recalling good things which may decrease with practice. It is also possible that the exercise would be more effective if the number of things to be recalled was initially lower and increased later on or if it was adjusted to particular conditions of each person.

The goal of the present study was to test this possibility. We hypothesized that having people write more good things may result in lower improvement of life satisfaction than having people write less good things. Our hypothesis was, however, not confirmed by the results. We also explored various further questions related to difficulty of the exercise and its effectiveness.

Methods

Pretest

To explore whether it is possible that people find it hard to recall three good things that happened to them during a day, we conducted a pretest in which we asked participants to report as many good (or bad) things that happened to them during a previous day as possible. Then we asked them how many things it was still easy to recall and how satisfied they were with their life on a scale from 1 (not satisfied) to 10 (satisfied).

One hundred and eighty three anonymous participants volunteered to fill in a short questionnaire on a Czech web survey platform. The median age of the participants was 23 years and 87% percent of them were female. We initially considered using both good and bad things in our main study; therefore, we randomly assigned people to one of two groups – reporting either good, or bad things that happened to them during a previous day. Although we did not use the bad things condition in the main study, it enabled us to test whether recalling bad things results in higher reported life satisfaction than recalling good things. This might happen because people were asked to provide as many events as possible, which might have felt difficult. As seen in O'Brien (2013), this should result in higher reported life satisfaction by participants recalling bad events in comparison with participants recalling good events.

One participant was removed from analysis because she provided an invalid answer for one of the questions. In general, we found that it was easier for participants to list good things than bad things. Participants recalling good things wrote more events than participants recalling bad things, $t(180) = 4.58$, $p < .001$, $d = 0.68$, 95% CI [0.38, 0.98], $M_{\text{positive}} = 4.53$ ($SD = 3.34$), $M_{\text{negative}} = 2.60$ ($SD = 2.18$), and reported that it was still easy to write more things, $t(180) = 4.16$, $p < .001$, $d = 0.62$, 95% CI [0.32, 0.91], $M_{\text{positive}} = 2.83$ ($SD = 2.49$), $M_{\text{negative}} = 1.56$ ($SD = 1.48$). Importantly, 52% of participants recalling good things answered that it was easy for them to provide only 2 or fewer events. This suggests that writing three good things as required by the TGT exercise might be difficult for some people.

Participants recalling good things reported somewhat lower life satisfaction than participants recalling bad things, $t(180) = -1.66$, $p = .10$, $d = -0.25$, 95% CI [-0.54, 0.05], $M_{\text{positive}} = 6.40$ ($SD = 2.36$), $M_{\text{negative}} = 6.95$ ($SD = 2.11$). While the difference was not significant, the results were in the direction expected from the results of O'Brien (2013).

Main study

Design, procedure, materials, hypothesis and analysis plan were registered before the beginning of data collection on the Open Science Framework (osf.io/buqh7/). Full wording of materials, analysis scripts, and data are also included therein.

Participants

Two hundred and four students (74% female) of business administration programs at two Czech universities registered for the study and completed a pre-exercise questionnaire. Their average age was 21.8 (*SD* = 2.5). Participants were offered partial course credit for their participation. After the exclusions described in the Analysis section, 172 participants remained. We aimed for a sample as large as possible given our limited resources. A sample of 172 participants results in statistical power $1 - \beta = .80$ for a correlation $r = .21$ given an α level of .05, which means that the study had sufficient power to find a small-to-moderate effect.

Design

Participants were randomly assigned to one of ten groups at the time of their registration. The groups differed only in the number of good things that participants were supposed to report each day (from 1 to 10). The number of things to report stayed constant throughout the two weeks of the exercise. All other aspects of the study were the same for everyone. Life satisfaction measured immediately after the two weeks of the exercise served as the primary dependent variable.

Procedure

The study was introduced in a classroom setting, where participants were informed about the conditions of participation and asked to register at a study website. Participants were able to give their informed consent and complete the entry questionnaire on the study website for 5 days, after which the registration closed. The exercise began the following day. Participants were asked to log in on the web page and fill in the given number of good things (from 1 to 10 as specified by their random assignment to one of the ten respective conditions) that happened to them during the day each day for two weeks. Before submitting their answers, all participants were reminded about the importance of following the instructions closely (nevertheless, they were allowed to report a lower number of good things than stated in the instructions). The web page was accessible each day from 20:00 to 12:00 of the following day. After two weeks, the exercise ended and the participants were sent the first post-exercise questionnaire the following day. The questionnaire contained measures described in Materials section and suggested that the participants continue with the exercise even after the end of the study. Participants were sent two further questionnaires one and six weeks after the end of the exercise. The time schedule along with the numbers of participants completing each part of the study can be found on osf.io/shtkf/.

Materials

As described above, participants filled four questionnaires in total – one before and three after the completion of the exercise. All questionnaires were presented online in Czech on the website specifically designed for the study.

For measuring the main dependent variable of interest, we used an altered version of the Satisfaction with Life Scale (SWLS) (Diener, Emmons, Larsen, & Griffin, 1985). The altered version contained three questions instead of the original five and was focused on current life satisfaction instead of long-term life satisfaction (e.g. "*In most ways my **current** life is close to my ideal.*" – the word "current" was added). Answers were provided on a scale from 1 (strongly disagree) to 7 (strongly agree). The three questions showed good internal consistency, Cronbach's α = .84, 95% CI [.73, .96].

Apart from the SWLS, we used a short form of the Positive and Negative Affect Schedule (PANAS) (Thompson, 2007) for measuring current affective state of participants. The PANAS asks participants to rate to what extent they have felt 10 different emotions – 5 negative and 5 positive (we used only 4 positive emotions in analyses due to an error in translation of an item "alert") on a scale from 1 (very slightly or not at all) to 5 (extremely). Answers for negative and positive emotions are averaged to form two separate measures of negative and positive affect. Both measures showed reasonably good internal consistency, Cronbach's α = .71, 95% CI [.60, .83] for positive affect, and Cronbach's α = .70, 95% CI [.60, .81] for negative affect. The PANAS was intended for exploratory analyses and had only a supportive role for the SWLS.

We administered the SWLS and PANAS in all four questionnaires. In the questionnaire filled immediately after the two weeks of exercise, we also asked participants how difficult they had found recalling good things (on a 7-point scale ranging from *not at all* to *very*), how many minutes per day they had spent on the exercise, whether the exercise had been getting easier with practice, and whether they thought the exercise could have made them more satisfied and happier. The last two question used a 5-point scale from *no* to *yes* with an *I don't know* midpoint. We report the two answers above the midpoint (*rather yes* and *yes*) as positive answers (analogously for the two answers below the midpoint).

In the questionnaires completed one and six weeks after the exercise, in addition to the SWLS and PANAS, we asked participants whether they had continued with the exercises on their own and if so, how often.

Analysis

Following pre-registered exclusion criteria, data from 26 participants who failed to provide at least one good event in their daily exercise questionnaire at least on five occasions were excluded from analysis. Whether a participant provided at least one good event was determined after the end of the exercise in the following manner: All reported events from all participants were pooled and shuffled. Afterwards, an assistant coded whether the events referred to good things (i.e. whether they made sense and could be reasonably considered good things). The information about validity of events was then connected back to the data from participants. Altogether, participants provided 13194 events, out of which 2.9% were not valid (mainly answers such as "nothing" or "I can't remember."). Additionally, we excluded data from 6 participants who failed to fill the SWLS in the entry questionnaire or in the post-study questionnaire. This left 172 participants for the analysis. Data from participants who did not answer the SWLS in the two remaining questionnaires were not excluded because they were (together with the PANAS and other additional questions) intended only for exploratory analyses. Some of these exploratory analyses were therefore conducted with data from a lower number of participants.

Confirmatory pre-registered analysis of the primary hypothesis was conducted with a linear regression with life satisfaction measured immediately after the exercise as a dependent variable. Pre-exercise life satisfaction and the number of good things to be reported each day served as independent variables. We did not have any reason to expect that there is any particular threshold where the recollection of good things leads to more negative life satisfaction. Therefore, we used linear and quadratic contrasts for the number of good things to be reported. A negative parameter for either of the contrasts would be consistent with our hypothesis.

The remaining measures were used for exploratory analyses which were not pre-registered and aimed only to clarify the results.

Results

Confirmatory analysis

Linear regression with post-study life satisfaction as a dependent variable found an effect of pre-study life satisfaction, $t(168) = 14.57$, $p < .001$, $\beta = 0.75$, 95% CI = [0.65, 0.85], but no effect of either linear, $t(168) = 0.44$, $p = .66$, $\beta = 0.02$, 95% CI = [-0.08, 0.13], or quadratic, $t(168) = 0.45$, $p = .66$, $\beta = 0.02$, 95% CI = [-0.08, 0.12], contrasts for the number of things reported each day. This shows that the number of things had no influence on effectiveness of the exercise. The results are virtually unchanged when we use data from all 184 participants who completed both measures without excluding those who failed to complete the exercise five or more times.

Exploratory analyses

Long-term effect

Seligman et al. (2005) found an effect of the exercise only after one month. We therefore analyzed whether there was any effect of the number of good things on life satisfaction one and six weeks after the end of the exercise. However, life satisfaction did not depend on the number of things after one or six weeks. For life satisfaction after one week, the effect of pre-exercise life satisfaction was significant, $t(137) = 12.40$, $p < .001$, $\beta = 0.75$, 95% CI = [0.63, 0.87], but linear , $t(137) = -1.54$, $p = .13$, $\beta = -0.09$, 95% CI = [-0.21, 0.03], or quadratic trends for the number of things, $t(137) = -0.98$, $p = .33$, $\beta = -0.06$, 95% CI = [-0.18, 0.06], were not. Similarly, pre-exercise life satisfaction was associated with life satisfaction after six weeks, $t(115) = 10.08$, $p < .001$, $\beta = 0.69$, 95% CI = [0.55, 0.82], but linear, $t(115) = -0.51$, $p = .61$, $\beta = -0.04$, 95% CI = [-0.17, 0.10], or quadratic trend for the number of things, $t(115) = 0.14$, $p = .89$, $\beta = 0.01$, 95% CI = [-0.13, 0.15], were not. However, it should be noted that life satisfaction correlated highly (.68 < *r*s < .87) between all SWLS measurements, so the results are not independent.

Positive and negative affect

While we did not find an effect of the number of things on life satisfaction, it is possible that it could have influenced positive or negative affect. We tested this hypothesis using the measures of positive and negative affect obtained from the PANAS. We conducted separate regressions for the two measures of affect and three post-exercise questionnaires, which resulted in total six regressions. The measures of positive and negative affect correlated with life satisfaction, .50 < *r*s < .60 for positive affect, and -.55

< *r*s < -.37 for negative affect, and between themselves, -.51 < *r*s < -.29. The effects of linear or quadratic trends for the number of things provided did not approach statistical significance in any analysis with the exception of the linear trend for negative affect measured one week after the end of the exercise, $t(137) = 1.93$, $p = .06$, $\beta = 0.12$, 95% CI = [-0.00, 0.25]. This effect would suggest that reporting more things led to higher negative affect. However, since the result is inconsistent with all other analyses, it is possible it could be a false positive. Detailed results of the analyses can be found on osf.io/m8ti5/.

Following instructions

One possible reason why we did not find any effect of the number of good things could have been that participants did not follow the instructions. That is, that they did not write down the number of things they were asked to. However, participants that were asked to recall more things were only slightly less likely not to do the exercise on any given day. The average number of times participants did not do the exercise ranged from 1.08 for the group writing one thing to 2.25 for the group writing 8 things. Furthermore, participants wrote the full required number of things only slightly less often in groups recalling more things. The average number of days when participants provided the full number of things ranged from 7.57 for the group writing 6 things to 9.31 for the group writing 1 thing. The summary results can be found in Table 1.

Group [N of things]	Number of participants	Difference in pre- and post-exercise life satisfaction	Did not do the exercise [number of days]	Wrote all things [number of days]	Perceived difficulty [scale 1 (not at all) - 7 (very)]	Recollection becoming easier [scale 1 (no) - 5 (yes)]	Belief in exercise causing happiness [scale 1 (no) - 5 (yes)]	Estimated duration the exercise took daily [minutes]
1	13	0.15 (0.94)	1.08 (1.12)	9.31 (0.75)	3.62 (1.66)	3.38 (0.96)	2.62 (1.04)	2.69 (1.65)
2	16	-0.18 (0.70)	1.69 (1.25)	8.56 (1.26)	4.56 (1.50)	3.12 (1.09)	2.88 (1.15)	5.31 (2.89)
3	18	0.22 (0.65)	1.50 (1.38)	9.00 (1.14)	4.28 (1.90)	3.06 (1.16)	2.50 (1.10)	5.22 (3.32)
4	21	0.22 (0.77)	1.81 (1.29)	8.62 (1.02)	4.52 (1.75)	3.19 (0.98)	2.52 (1.21)	5.05 (4.65)
5	13	0.03 (0.98)	1.23 (1.09)	9.08 (1.19)	3.85 (1.41)	3.46 (0.97)	2.62 (0.96)	5.15 (4.12)
6	23	0.10 (1.14)	2.13 (1.22)	7.57 (2.86)	4.74 (1.66)	3.00 (1.28)	2.52 (1.34)	8.09 (5.79)
7	14	-0.21 (0.66)	2.00 (1.41)	8.43 (1.45)	3.50 (1.34)	3.50 (1.16)	3.36 (1.22)	6.71 (4.86)
8	12	0.08 (0.90)	2.25 (1.60)	8.33 (1.61)	4.33 (1.23)	4.17 (0.39)	3.17 (1.03)	7.33 (4.25)
9	19	-0.07 (0.96)	2.05 (1.18)	8.05 (2.53)	4.47 (1.68)	3.74 (1.05)	2.79 (1.03)	7.53 (7.54)
10	23	0.22 (0.74)	1.91 (1.16)	8.26 (2.09)	5.09 (1.28)	3.17 (1.27)	2.78 (1.24)	7.78 (3.94)
r_S	–	-.01 [-.15, .15]	.18 [.04, .33]	-.12 [-.27, .02]	.11 [-.05, .26]	.12 [-.03, .27]	.09 [-.05, .24]	.34 [.21, .46]

Table 2: Summary results of selected variables for the ten experimental groups. The table contains averages and standard deviations for selected measures separately for each of the ten experimental groups. The last row shows Spearman correlation coefficient (along with 95% confidence intervals) between the number of good things recalled each day and a given measure.

Difficulty

Another possibility why we did not find the predicted effect of the number of good things may be that participants did not actually consider the task to be more difficult when they were supposed to write more things. Not surprisingly, participants in groups recalling more things reported spending more time doing the exercise per day, $r_S = .34$, 95% CI = [.21, .46], $p < .001$ (see Table 2), and the perceived difficulty seemed to be dependent on reported time spent working on the exercise, $r_S = .37$, 95% CI = [.23, .50], $p < .001$. However, while people writing more good things considered the exercise more difficult, the association was small and not significant, $r(170) = .13$, 95% CI = [-.02, .27], $p = .09$.

Most of the participants (*n* = 99) reported that the exercise became easier in the course of the study (15 answered that they don't know whether the exercise became easier and 59 answered that it did not). Participants from groups listing more things were somewhat more likely to report that the exercise became easier during the days, r_S = .12, 95% CI = [-.03, .27], *p* = .12, which may partially explain why the difference in perceived difficulty was not larger.

Improvement

The study did not use a control group with a placebo exercise. However, it is still of interest to see whether life satisfaction and affect changed during the exercise. To explore the improvement in life satisfaction and affect, we compared pre-exercise measures with post-exercise measures using paired t-tests. The results are displayed in Figure 9.

We found no difference in life satisfaction and positive affect before and after the exercise. However, a decrease of negative affect can be seen after the exercise. While we cannot confidently attribute the change to the exercise, the results suggest that if the exercise work, it might be primarily by reducing negative affect. This is compatible with the results of Seligman et al. (2005), where the effect of the exercise was especially visible in a decrease of depressive symptoms.

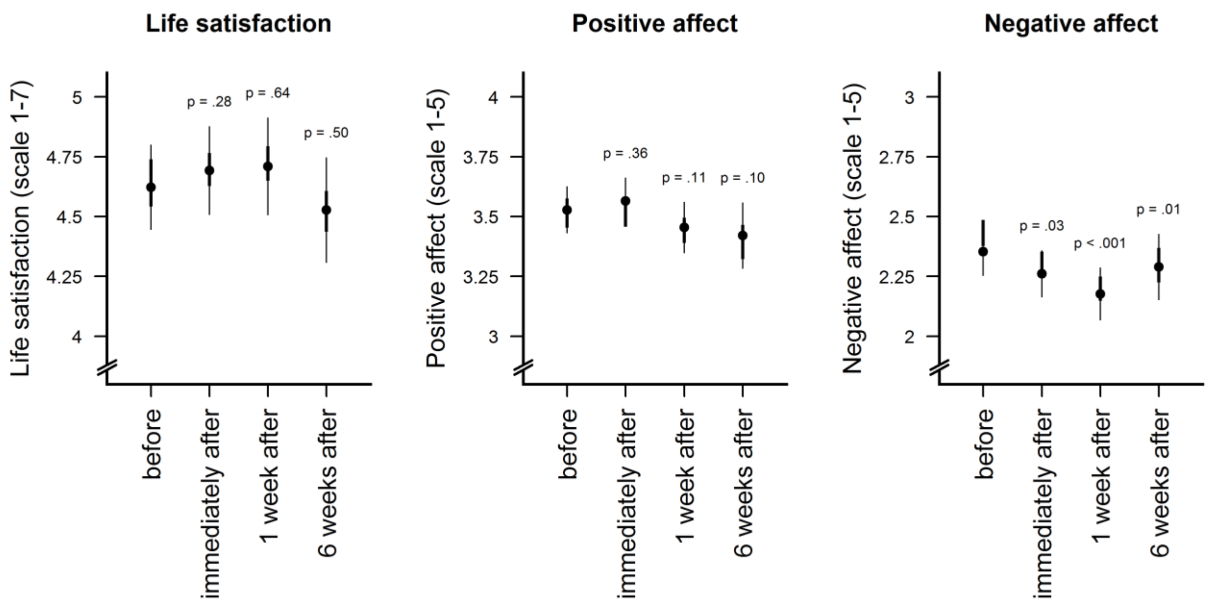

Figure 9. Life satisfaction and affect measured at four points during the study. The figure displays mean values of measures of life satisfaction and affect measured before the exercise and immediately, one week and six weeks after the exercise. Thin error bars represent unadjusted 95% confidence intervals around means. Thick error bars show 95% confidence intervals around means corrected for dependence in measurements (Cousineau-Morey intervals; see Baguley, 2012). These intervals were computed using only data from participants with no missing values for a given measure, so inference from them can differ from a result of comparison between measurements before and after the exercise. The results of such comparisons using paired t-tests are displayed above the error bars. Note that a range of values displayed on ordinate does not contain all possible values obtainable using a given scale.

Continuing the exercise

Seligman et al. (2005) reported that the effect of the exercise was more pronounced for people who continued to do the exercise even after the end of the week they were supposed to do it. We also asked our participants whether they continued to do the exercise in the questionnaires one week and six weeks after the exercise. Thirty two participants (23%) reported continuing the exercise after one week and 20 participants

(17%) after six weeks. This shows that participants in our study did not generally continue with the exercise when they had no external incentive to do so. The participants continuing the exercise did so on average 3.56 times in the one week following the end of the exercise and on average 3.67 times per week in the five weeks afterwards.

Next, we tested whether life satisfaction of participants continuing the exercise improved during the course of the study. We found no indication of improvement in life satisfaction either after one week, $t(31) = -0.96$, $p = .34$, $d = -0.17$, 95% CI [-0.52, 0.18], or after six weeks, $t(19) = 0.08$, $p = 0.94$, $d = 0.02$, 95% CI [-0.42, 0.46].

Belief in the exercise

Unlike in the study of Seligman et al. (2005), our participants had an external motivation to do the exercise. Participants in Seligman et al. were volunteers interested in positive psychology, so it is probable that they were more likely to believe in the effectiveness of the exercise. Indeed, less than one third of our participants ($N = 51$) reported that they thought that the exercise could have caused them to feel more content and happier (34 reported that they did not know whether the exercise could have caused them to feel more content and happier and 87 reported that they did not think so). The belief was not associated with the number of things to be recalled each day, $r_S = .09$, 95% CI = [-.05, .24], $p = .22$, and with reported time spent doing the exercise daily, $r_S = .01$, 95% CI = [-.13, .16], $p = .88$. However, the participants who reported that they believed that the exercise could have caused them to feel more content and happier were less likely to consider the exercise to be difficult, $r_S = -.34$, 95% CI = [-.47, -.20], $p < .001$, and were more likely to do the exercise even after the end of the two weeks, $r_S = .41$, 95% CI = [.25, .57], $p < .001$.

The participants who believed that the exercise could have caused them to feel more content and happier reported higher life satisfaction than before the exercise, immediately after the exercise, $t(50) = 2.79$, $p = .007$, $d = 0.39$, 95% CI [0.10, 0.67], one week after the exercise, $t(36) = 2.53$, $p = .02$, $d = 0.42$, 95% CI [0.08, 0.75], but not six weeks after the exercise, $t(35) = 0.17$, $p = .87$, $d = 0.03$, 95% CI [-0.30, 0.35]. However, it is not clear whether it was the belief that influenced participants' life satisfaction or whether the improvement in life satisfaction affected their belief in the exercise.

Discussion

The main research question of the present study was whether the number of good things to be recalled each day can influence the effectiveness of the TGT exercise. Intuitively, one could believe that recalling a higher number of good things might lead to a higher improvement of life satisfaction. However, following the previous study by O'Brien (2013), we hypothesized that recalling a higher number of good events can be perceived as difficult by some participants. The difficulty (also termed as experienced disfluency) may lead them to infer that their lives do not contain many positive experiences, and thus induce negative affect. This negative affect would then in turn lower or cancel any positive influence of the TGT exercise. If our hypothesis was correct, we would expect to find a negative relationship between the number of things that participants were asked to recall each day (ranging from 1 to 10) and the change in their life satisfaction. However, we found no relationship. Moreover, we did not find such a relationship even with the later post-exercise measurements of life satisfaction or with measures of positive and negative affect. Even though the results constitute a complete lack of support for our hypothesis, there are a few limitations of our study that prevent us from making definitive conclusions.

One possible reason why we did not find any effect might have been that writing more good things was not perceived as much more difficult than writing fewer things. Based on the results of the pretest, it seemed plausible that as much as half of participants would consider recalling even three things difficult. Yet, this estimate concerned only a recall on a single occasion and it is possible that with practice even recalling as much as ten things each day might become relatively easy. Participants in groups writing more good things might have learned to better notice and remember good things or they might have simply listed more mundane things (Folkman & Moskowitz, 2000). The results from exploratory analyses generally support this possibility: There was no strong association between the number of things to be recalled and the rating of difficulty of the exercise, and most people thought that the exercise was getting easier with time – especially participants in the groups recalling more things. These findings fit with the notion that the recall of more things is difficult only initially and is becoming easier afterwards. It is therefore still possible that difficulty of recalling good things could negatively affect change of life satisfaction, but this possibility does not seem to have any practical consequences for the TGT exercise.

We were not primarily interested in the effectiveness of the exercise and for that reason our study did not use a placebo control group. Given the lack of the effect of the number of good things to be recalled, we also evaluated the effectiveness of the exercise itself, although only in an explorative fashion without attempting to come to any strong general conclusions.

We found no difference between life satisfaction and positive affect before and after the exercise at any of the three post-exercise measurements. However, participants reported less negative affect at all measurements after the exercise. Therefore, it seems that the exercise could have worked by reducing negative affect. Still, as we mentioned above, without a placebo control group, any difference or lack of difference before and after the exercise might be due to other factors than the exercise itself such as change of workload due to the progression of an academic term.

Up to this date, there has only been one successful direct replication of the original TGT study by Mongrain and Anselmo-Matthews (2012), who however failed to find any effect of the exercise on depressive symptoms. Another replication study by Gander et al. (2013) brought only mixed results for the original one week version of the exercise and no evidence of effectiveness for the two week version. The authors hypothesized that asking people to do the exercise for two weeks could be counterproductive as it could feel too forced, and thus evoke negative affect. If true, this issue could have had even more serious consequences in the present study as our participants' reward of partial course credit was conditioned on their adherence to the prescribed task, making the possible negative effect more pronounced.

This brings us to probably the most important difference between the present and the original study, which is the recruitment and composition of the study sample. Seligman and colleagues, similarly to the authors of replication studies, used a self-selected sample recruited by advertising of an exercise aimed to rise happiness or build character strengths (Gander et al., 2013; Mongrain & Anselmo-Matthews, 2012; Seligman et al., 2005). Lyubomirsky, Dickerhoof, Bohem, and Sheldon (2011) showed that self-selection is one of the key moderators of the overall success of happiness-increasing interventions, although it is not clear what exactly causes the moderating effect. It is therefore possible that the self-selected sample of participants in Seligman et al. (2005) was more intrinsically motivated to follow the instruction of the exercise and believed more strongly in its effectiveness than participants in our study. This may be important because Seligman and colleagues found that the effect of the exercise was especially pronounced in participants who continued the exercise on their own. Less than one quarter of the participants in our study continued the exercise on their own and participants generally did not state that they had believed that the exercise could have made them more content or happier. Additionally, the participants that believed in the effectiveness of the exercise reported higher life satisfaction after the two weeks of the exercise, which further indicates that the effectiveness of the exercise can be dependent on belief and motivation. This result is in accord with the study of Odou and Vella-Brodrick (2013), that found that motivation and the number of times participants voluntarily completed the TGT exercise during a week were positively correlated with its effects. Given the expectancy effect, selection bias, and high attrition rates, previous studied might have overestimated the effect of the exercise. Our study may have been less likely to suffer from this problem due to the low attrition resulting from the strong external motivation.

It is necessary to note a few additional differences between the present and the original study by Seligman et al. (2005). The original TGT exercise included writing causes and explanations of the good things, an aspect which was not used in our study since we were primarily interested in the effect of the number of good things. It is possible that the effectiveness of the exercise might be due to thinking about reasons for the good things to some extent and not due to recalling and writing them per se. On the other hand, Mongrain and Anselmo-Matthews (2012) found that writing about happy childhood memories without any causal explanation had a similar effect to the TGT exercise, which would suggest that thinking about reasons and their explanations is not essential for effectiveness of the exercise. Second, our measure of life satisfaction differed from the measure used by Seligman and colleagues, who used the Steen Happiness Index (Seligman et al., 2005). However, the SHI has similar psychometric properties to the SWLS and both measure a similar construct (Kaczmarek, Bujacz & Eid, 2015). We used a shortened version of the SWLS in our study, but this did not seem to significantly affect its reliability.

In conclusion, we found that the number of good things recalled does not have any practical impact on effectiveness of the TGT exercise. We did not use a placebo control group, so we cannot confidently infer effectiveness of the exercise from our study. Nevertheless, our results further corroborated previous findings related to the importance of expectations for the effects of the TGT exercise. Further studies that can carefully differentiate a placebo effect from the genuine contribution of recalling good things are needed in order to evaluate the true value of the exercise. And whatever the genuine benefits of recalling good things might be, the number of things, as our study suggests, is not an important factor.

CHAPTER V – Discussion

The study reported in Chapter II showed that participants were faster to make liking decisions for hard-to-pronounce words than for words moderate in pronounceability. The effect suggests that strong feelings emanating from felt fluency or disfluency serve as a strong cue and make the judgment easier and thus faster. While the effect of pronounceability on liking seems to be strong and robust, as shown also in Chapter II, the study reported in Chapter III showed that the effect of pronounceability on judgment of riskiness does not replicate easily. In fact, it may be limited to the stimuli used in the original study demonstrating the effect. While the studies in Chapters II and III were conducted using usual laboratory methods, the study reported in Chapter IV tried to apply a finding from the processing fluency literature to a real-world domain. The results did not show any effect of the manipulation of fluency. This illustrates that any application of processing fluency in a real-world domain requires cautious testing since usual laboratory experiments provide much more controlled and poorer decision environment where fluency may play a role, even though, its effect would be limited in the real-world.

The studies described in the present work explored effects of different types of fluency and on different types of judgments. Chapter II clearly demonstrated a large effect of fluency on liking and Chapter III showed that fluency influences perceived novelty. Both of these effects have been demonstrated a number of times before (e.g., Tversky & Kahneman, 1973; Winkielman et al., 2003) and the present work further illustrates their robustness. On the other hand, Chapter III showed that the previously found effect of fluency on judgment of riskiness (Song & Schwarz, 2009) might not be as robust as previously thought, or may not even exist at all. Even though it was possible to obtain the effect repeatedly with the set of stimuli used in the study demonstrating the effect for the first time, the effect disappeared when newly constructed stimuli were used. It is still possible that the association between fluency and safety may exist. However, as the final experiment in Chapter III shows, it seems that the association is not causal. That is, fluency by itself is not the cause of perceived safety, but both the perception of safety and better pronounceability seem to be a result of another underlying factor. It is, of course, possible that fluency may be used as a cue in judgment of riskiness under certain circumstances. However, Chapter III shows that this cannot be inferred just from the study by Song and Schwarz (2009) and other evidence would be needed before reaching this conclusion.

The studies in chapters II and III both used pronounceability as a manipulation of fluency. Apart from the phonological fluency manipulated by pronounceability, the study in Chapter II also used previous exposure for manipulating fluency. The study in Chapter IV manipulated retrieval fluency by varying the number of things to-be-recalled. In all studies, the results suggested that fluency had some impact on judgment. In Chapter II fluency influenced both judgment of liking and recognition, in Chapter III judgment of novelty, and in the pretest of Chapter IV fluency seemed to influence judgment of life satisfaction. While it has been argued that the different types of processing fluency lead to the same judgmental effects, Chapter II suggests that this may not be always the case. Even though previous exposure influenced recognition judgment as expected, there was no effect of pronounceability on recognition judgment contrary to previous findings. A recent study similarly found that different types of processing fluency may lead to different effect on recognition judgment (Lanska, Olds, & Westerman, 2014). The study suggested that the contribution of the different types of processing fluency to the fluency effect on judgment may depend on their diagnosticity. In the context of the study reported in Chapter II, pronounceability might have been perceived to have little diagnosticity for recognition judgment, which could explain the lack of its effect.

The study reported in Chapter II found a strong effect of pronounceability on liking. While the effect is easy to explain from the perspective of the hedonic marking hypothesis (Winkielman, Schwarz, Fazendeiro, & Reber, 2003), the fluency amplification model (Albrecht & Carbon, 2014) has more troubles explaining the finding. The stimuli used in the study were neutral in valence and yet fluency led to a large effect on liking. One possible explanation is that the theories are not exclusive. It is possible that fluency generally leads to positive affect, but it also amplifies the affect already associated with the stimuli. This can explain both the general shift towards liking for fluent neutral and positive stimuli and the negative effect of fluency on liking for negatively valenced stimuli observed by Albrecht and Carbon (2014). Fluency may thus lead to decrease in liking if the amplification of the negative affect is stronger than the positive affect elicited by fluency itself. In the present study, the stimuli were prepared to not have any connotation. Therefore, they probably did not elicit any negative or positive affect apart from the affect elicited by the associated fluency. Since participants had little else than fluency to base their judgment on, a strong fluency effect was observed.

Even though fluency is often associated with speed of a process, the study in Chapter II did not find the effect of pronounceability on reading speed despite its effect on liking. On the other hand, previous exposure somewhat decreased reading speed. Nevertheless, speed of reading did not seem to mediate the pronounceability effect on liking. It is still possible that pronounceability influenced reading speed, but this did not show in the response time measure used for estimating reading speed. The literature on the association of processing fluency and speed of the process is still limited and future studies may elucidate the lack of an effect of pronounceability on reading speed in the present study.

The initial studies in Chapter II explored the effect of naïve theories on the association between fluency and perceived riskiness. Even though we found that people hold different naïve theories about the association between fluency and riskiness for different categories of objects when we asked them directly, we found little support for the hypothesis that naïve theories play a role in the effect of fluency on actual judgment. The lack of an effect was especially interesting given the number of participants recruited and the resulting statistical power of the experiments. However, possible conclusions from the studies are limited given that the subsequent studies found that the effect of fluency on judgment of riskiness is not reliable to begin with. It is therefore possible that naïve theories would have mediated the effect of fluency on judgment as in other studies (e.g., Galak & Nelson, 2011) if a different target attribute of judgment was studied.

The study reported in Chapter IV tried to improve the effectiveness of a positive psychology exercise by manipulating the ease with which it is possible to do the exercise. However, we did not find any effect of the manipulation on life satisfaction or positive and negative affect. The manipulation influenced the rating of difficulty of the exercise and the time spent on the exercise each day, so it seems that it successfully manipulated the ease of recollection of positive things as it aimed to do. Given that participants listing more good things were more likely to say that the exercise had become easier with time, it is possible that the initial feeling of difficulty decreased in the later days and retrieval fluency was no longer felt once participants adapted to the number of things they were asked to write down, for example, by reporting more menial things. The study demonstrated the issues present when applying laboratory experiments in the real-world settings. While it is possible to manipulate fluency while holding all other aspects constant in the laboratory, as was done in studies in chapters II and III, a lot of other factors than fluency may also play a role in the real-world settings. The effect of fluency is thus bound to be smaller than effects that are normally seen in laboratory studies. Furthermore, it is possible that fluency might not influence only the intended variable, but may affect other factors as well. For example, in the present study the manipulation of the number of things might have influenced not only the difficulty of their retrieval, but also their content. The manipulation seemed to affect the number of days participants did not do the exercise and the belief in the exercise, possibly confounding the intended effect of fluency. Moreover, the dependent variable is likely to be more causally distant from the independent variable in the real-world study. While the laboratory studies reported in the present work explored the effect of fluency on judgment related directly to the stimulus for which fluency was manipulated, the study in Chapter IV manipulated fluency of retrieval for the duration of two weeks in order to influence life satisfaction, which is clearly not influenced by fluency directly and which is determined by many other unrelated factors.

Apart from the implications for processing fluency research, the present studies have also several methodological implications. The topic of replications has become discussed increasingly more in psychology in general (Klein et al., 2014; Open Science Collaboration, 2015) as well as in relation to processing fluency in particular (Meyer et al., 2015; Thompson et al., 2013; Westerman et al., 2015). Chapter III showed an interesting pattern of results in the replication of the study by Song and Schwarz (2009). While the effect seemed to replicate with the original stimuli, it did not replicate with new randomly sampled stimuli (in studies 5 and 6). The discussion in psychology has often focused on direct replications as means of evaluating a previously found effect. Some studies even recruit many laboratories to study the same effect with the original materials that were used in the replicated study (e.g., Alogna et al., 2014). Chapter III, however, shows that such studies may overestimate the replicability of the effect in question because the effect can be limited only to the original materials. If a replication of the study by Song and Schwarz (2009) was conducted using the methods of the Reproducibility project (Open Science Collaboration, 2015) or the Many labs project (Klein et al., 2014), it is likely that the replications would have obtained the original effect. However, as Chapter III shows, this does not mean that the effect is generalizable to other contexts (see also Monin & Oppenheimer, 2014; Westfall et al., 2015), which is clearly of higher importance than just obtaining the original effect. While a majority of commentaries on the replication initiatives focused on the risk of false negative results (e.g., Gilbert, King, Pettigrew, & Wilson, 2016; Schnall, 2014; Schwarz,

& Strack, 2014), Chapter III shows that it is important to consider the risk of false positive results as well.

As described in Chapter III, the results of the study reported therein show the importance of random sampling of stimuli in psychological experiments (see also Judd et al., 2012; Westfall et al., 2014). Unfortunately, the problem of stimuli sampling is ubiquitous in studies of processing fluency. A large part of studies use only a limited number of items and do not treat them as a random factor. Moreover, the same problem applies to manipulations, which often may have just two different levels. For example, when manipulating fluency using two fonts – one of which is better legible and the other harder-to-read – it is not clear whether any observed effect can be indeed attributed to fluency or whether it should be attributed to some other feature that differs between the two fonts. This problem is further exacerbated by the fact that fluency effects may not be linear as Chapter II shows. When the independent variable has only two levels, any nonlinear effect is obscured and any observed effect may not be generalizable to other levels of the independent variable. If the pronounceability range was restricted in the study reported in Chapter II, it would be possible to obtain anything from the effect that fluency leads to faster liking decisions, through no effect of fluency on the speed of liking decision, to disfluency leading to faster liking decisions. It is important to heed the discussed methodological issues in order to achieve robust and generalizable knowledge about processing fluency effects on judgment.

References

Albrecht, S., & Carbon, C. C. (2014). The Fluency Amplification Model: Fluent stimuli show more intense but not evidently more positive evaluations. *Acta Psychologica*, *148*, 195-203.

Alogna, V. K., Attaya, M. K., Aucoin, P., Bahník, Š., Birch, S., Birt, A. R., … Zwaan, R. A. (2014). Registered Replication Report: Schooler and Engstler-Schooler (1990). *Perspectives on Psychological Science, 9*, 556-578.

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.

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.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*, 390-412.

Baayen, R. H., & Milin, P. (2015). Analyzing reaction times. *International Journal of Psychological Research*, *3*, 12-28.

Baguley, T. (2012). Calculating and graphing within-subject confidence intervals for ANOVA. *Behavior Research Methods, 44*, 158–175.

Bahník, Š., Englich, B., & Strack, F. (in press). Anchoring effect. In R. F. Pohl (Ed.). *Cognitive Illusions: Intriguing Phenomena in Thinking, Judgment, and Memory (2nd ed.)*. Hove, UK: Psychology Press.

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.

Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015). Parsimonious mixed models. *ArXiv e-print; 1506.04967*.

Bates, D., Maechler, M., Bolker, B.M., & Walker, S. (2014). lme4: Linear mixed-effects models using Eigen and S4. *ArXiv e-print; 1406.5823*.

Beilock, S. L., & Holt, L. E. (2007). Embodied Preference Judgments: Can Likeability Be Driven by the Motor System? *Psychological Science, 18*, 51-57.

Bolier, L., Haverman, M., Westerhof, G. J., Riper, H., Smit, F. & Bohlmeijer, E. (2013). Positive psychology interventions: A meta-analysis of randomized controlled studies. *BMC Public Health*, *13*, 119–139.

Briñol, P., Petty, R. E., & Tormala, Z. L. (2006). The Malleable Meaning of Subjective Ease. *Psychological Science, 17*, 200-206.

Burton, C. M. & King, L. A. (2004). The health benefits of writing about intensely positive experiences. *Journal of Research in Personality*, *38*, 150–163.

Casasanto, D. (2014). Bodily relativity. In L. Shapiro (Ed.), *Routledge handbook of embodied cognition* (pp. 108–117). London, UK: Routledge.

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

de la Vega, I., De Filippis, M., Lachmair, M., Dudschig, C., & Kaup, B. (2012). Emotional valence and physical space: Limits of interaction. *Journal of Experimental Psychology: Human Perception and Performance*, *38*, 375-385.

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.

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.

Diener, E. D., Emmons, R. A., Larsen, R. J., & Griffin, S. (1985). The satisfaction with life scale. *Journal of Personality Assessment*, *49*, 71–75.

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.

Folkman, S. & Moskowitz, J. T. (2000). Positive affect and the other side of coping. *American Psychologist*, *55*, 647–654.

Forster, M., Gerger, G., & Leder, H. (2015). Everything's relative? Relative differences in processing fluency and the effects on liking. PloS one, 10(8), e0135944.

Forster, M., Leder, H., & Ansorge, U. (2013). It felt fluent, and I liked it: Subjective feeling of fluency rather than objective fluency determines liking. *Emotion, 13*, 280-289.

Fox, C. R. (2006). The availability heuristic in the classroom: How soliciting more criticism can boost your course ratings. *Judgment and Decision Making, 1*, 86-90.

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.

Gander, F., Proyer, R. T., Ruch, W. & Wyss, T. (2013). Strength-based positive interventions: Further evidence for their potential in enhancing well-being and alleviating depression. *Journal of Happiness Studies, 14*, 1241–1259.

Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. New York, NY: Cambridge University Press.

Gigerenzer, G. (2008). Why Heuristics Work. *Perspectives on Psychological Science, 3*, 20-29.

Gigerenzer, G., Czerlinski, J., & Martignon, L. (2002). How Good Are Fast and Frugal Heuristics? In T. Gilovich, D. Griffin & D. Kahneman (Eds.). *Heuristics and Biases: The Psychology of Intuitive Judgment* (pp. 559-581). New York: Cambridge University Press.

Gilbert, D., King, G., Pettigrew, S., & Wilson, T. (2016). Comment on 'Estimating the reproducibility of psychological science'. *Science, 351*, 1037.

Gilovich, T., & Griffin, D. (2002). Introduction – Heuristics and Biases: Then and Now. In T. Gilovich, D. Griffin & D. Kahneman (Eds.). *Heuristics and Biases: The Psychology of Intuitive Judgment* (pp. 1-18). New York: Cambridge University Press.

Greenberg, M. A., Wortman, C. B. & Stone, A. A. (1996). Emotional expression and physical heath: Revising traumatic memories or fostering self-regulation? *Journal of Personality and Social Psychology*, *71*, 588–602.

Greifeneder, R., Alt, A., Bottenberg, K., Seele, T., Zelt, S., & Wagener, D. (2010). On Writing Legibly: Processing Fluency Systematically Biases Evaluations of Handwritten Material. *Social Psychological and Personality Science, 1*, 230-237.

Hansen, J., Dechêne, A., & Wänke, M. (2008). Discrepant Fluency Increases Subjective Truth. *Journal of Experimental Social Psychology, 44*, 687-691.

Hertwig, R., Herzog, S. M., Schooler, L. J., & Reimer, T. (2008). Fluency heuristic: A model of how the mind exploits a by-product of information retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*, 1191-1206.

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.

Kaczmarek, L. D., Bujacz, A., & Eid, M. (2015). Comparative Latent State-Trait Analysis of Satisfaction with Life Measures: The Steen Happiness Index and the Satisfaction with Life Scale. *Journal of Happiness Studies, 16*, 443-453.

Kahneman, D. (1991). Judgment and Decision Making: A Personal View. *Psychological Science, 2*, 142-145.

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

Klein, R. A., Ratliff, K. A., Vianello, M., Adams, R. B., Jr., Bahník, Š., Bernstein, M. J., … Nosek, B. A. (2014). Investigating Variation in Replicability: A “Many Labs” Replication Project. *Social Psychology, 45*, 142-152.

Labroo, A. A., Lambotte, S., & Zhang, Y. (2009). The „Name-Ease“ Effect and Its Dual Impact on Importance Judgments. *Psychological Science, 20*, 1516-1522.

Laham, S. M., Koval, P., & Alter, A. L. (2012). The name-pronunciation effect: Why people like Mr. Smith more than Mr. Colquhoun. *Journal of Experimental Social Psychology*, *48*, 752-756.

Lanska, M., Olds, J. M., & Westerman, D. L. (2014). Fluency effects in recognition memory: Are perceptual fluency and conceptual fluency interchangeable?. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40*, 1-11.

Lyubomirsky, S., Dickerhoof, R., Boehm, J. K., & Sheldon, K. M. (2011). Becoming happier takes both a will and a proper way: An experimental longitudinal intervention to boost well-being. *Emotion, 11*, 391-402.

Lyubomirsky, S., Sheldon, K. M. & Schkade, D. (2005). Pursuing Happiness: The Architecture of Sustainable Change. *Review of General Psychology*, *9*, 111–131.

McGlone, M. S., & Tofighbakhsh, J. (1999). The Keats Heuristic: Rhyme as Reason in Aphorism Interpretation. *Poetics, 26*, 235-244.

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.

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.

Mongrain, M., & Anselmo-Matthews, T. (2012). Do Positive Psychology Exercises Work? A Replication of Seligman et al.(2005). *Journal of Clinical Psychology, 68*, 382– 389.

Monin, B., & Oppenheimer, D. M. (2014) The Limits of Direct Replications and the Virtues of Stimulus Sampling. *Social Psychology, 45*, 299-300.

O'Brien, E. (2013). Easy to Retrieve but Hard to Believe: Metacognitive Discounting of the Unpleasantly Possible. *Psychological Science, 24*, 844–851.

Odou, N., & Vella-Brodrick, D. A. (2013). The efficacy of positive psychology interventions to increase well-being and the role of mental imagery ability. *Social Indicators Research*, *110*, 111–129.

Olds, J. M., & Westerman, D. L. (2012). Can fluency be interpreted as novelty? Retraining the interpretation of fluency in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38*, 653-664.

Open Science Collaboration (2015). Estimating the Reproducibility of Psychological Science. *Science, 349*, 943.

Oppenheimer, D. M. (2004). Spontaneous Discounting of Availability in Frequency Judgment Tasks. *Psychological Science, 15*, 100-105.

Oppenheimer, D. M., & Monin, B. (2009). Investigations in Spontaneous Discounting. *Memory & Cognition, 37*, 608-614.

Pennebaker, J. W. & Beall, S. K. (1986). Confronting a traumatic event: Toward an understanding of inhibition and disease. *Journal of Abnormal Psychology*, *95*, 274–281.

Pocheptsova, A., Labroo, A. A., & Dhar, R. (2010). Making Products Feel Special: When Metacognitive Difficulty Enhances Evaluation. *Journal of Marketing Research, 47*, 1059-1069.

Radloff, L. S. (1977). The CES-D scale a self-report depression scale for research in the general population. *Applied Psychological Measurement*, *1*, 385–401.

Ratcliff, R., Smith, P. L., Brown, S. D., & McKoon, G. (2016). Diffusion Decision Model: Current Issues and History. *Trends in Cognitive Sciences*, *20*, 260-281.

Reber, R., & Unkelbach, C. (2010). The epistemic status of processing fluency as source for judgments of truth. *Review of Philosophy and Psychology, 1*, 563-581.

Reber, R., Winkielman, P., & Schwarz, N. (1998). Effects of Perceptual Fluency on Affective Judgments. *Psychological Science, 9*, 45-48.

Reber, R., Wurtz, P., & Zimmermann, T. D. (2004). Exploring “fringe” consciousness: The subjective experience of perceptual fluency and its objective bases. *Consciousness and Cognition*, *13*, 47-60.

Ross, M., & Sicoly, F. (1979). Egocentric Biases in Availability and Attribution. *Journal of Personality and Social Psychology, 37,* 322-336.

Savitsky, K., Van Boven, L., Epley, N., & Wight, W. M. (2005). The unpacking effect in allocations of responsibility for group tasks. *Journal of Experimental Social Psychology, 41*, 447-457.

Schnall, S. (2014). Clean Data: Statistical Artifacts Wash Out Replication Efforts. *Social Psychology, 45*, 315-317.

Schwarz, N., Bless, H., Strack, F., Klumpp, G., Rittenauer-Schatka, H., & Simons, A. (1991). Ease of Retrieval as Information: Another Look at the Availability Heuristic. *Journal of Personality and Social Psychology, 61*, 195–202.

Schwarz, N., & Strack, F. (2014). Does Merely Going Through the Same Moves Make for a ‘‘Direct’’ Replication? Concepts, Contexts, and Operationalizations. *Social Psychology, 45*, 305-306.

Seear, K. H. & Vella-Brodrick, D. A. (2013). Efficacy of positive psychology interventions to increase well-being: Examining the role of dispositional mindfulness. *Social Indicators Research*, *114*, 1125–1141.

Sekizawa, Y., & Yoshitake, N. (2015). *Does the Three Good Things Exercise Really Make People More Positive and Less Depressed? A study in Japan*. (RIETI Discussion Paper Series 15-E-001)

Seligman, M. E. & Csikszentmihalyi, M. (2000). Positive psychology: An introduction. *American Psychologist*, *55*, 5–14.

Seligman, M. E., Steen, T. A., Park, N., & Peterson, C. (2005). Positive psychology progress. *American Psychologist, 60*, 410-421.

Shah, A. K., & Oppenheimer, D. M. (2007). Easy Does It: The Role of Fluency in Cue Weighting. *Judgment and Decision Making, 2*, 371-379.

Shah, A. K., & Oppenheimer, D. M. (2008). Heuristics Made Easy: An Effort-Reduction Framework. *Psychological Bulletin, 134*, 202-222.

Shah, A. K., & Oppenheimer, D. M. (2009). The Path of Least Resistance: Using Easy-to-Access Information. *Current Directions in Psychological Science, 18*, 232-236.

Simmons, J. P., & Nelson, L. D. (2006). Intuitive Confidence: Choosing Between Intuitive and Nonintuitive Alternatives. *Journal of Experimental Psychology: General, 135*, 409-428.

Sin, N. L. & Lyubomirsky, S. (2009). Enhancing well-being and alleviating depressive symptoms with positive psychology interventions: A practice-friendly meta-analysis. *Journal of Clinical Psychology*, *65*, 467–487.

Song, H., & Schwarz, N. (2008). Fluency and the detection of misleading questions: Low processing fluency attenuates the Moses illusion. *Social Cognition*, *26*, 791-799.

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.

Strack, F., Bahník, Š., & Mussweiler, T. (in press). Anchoring: Accessibility as a Cause of Judgmental Assimilation. *Current Opinion in Psychology*.

Thomas, M., & Morwitz, V. G. (2009). The Ease-of-Computation Effect: The Interplacy of Metacognitive Experiences and Naive Theories in Judgments of Price Differences. *Journal of Marketing Research, 46*, 81-91.

Thompson, E. R. (2007). Development and validation of an internationally reliable short-form of the positive and negative affect schedule (PANAS). *Journal of Cross-Cultural Psychology*, *38*, 227–242.

Thompson, S. C., & Kelley, H. H. (1981). Judgments of Responsibility for Activities in Close Relationships. *Journal of Personality and Social Psychology, 41*, 469-477.

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.

Topolinski, S., & Strack, F. (2009a). Motormouth: Mere Exposure Depends on Stimulus-Specific Motor Simulations. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*, 423-433.

Topolinski, S., & Strack, F. (2009b). The architecture of intuition: Fluency and affect determine intuitive judgments of semantic and visual coherence and judgments of grammaticality in artificial grammar learning. *Journal of Experimental Psychology: General*, 138, 39-63.

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.

Tversky, A., & Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases. *Science, 4157*, 1124-1131.

Tversky, A., & Kahneman, D. (1982). Judgment of and by Representativeness. In D. Kahneman, P. Slovic & A. Tversky (Eds.). *Judgment under Uncertainty: Heuristics and Biases* (pp. 84-98). New York: Cambridge University Press.

Tversky, A., & Kahneman, D. (1983). Extensional versus Intuitive Reasoning: The Conjunction Fallacy in Probability Judgment. *Psychological Review, 91*, 239-315.

Unkelbach, C. (2006). The Learned Interpretation of Cognitive Fluency. *Psychological Science, 17*, 339-345.

Unkelbach, C. (2007). Reversing the Truth Effect: Learning the Interpretation of Fluency in Judgments of Truth. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*, 219-230.

Unkelbach, C., & Greifeneder, R. (2013). A general model of fluency effects in judgment and decision making. In C. Unkelbach & R. Greifeneder (Eds.), *The Experience of Thinking* (pp. 11-32). New York: Psychology Press.

Vella-Brodrick, D. A. (2013). Positive Psychology Interventions: Research Evidence, Practical Utility, and Future Steps. In C. L. M. Keyes (Ed.) *Mental Well-Being: International Contributions to the Study of Positive Mental Health.* (pp. 331–353). Dordecht, Netherlands: Springer.

Voss, A., Nagler, M., & Lerche, V. (2013). Diffusion models in experimental psychology. *Experimental Psychology, 60*, 385-402.

Westerman, D. L., Klin, C. M., & Lanska, M. (2015). On the (elusive) role of oral motor-movements in fluency-based memory illusions. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 41*, 1003-1013.

Westerman, D. L., Lanska, M., & Olds, J. M. (2015). The effect of processing fluency on impressions of familiarity and liking. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 41*, 426-438.

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.

Winkielman, P., & Cacioppo, J. T. (2001). Mind at ease puts a smile on the face: Psychophysiological evidence that processing facilitation elicits positive affect. *Journal of Personality and Social Psychology, 81*, 989-1000.

Winkielman, P., & Schwarz, N. (2001). How Pleasant Was Your Childhood? Beliefs About Memory Shape Inferences From Experienced Difficulty of Recall. *Psychological Science, 12*, 176-179.

Winkielman, P., Schwarz, N., & Belli, R. F. (1998). The Role of Ease of Retrieval and Attribution in Memory Judgments: Judging Your Memory as Worse Despite Recalling More Events. *Psychological Science, 9*, 124-126.

Winkielman, P., Schwarz, N., Fazendeiro, T., & Reber, R. (2003). The hedonic marking of processing fluency: Implications for evaluative judgment. In J. Musch & K. C. Klauer (Eds.), *The psychology of evaluation: Affective processes in cognition and emotion* (pp. 189–217). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Wurtz, P., Reber, R., & Zimmermann, T. D. (2008). The feeling of fluent perception: A single experience from multiple asynchronous sources. *Consciousness and Cognition, 17*, 171-184.

Zajonc, R. B. (1968). Attitudinal effects of mere exposure. *Journal of Personality and Social Psychology*, *9*, 1-27.

Appendix A: Affidavit

I hereby confirm that my thesis entitled “Processing fluency and judgment” is the result of my own work. I did not receive any help or support from commercial consultants. All sources and / or materials applied are listed and specified in the thesis.

Furthermore, I confirm that this thesis has not yet been submitted as part of another examination process neither in identical nor in similar form.

Würzburg, 23th May 2016 Štěpán Bahník

Eidesstattliche Erklärung

Hiermit erkläre ich an Eides statt, die Dissertation “Verarbeitungsflüssigkeit und Urteilsbildung” eigenständig, d.h. insbesondere selbständig und ohne Hilfe eines kommerziellen Promotionsberaters, angefertigt und keine anderen als die von mir angegebenen Quellen und Hilfsmittel verwendet zu haben.

Ich erkläre außerdem, dass die Dissertation weder in gleicher noch in ähnlicher Form bereits in einem anderen Prüfungsverfahren vorgelegen hat.

Würzburg, 23. Mai 2016 Štěpán Bahník

Appendix B: Curriculum Vitae

Štěpán Bahník

E-mail: bahniks@seznam.cz
Web: bahniks.com

EDUCATION

M.Sc. Biology, Charles University in Prague (2013)
M.A. Psychology, Charles University in Prague (2011)
B.Sc. Biology, Charles University in Prague (2011)
M.Sc. Corporate Economics and Management, University of Economics, Prague (2010)
B.Sc. Corporate Economics and Management, University of Economics, Prague (2009)

RESEARCH EXPERIENCE

10/2013 – present	University of Würzburg, Department of Psychology II, Social Psychology
11/2010 – 9/2013	Institute of Physiology, Academy of Sciences of the Czech Republic, Department of Neurophysiology of Memory

PUBLICATIONS

in press

Bahník, Š., Englich, B., & Strack, F. (in press). Anchoring effect. In R. F. Pohl (Ed.). *Cognitive Illusions: Intriguing Phenomena in Thinking, Judgment, and Memory (2nd ed.)*. Hove, UK: Psychology Press.

Bahník, Š., & Rubínová, E. (in press). Psychological methodology in practice of a clinical neuropsychologist. In P. Kulišťák (Ed.). *Clinical neuropsychology in practice*. [in Czech]

Strack, F., **Bahník, Š.**, & Mussweiler, T. (in press). Anchoring: Accessibility as a Cause of Judgmental Assimilation. *Current Opinion in Psychology*.

Vranka, M., & **Bahník, Š.** (in press). Differences in autonomy of humans and ultrasocial insects. *Behavioral and Brain Sciences*.

Vranka M., & **Bahník, Š.** (in press). Is the emotional dog blind to its choices? An attempt to reconcile the Social intuitionist model and the choice blindness effect. *Experimental Psychology*.

2016

Anderson, C. J., **Bahník, Š.**, Barnett-Cowan, M., Bosco, F. A., Chandler, J., Chartier, C. R., … Zuni, K. (in press). Response to a comment on “Estimating the reproducibility of psychological science”. *Science, 351*, 1037.

Bahník, Š., & Strack, F. (2016). Overlap of accessible information undermines the anchoring effect. *Judgment and Decision Making, 11*, 92-98.

2015

Bahník, Š., & Stuchlík, A. (2015). Temporal and spatial strategies in an active place avoidance task on Carousel: A study of effects of stability of arena rotation speed in rats. *PeerJ,* 3:e1257.

Bahník, Š., Vranka, M., & Dlouhá, J. (2015). X good things in life: Processing fluency effects in the "Three good things in life" exercise. *Journal of Research in Personality, 55*, 91-97.

Open Science Collaboration (2015). Estimating the Reproducibility of Psychological Science. *Science, 349*, 943.

2014

Alogna, V. K., Attaya, M. K., Aucoin, P., **Bahník, Š.**, Birch, S., Birt, A. R., … Zwaan, R. A. (2014). Registered Replication Report: Schooler and Engstler-Schooler (1990). *Perspectives on Psychological Science, 9*, 556-578.

Klein, R. A., Ratliff, K. A., Vianello, M., Adams, R. B., Jr., **Bahník, Š.**, Bernstein, M. J., … Nosek, B. A. (2014). Data from Investigating Variation in Replicability: A "Many Labs" Replication Project. *Journal of Open Psychology Data, 2(1): e4,* doi: http://dx.doi.org/10.5334/jopd.ad

Klein, R. A., Ratliff, K. A., Vianello, M., Adams, R. B., Jr., **Bahník, Š.**, Bernstein, M. J., … Nosek, B. A. (2014). Investigating Variation in Replicability: A "Many Labs" Replication Project. *Social Psychology, 45*, 142-152.

Klein, R. A., Ratliff, K. A., Vianello, M., Adams, R. B., Jr., **Bahník, Š.**, Bernstein, M. J., … Nosek, B. A. (2014). Theory Building Through Replication: Response to Commentaries on the "Many Labs" Replication Project. *Social Psychology, 45*, 307-310.

Open Science Collaboration. (2014). The Reproducibility Project: A model of large-scale collaboration for empirical research on reproducibility. In V. Stodden, F. Leisch, & R. D. Peng (Eds.), *Implementing Reproducible Computational Research* (A Volume in The R Series), NY, NY: Taylor & Francis.

Petrasek, T., Prokopova, I., **Bahnik, S.**, Schonig, K., Berger, S., Vales, K., … Stuchlik, A. (2014). Nogo-A downregulation impairs place avoidance in the Carousel maze but not spatial memory in the Morris water maze. *Neurobiology of Learning and Memory, 107*, 42-49.

Petrasek, T., Prokopova, I., Sladek, M., Weissova, K., Vojtechova, I., **Bahnik, S.**, … Stuchlik, A. (2014). Nogo-A-deficient transgenic rats show deficits in higher cognitive functions, decreased anxiety and altered circadian activity patterns. *Frontiers in Behavioral Neuroscience, 8:90*. doi: 10.3389/fnbeh.2014.00090

2012

Prokopova, I., **Bahnik, S.**, Doulames, V., Vales, K., Petrasek, T., Svoboda, J., & Stuchlik, A. (2012). Synergistic effects of dopamine D2-like receptor antagonist sulpiride and beta-blocker propranolol on learning in the Carousel maze, a dry-land spatial navigation task. *Pharmacology Biochemistry and Behavior, 102*, 151-156.

Bahník, Š. (2012). Anchoring in the courtroom. *Psychologie pro praxi, 47*, 93-102. [in Czech]

CONFERENCE PRESENTATIONS

Bahník, Š., & Vranka, M. (2015, November). *If it's difficult to pronounce, it might not be risky: Fluency, risk perception, and random sampling of stimuli.* Poster session presented at the meeting of the Society of Judgment and Decision Making, Chicago, IL.

Bahník, Š., & Strack, F. (2015, August). *Overlap of Accessible Information Undermines the Anchoring Effect.* Paper presented at the Subjective Probability, Utility, and Decision Making Conference, Budapest, Hungary.

Vranka, M., & Bahník, Š. (2015, May). *The difference in prescriptive and normative moral judgment: Moral asymmetry in self- and other- sacrifice judgments in trolley-like dilemmas.* Poster session presented at the meeting of the Association for Psychological Science, New York, NY.

Bahník, Š., & Strack, F. (2014, November). *Overlap of accessible information undermines the anchoring effect.* Poster session presented at the meeting of the Society of Judgment and Decision Making, Long Beach, CA.

Bahník, Š., & Vranka, M. (2014, July). *If it's easy to pronounce, it might be risky: The effect of fluency on judgment of risk depends on a category of risk.* Poster session presented at the meeting of the European Association of Social Psychology, Amsterdam, Netherlands.

TEACHING

Introduction to empirical and experimental research methods (winter 2015, summer 2016)
Moral judgment (summer 2015, summer 2016)

SOFTWARE

Carousel Maze Manager	Python program for analyzing data from widely used behavioral tasks bahniks.com/cmm/