ARTICLE IN PRESS

Available online at www.sciencedirect.com
ScienceDirect

Cognition xxx (2007) xxx–xxx

COGNITION

www.elsevier.com/locate/COGNIT

A rose in any other font would not smell as sweet: Effects of perceptual fluency on categorization ☆

Daniel M. Oppenheimer [a,*], Michael C. Frank [b]

[a] Department of Psychology, Princeton University, Princeton, NJ 08544-1010, USA
[b] Massachusetts Institute of Technology 77 Massachusetts Avenue Cambridge, MA 02139, USA

Received 27 May 2005; revised 19 May 2007; accepted 25 May 2007

Abstract

Fluency – the ease with which people process information – is a central piece of information we take into account when we make judgments about the world. Prior research has shown that fluency affects judgments in a wide variety of domains, including frequency, familiarity, and confidence. In this paper, we present evidence that fluency also plays a role in categorization judgments. In Experiment 1, participants judged a variety of different exemplars to be worse category members if they were less fluent (because they were presented in a smaller typeface). In Experiment 2, we found that fluency also affected judgments of feature typicality. In Experiment 3, we demonstrated that the effects of fluency can be reversed when a salient attribution for reduced fluency is available (i.e., the stimuli are hard to read because they were printed by a printer with low toner). In Experiment 4 we replicated these effects using a within-subject design, which ruled out the possibility that the effects were a statistical artifact caused by aggregation of data. We propose a possible mechanism for these effects: if an exemplar and its category are closely related, activation of one will cause priming of the other, leading to

☆ This material is based upon work supported by the Norman Anderson research fund and by Grant # 0518811 from the National Science Foundation. The authors would like to thank two anonymous reviewers, Sara Etchison, David Collister, Anuj Shah, Steve Sloman, Melissa Miller, Jessica Laughlin, Benoit Monin, Rebecca Norwick, Max Abelev, Teenie Matlock, and Pamela Sawyer for their help in all aspects of this research, and thanks to mom and dad on principle.

* Corresponding author. Tel.: +1 609 258 7465.
E-mail address: doppenhe@princeton.edu (D.M. Oppenheimer).
URL: www.princeton.edu/sites/opplab (D.M. Oppenheimer).

doi:10.1016/j.cognition.2007.05.010

increased fluency. Over time, feelings of fluency come to be used as a valid cue that can become confused with more traditional sources of information about category membership.

Keywords: Perceptual fluency; Categorization; Priming; Categorical induction

1. Effects of perceptual fluency on categorization

Imagine the subjective experience of reading this article in a poor photocopy: words are blurry, in places illegible and obscured at the sides of the page by the rise of the journal's binding. In contrast, imagine the article printed crisply off a new laser printer: the text is clear, centered on the page, high in contrast, and easy to read. Naively we might imagine that these superficial factors involved in presentation will not affect readers' perception of our work. However, a growing psychological literature suggests that the ease with which we are able to access information, the meta-cognitive experience of ease of processing that we call fluency, can affect a wide variety of our judgments. Effects of fluency have been shown in domains as disparate as intelligence (Oppenheimer, 2005), frequency (Tversky & Kahneman, 1973), familiarity (Monin, 2003), duration (Witherspoon & Allan, 1985), risk for disease (Rothman & Schwarz, 1998), and attractiveness (Reber, Winkielman, & Schwarz, 1998). In fact, the fluency caused by the presentation of this article might influence a wide variety of judgments about its contents.

Fluency may also affect categorization. The most typical exemplars of a category are often the most frequent in our experience, the most easily accessible in our memories, and the most primed by related experiences. Because of this, the experience of fluency and the judgment of good category membership frequently co-occur over the course of a lifetime's experience. These correlations make fluency a highly reliable cue to category membership, one that our experiments suggest can influence decision-makers. More generally, our experiments suggest that individuals do not fully distinguish between the cues for category membership cited by traditional theories of categorization – such as similarity to a prototype (e.g., Rosch & Mervis, 1975) or to other members of that category (e.g., Medin & Schaffer, 1978) – and experiential cues such as fluency. In this article, we present evidence that changing the fluency of a category member can alter participants' judgments of its typicality (Experiments 1 and 4); that changing the fluency of category features can change judgments of the features' typicality (Experiment 2); and that, given a salient alternative explanation, participants can discount the effects of fluency (Experiment 3). In our discussion, we present several suggestions for how such experiential cues could be integrated into a variety of existing models of categorization.

Perceptual fluency is generally defined as the subjective experience of ease or speed in processing perceptual information (Jacoby & Dallas, 1981; Benjamin, Bjork, &

Hirshman, 1998; Reber, Wurtz, & Zimmerman, 2004).[1] Experimenters have manipulated fluency in a variety of ways, most often by changing font size of text (e.g., Norwick & Epley, 2003; Epley & Norwick, submitted for publication), adding visual noise for images and text (e.g., Kleider & Goldinger, 2004; Buchner, 1994), or manipulating phonological properties for nonce words (e.g., Whittlesea & Leboe, 2002).

Fluency can be thought of as a metacognitive experience, in which one has a subjective feeling that a cognitive process is running smoothly. When people have such an experience, they often attempt to interpret it, and those interpretations can influence their judgments (for a review see Schwarz, 2004). This is especially true when the fluency experience is surprising (Whittlesea & Williams, 2000). People often have difficulty determining the source of a given fluency experience and, as such, are biased in their judgments (Jacoby & Dallas, 1981). For example, Jacoby and Dallas showed that people would misattribute fluency due to recent exposure to fluency caused by fame, and that fame judgments were biased accordingly.

As the widespread impact of fluency on a variety of cognitive phenomena grows, researchers have begun to accumulate evidence that fluency may also affect categorization judgments. However, most common accounts of categorization do not currently account for the effects of metacognitive attributes such as fluency.

The issue of how we represent the world and break down continuous perceptual stimuli into discrete categories has been historically central to the field of cognitive science (Goldstone & Kersten, 2002). In the modern literature on categorization, most theories can be assigned to one of three groups: prototype models; exemplar models; and theory- or knowledge-based models. In prototype theories, the central tendency – or prototype – of a category is extrapolated by finding the common features of the known exemplars of the category. Further exemplars are judged to be category members if they are closer to the prototype of that category than to the prototype of any other (e.g., Rosch & Mervis, 1975). In contrast, exemplar theories do not explicitly represent a central tendency – instead judging category membership by comparing novel exemplars to all previous examples in that category (e.g., Medin & Schaffer, 1978). Finally, theory- and knowledge-based models of categorization take into account the semantic content of categories and posit that the process of categorization is governed by background knowledge about the domain of categorization (e.g., Keil, 1989; Murphy & Medin, 1985). For example, exemplars may be judged to be members of a category to the extent that they can be generated by the category's causal model (e.g., Rehder & Hastie, 2001).

[1] In addition to perceptual fluency, there are other cognitive operations that can have an influence on fluency. Retrieval fluency, also known as availability, reflects how easily instances of an item are brought to mind (Tversky & Kahneman, 1973). Conceptual fluency is defined by the contextual situation and semantic predictability of a stimulus (Whittlesea, 1993). While these variations in fluency are subtly distinct, all of them have been shown to have similar effects on processing (Alter & Oppenheimer, submitted for publication; Winkielman, Schwarz, Fazendeiro, & Reber, 2003). As such, for the remainder of this paper, we shall refer to them as a single construct.

Taken as a whole, theories of categorization remain largely silent on issues of presentation. Presumably, under a standard exemplar or prototype model of categorization, categorization of a novel exemplar should not depend on properties of the presentation format of the stimulus – such as the font size in which the exemplar's name was written – that are not relevant to the feature space within which it is being categorized. Theory- and knowledge-based models of categorization also currently fail to address issues of presentation, unless presentation bears a specific causal relationship to why exemplars should be judged to be category members. However, presentation can have a large impact on fluency (for a review, see Schwarz, 2004), and to the extent that fluency is used as a cue in categorization, can play a role in categorization judgments as well. Overall, current models of categorization do not parsimoniously account for the effects of fluency.

2. Why there might be a link between fluency and category membership

There are a variety of reasons to suspect that there may be significant effects of fluency on categorization. First, researchers have suggested that fluency heuristics are used in a variety of different situations as a decision-making tool (e.g., Tversky & Kahneman, 1973; Whittlesea, Jacoby, & Girard, 1989; Lindsay & Kelley, 1996). Given the pervasive effects of fluency on these other aspects of decision-making, one might well suppose that categorization judgments would be anomalous if they were not affected by fluency.

Second, fluency can act as a robustly accurate way of making decisions about prior experiences (Reber et al., 1998). Exemplars that are highly typical of a category are primed more by the category than less typical exemplars. For example, if you are looking for an exemplar of the category "bird", "robin" should be more primed than "penguin." This greater priming in turn leads to a greater increase in fluency for the primed exemplar; that is, the primed exemplar will be easier to process in a variety of different ways (Meyer & Schvaneveldt, 1971; Stark & McClelland, 2000). If "robin" is more fluent – easier to produce or to remember – it might be in part due to the fact that it has been differentially primed by the activation of "bird". Over time, ease of processing should become correlated with category member typicality. This correlation may cause exemplar fluency to be taken into account in judgments of category typicality, whether by raising the judged typicality of fluent exemplars or lowering the judged typicality of unusually disfluent ones. In other words, the fluency of an exemplar ("robin") might affect how typical of the category ("bird") we think it is.

There have been several documented attempts to test whether fluency does, in fact, affect categorization, but to date the evidence is mixed. In two experiments, Buchner (1994) trained participants on consonant strings produced by a particular grammar and then tested them on the grammaticality of novel strings. Fluency was manipulated by gradually removing visual noise from each string in training until participants indicated that they could identify the string. Buchner found that although fluent strings – those strings which were identified more quickly – were judged to be more familiar, they were not classified as legal or illegal strings more easily. However, the grammaticality judgment that Buchner asked his participants to make differs from a categorization task in an important way. In contrast to judging category typicality, succeeding in a grammaticality judgment task requires knowing the correct answer. Greater fluency may have led to more false positives as well as more correct grammaticality judgments, leaving the overall rate of correct performance virtually unchanged (c.f. Whittlesea & Leboe, 2002).

Whittlesea and Dorken (1993) presented similarly indirect – although positive – evidence for effects of fluency on categorization. In one of a number of studies, they presented participants with nonce words generated by an artificial grammar. During the training phase, participants were asked to spell or pronounce these words, and then to repeat the pronounciation or spelling of the word from memory immediately afterwards. In the subsequent test phase, participants were told that the items they had seen were generated by a complex set of rules and were again asked to pronounce or spell a word before judging it to be either a ''good'' or ''bad'' word, depending on whether it conformed to those rules. They found that if an item was processed the same way at training and at test – that is, if it was either pronounced in both parts of the experiment or spelled in both – participants tended to classify it more accurately than if it was processed differently in the two different parts. However, since Whittlesea and Dorken reported no direct measure of the fluency of mismatched objects, this evidence must be taken as only indirectly supporting the hypothesis that fluency affects categorization. Westerman, Miller, and Lloyd (2003) have since reported that changes in modality or task can reduce the effects of fluency on recognition memory, but it remains an open question whether a comparable reduction in fluency across mismatched tasks was responsible for the effects reported by Whittlesea and Dorken.

Most recently, Whittlesea and Leboe (2002) ran a series of elegant studies giving evidence of strong effects of fluency on the classification of nonce words. They asked participants to learn nonce words with either glide or non-glide consonant transitions (i.e., BAMDEN vs. BAGDEN) by pronouncing them several times over the course of a training phase. They then asked their participants to make a forced choice decision about a set of similar but novel words, judging them to be either legal or illegal according to the abstract pattern found in the initial word set. They found that participants judged the more fluent, more easily pronounced glide stimuli as legal to the category more often than the less fluent, non-glide stimuli, even when participants had no knowledge about what the category was. In a separate experiment, Whittlesea and Leboe asked participants to perform the same classification task on stimuli that were presented either in uniform or alternating case (i.e., bAmDeN vs. BAMDEN) and found that participants were more likely to classify uniform case stimuli as legal.

Taken together, Buchner (1994), Whittlesea and Dorken (1993), and Whittlesea and Leboe (2002) along with others (e.g., Kinder, Shanks, Cock, & Tunney, 2003; Newell & Bright, 2001) present fairly solid evidence about the role of fluency in categorization judgments. However, all of these experiments use exclusively novel stimuli, about which participants had no prior experience or knowledge. None of the previous evidence directly addresses the issue of whether these effects of fluency

Please cite this article in press as: Oppenheimer, D. M., & Frank, M. C., A rose in any other font would not smell as sweet: ..., *Cognition* (2007), doi:10.1016/j.cognition.2007.05.010

are large enough to play a role in categorization decisions in which participants have a large body of conceptual knowledge about the exemplars, although there are solid theoretical reasons to believe that it should.

We propose that by manipulating the fluency of well-known category exemplars we can influence judgments about their pre-existing category memberships. Based on the priming relationship between typicality and fluency, good category members should ''feel'' more fluent. When fluency is independently manipulated, people are likely to misread their experience of fluency as resulting from high co-activation. Therefore, we should observe effects of fluency on typicality if participants use fluency as one of the variety of cues aiding in their categorization judgments.

3. Experiment 1

To test whether fluency would have an impact on category judgments in familiar domains, we used a standard font manipulation (Epley & Norwick, submitted for publication) and had participants rate exemplars on their typicality as members of a given category (e.g., how good a category member ''robin'' was for the category ''bird''). We predicted that making the names of the category members more difficult to read would decrease participants' typicality ratings.

3.1. Method

3.1.1. Participants

Seventy one Stanford University undergraduates participated as part of a course requirement. The survey was included in a packet of approximately 20 unrelated one-page questionnaires. Packets were distributed in class, and participants were given a week to complete the entire packet.

3.1.2. Design, stimuli and procedure

Participants were presented with a set of exemplars and a target category, and asked rate how good a category member each exemplar was on a scale of 1–9, (with 1 being least typical and 9 being most typical). Four categories – comprised of two natural kinds (bird and mammal), one artifact (vehicle), and one ad-hoc (unusual foods) – with 15 exemplars each were used. Exemplars were selected so as to vary both on the degree to which they were typical of the category and their commonness in everyday life (Stimulus materials are listed in Table 1). In the lowered fluency condition, the text of the study – including category members' names – was printed using 10 point Mistral type, while in the control condition the study was printed in 12 point Times New Roman type. In both conditions, instructions were presented in the same font as the category members in order to prevent participants from guessing the manipulation. The stimulus materials in the lowered fluency condition were still clearly legible, but significantly degraded from those in the control condition.

Table 1
The category exemplars used in Experiment 1

Bird	Mammal	Vehicle	Unusual food
Pigeon	Dog	Car	Kiwi
Hawk	Kangaroo	Airplane	Jicama
Ostrich	Elephant	Hang glider	Borsht
Robin	Human	Cart	Kumquat
Dodo	Dolphin	Submarine	Hamburger
Sparrow	Whale	Truck	Hummus
Eagle	Mole	Bicycle	Peanut butter
Turkey	Cow	Unicycle	Pasta
Goose	Koala	Carriage	Garbanzo beans
Hummingbird	Sloth	Tractor	Sausage
Wren	Platypus	Train	Pomegranate
Penguin	Bat	Scooter	Steak
Buzzard	Walrus	Skateboard	Chocolate
Mockingbird	Rabbit	Motorboat	Waffles
Parrot	Moose	Bus	Muffin

3.2. Results

Typicality ratings ranged from a mean of 1.6 in the case of from ‘‘pasta’’ in the unusual food category, to a mean of 8.9 for ‘‘car’’ in the vehicle category, confirming that the stimuli indeed varied in typicality.

The mean of the typicality ratings across categories was 6.11 out of 9 for low fluency exemplars and 6.26 out of 9 for control exemplars. A paired sample *t*-test compared mean typicality ratings for each exemplar in the low-fluency condition against mean typicality ratings for the corresponding exemplar in the control condition, and the differences were found to be reliable. ($t(59) = 2.00$, $p < 0.014$, Cohen’s $d = .52$).

3.3. Discussion

Simply changing the font of an exemplar can make people view it as less typical of a given category. This finding held over a wide range of categories and exemplars. Our findings were congruent with the results of Whittlesea and Leboe (2002), providing a demonstration that the effects they reported generalize to stimuli about which participants already know a great deal. Even when participants have pre-experimental notions about what features of an exemplar are relevant to category membership (e.g., ‘‘has wings’’ rather than ‘‘printed in a legible typeface’’), fluency still plays a significant role.

The proposed mechanism of these effects – namely that the priming relationship between categories and their members leads to fluency, which, in turn, serves as a cue for category membership – implies that fluency should be used as a cue whenever an individual rates the relationship between two items that prime one another. In experiment two, we test the generality of these effects by examining whether a similar

Please cite this article in press as: Oppenheimer, D. M., & Frank, M. C., A rose in any other font would not smell as sweet: ..., *Cognition* (2007), doi:10.1016/j.cognition.2007.05.010

font manipulation will influence ratings of the likelihood that category members will have a particular feature. If the effects in Experiment 1 were caused by some mechanism specific to the process of categorizing exemplars, then we would predict that varying fluency would not influence judgments of feature likelihood from categories. If, on the other hand, the effects are due to the fluency account then varying the fluency of features should also vary judgments of the features' likelihood, because features of a category should prime the category and vice versa.

4. Experiment 2

4.1. Method

4.1.1. Participants

Forty three Princeton University undergraduates were recruited from an introductory psychology course and asked to complete a single-page questionnaire as part of a group of studies done in exchange for course credit.

4.1.2. Design, stimuli and procedure

Participants were asked to rate on a scale of 1–9 (1 being least likely and 9 being most likely) the likelihood that any given member of a category had a given feature. For example, they were asked, "Imagine you come across a fish. What are the odds it has fins?" Six separate categories were listed, (fish, dog, bird, tree, pig, and cat), with five features varying in their typicality presented for each category (see Table 2). There were two between-subject conditions: a control condition and a lowered fluency condition. In the control condition, stimuli were presented in a normal, 12 point Arial font, while in the lowered fluency condition, stimuli were presented in italic Blackadder ITC font, printed in 50% gray color, halfway between black and white text colors (see Fig. 1). Again, lowered fluency stimuli were fully legible, but significantly degraded from those in the control condition.

Table 2
Category features used in Experiment 2

Fish	Bird	Dog
Has fins	Flies	Has a tail
Is less than 1 foot long	Eats worms	Barks
Is considered edible	Lays eggs	Wears a collar
Is colorful	Builds nests	Weighs less than 100 lbs.
Lives in a stream	Chirps/sings	Chases cats
Tree	Pig	Cat
Has leaves	Is pink	Kills birds
Is in a forest	Has a squiggly tail	Gets stuck up in trees
Produces fruit or flowers	Oinks	Is someone's pet
Houses birds	Lives on a farm	Meows
Is over 10 feet tall	Is dirty/muddy	Has soft fur

Please cite this article in press as: Oppenheimer, D. M., & Frank, M. C., A rose in any other font would not smell as sweet: ..., *Cognition* (2007), doi:10.1016/j.cognition.2007.05.010

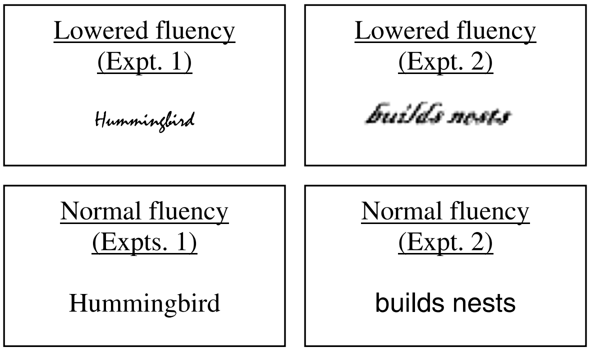

Fig. 1. Example stimuli from Experiments 1 and 2 in which lowered fluency was due to a font manipulation.

4.2. Results

Features ranged greatly in typicality from "gets stuck in trees" as a feature of cat (mean = 4.0) to "has fins" as a feature of fish (mean = 8.5). Features in the lowered fluency condition were judged to be, on average, less likely for any given exemplar (mean control = 6.70, mean lowered fluency = 6.54). A comparison of exemplar means in each condition in a paired-sample *t*-test revealed this difference to be reliable ($t(29) = 2.13$, $p = 0.042$, Cohen's $d = .79$).

4.3. Discussion

The differential priming model predicted that lowering fluency would lower exemplar typicality judgments (in Experiment 1) and would also lower judgments of feature likelihood (in Experiment 2). This prediction was borne out by our data. However, there are several alternative explanations that could also account for these data. For instance, participants might feel frustration with the difficulty of pronounciation (in the case of the Whittlesea and Leboe artificial words) or experience a general decrease in affect from the difficulty of reading the stimulus materials, which might cause them to give lower ratings.

To eliminate affective confounds, one must dissociate the predictions of an affective model and a fluency-based model. This can be done through the phenomenon of causal discounting. In explaining events, we tend to prefer a single cause to multiple causes (Einhorn & Hogarth, 1986; Kelley, 1973). Thus, the known presence of one cause for a particular outcome tends to lower people's likelihood estimates of the presence of other causes. Research has demonstrated that people often attempt to make causal attributions about their meta-cognitive experience of fluency (Oppenheimer, 2004a, 2005; Schwarz, 2004). As such, when there is a salient explanation for fluency that is not related to the judgment at hand, people tend to discount fluency as a cue.

Please cite this article in press as: Oppenheimer, D. M., & Frank, M. C., A rose in any other font would not smell as sweet: ..., *Cognition* (2007), doi:10.1016/j.cognition.2007.05.010

In a series of experiments, Oppenheimer (2004a) showed that individuals failed to use the availability heuristic (how easily something comes to mind) to make judgments about frequency when an alternate explanation for the availability of the particular stimulus was salient. He asked participants to judge surname frequency (a judgment usually made by using the availability heuristic) in cases in which some surnames were held by famous individuals (i.e., Clinton) and in cases in which they were not (i.e., Callaway). When surnames were famous, participants tended to rate them as less common, discarding availability information based on the salience of an alternate explanation for availability: fame. In other words, individuals disregarded an internal heuristic cue because there were specific alternate causes for that judgment.

This suggests that in the presence of a salient alternative explanation for decreased fluency, participants should actually overcompensate for lowered fluency, giving higher ratings of typicality than they might otherwise. However, in the case that these effects are due to lowered affect, and that unhappy participants give lower ratings, there is no reason to suppose that the presence of a salient explanation for lowered fluency should reverse the findings.[2]

As such, in Experiment 3 we dissociate fluency and affect by providing a salient but unstated explanation for the lowered fluency of the stimulus materials: the printer on which they were printed was running out of toner. This manipulation was based upon previous work demonstrating that participants observing low toner print tend to spontaneously attribute lack of fluency to the toner issues, and subsequently engage in discounting (Oppenheimer, 2005, Experiment 5). Across four studies, Oppenheimer demonstrated that the reduced fluency due to complex vocabulary, grammar, and font led to lower intelligence ratings of a text's author. However, when participants viewed texts in a "low toner" font, they rated the author as *more* intelligent. Participants spontaneously identified that the source of their disfluency was unrelated to the judgment at hand (i.e., the author's intelligence) and in an attempt not to be biased by their fluency experience ended up overcompensating and showing bias in the other direction (c.f. Oppenheimer, 2004a).

If participants accept this as a cause for lowered fluency, we predict that they will over-discount fluency as a factor in their typicality judgments, leading to ratings of typicality higher than those in the control condition. However, if the lowered fluency of the stimulus materials simply elicits a different affective response, typicality judgments in the low toner condition should be lower than typicality judgments in the control condition, congruent with Experiments 1 and 2.

[2] Note that Schwarz and Clore (1983) reported that participants in negative affective states are more likely than those in positive affective states to search for and explain their state based on available information. However, this type of affective discounting would not explain an increase in typicality judgments. Participants might use available information to explain a negative affective state, but they would still be in that affective state and hence still be predisposed to give lowered judgments of typicality or feature likelihood.

5. Experiment 3

5.1. Method

5.1.1. Participants

One hundred sixty four Stanford University undergraduates participated as part of a course requirement. The survey was included in a packet of approximately 20 unrelated one-page questionnaires. Packets were distributed in class, and participants were given a week to complete the entire packet.

5.1.2. Design, stimuli and procedure

The design and stimuli were identical to those of Experiment 1, save a change in the experimental condition. In the experimental, or 'low toner' condition, category members' names were printed in the same Times New Roman 12 point font as in the control condition, however the surveys were printed in such a way as to simulate having been printed on a laser printer running out of toner (see Fig. 2).

5.2. Results and discussion

Ratings of typicality were higher in the low toner condition (mean = 6.27) than in the control condition (mean = 6.12). A paired sample *t*-test compared mean typicality ratings for each exemplar in the low toner condition against mean typicality ratings for the corresponding exemplar in the control condition, and the differences were found to be reliable ($t(59) = 2.00$, $p = .01$, Cohen's $d = .52$).

The results from Experiment 3 suggest that these effects really are due to differential fluency: since hearing or reading the name of a category primes its most typical members, the fluency associated with this priming is a reliable cue for typicality. Based on Oppenheimer's (2004a) results we know that people will tend to overcompensate for the perceived influence of fluency when faced with a salient alternate explanation for that fluency. In other words, participants in Experiment 3 judge low-toner category members to be more typical because they are over-correcting for the (now salient) lowered fluency of the stimuli. The alternative, affective account of these effects predicts that typicality judgments would be decreased by the negative affect caused by lowered fluency stimuli.

Category of Vehicles:

___ Car	___ Truck	___ Train
___ Airplane	___ Bicycle	___ Scooter
___ Hang Glider	___ Unicycle	___ Skateboard
___ Cart	___ Carriage	___ Motorboat
___ Submarine	___ Tractor	___ Bus

Fig. 2. Example stimuli from Experiment 3 in which lowered fluency was due to a printer running out of toner.

Please cite this article in press as: Oppenheimer, D. M., & Frank, M. C., A rose in any other font would not smell as sweet: ..., *Cognition* (2007), doi:10.1016/j.cognition.2007.05.010

Up to this point, our studies have exclusively used a between-subject design. While this design is standard for investigations of fluency (for reviews, see Schwarz, 1998, 2004; Skurnik, Schwarz, & Winkielman, 2000) and has logistical advantages, it also suffers from certain drawbacks. Specifically, results from between-subject studies are often difficult to interpret because they are vulnerable to statistical artifacts caused by aggregating data across subjects (Monin & Oppenheimer, 2005; Nickerson, 1995). To remedy this shortcoming, in Experiment 4 we replicated the findings from Experiment 1 using a within-subject design.

6. Experiment 4

6.1. Method

6.1.1. Participants

Fifteen Princeton University and 24 MIT undergraduates were recruited from an upper division cognitive psychology course and an introductory psychology course, respectively. Each group participated in exchange for candy. Participants were tested in large groups, but were seated at different tables and as such were unable to see the materials and responses of the other participants.

6.1.2. Design, stimuli and procedure

The stimuli were identical to those in Experiment 1. The experiment was conducted in two phases, which took place exactly one week apart. In phase one, the participants were randomly assigned to either the fluent or disfluent condition and filled out a survey in the corresponding font. In phase two, participants completed an identical survey, only in a different font; those participants who had initially filled out the survey in a fluent font were given the same survey in a disfluent font, while those who had initially been given a disfluent font were now given the fluent font.

6.1.3. Results and discussion

Because several of the participants failed to complete the entire survey, complete data sets were only obtained for the "birds" and "mammals" categories, and so only those categories were included in the analysis. Participants' ratings of category membership at time 1 were strongly correlated with their ratings at time 2 (mean $r = .78$) indicating that participants were fairly consistent in their ratings across time. We conducted two analyses, averaging first across items and then across participants. We found that there was an average increase in typicality ratings of 0.46 points on our scale, which was highly reliable across items ($t(27) = 11.49$, $p = 6.67^{-12}$, Cohen's $d = 2.17$), replicating the results of our previous experiments. In the participants analysis, we also found a significant effect of fluency (mean = 0.46), which was also reliable ($t(38) = 3.01$, $p = 0.005$, Cohen's $d = 0.48$). These results show a replication of the previous experimental results in a more sensitive, within-subjects paradigm. Because of the greater power of this within-subjects paradigm, we found a very large effect for individual items. In addition, despite the large variability across participants, the effects of fluency were still visible within individual participants, providing evidence that the differences in typicality judgments we observed in Experiments 1–3 were not artifacts of averaging across participants.

7. General discussion

Four experiments have provided evidence that the perceptual fluency of a stimulus affects categorization judgments. Perceptual fluency appears to be used as one of several heuristic cues for categorization. The reason for the reliability of this cue is as follows: every categorization decision involves both an exemplar and a category. The more typical the exemplar is to the category, the stronger the representational link between them, and the more the activation of one will prime the other. Since the main consequence of priming is an increase in fluency (ease of processing and ease of production), there is a reliable relationship between fluency of a category (given that the category member is activated) and that member's typicality. Put differently, if you want to judge what the best exemplars of a category are, start naming them off the top of your head. The exemplars you name will be those most strongly linked to the category representation. Why? They are the most fluent exemplars, and this fluency comes as a result of their being primed by the activation of the original category representation. As such, while they are not foolproof indicators of category membership, heuristic cues such as fluency provide useful and easily available sources of information for decision-making.

Our proposal regarding the role of fluency in typicality is consistent with the differential judgments reported. In Experiment 1, we demonstrated the basic phenomenon of interest, namely that when we lowered the perceptual fluency of category exemplars by reducing the font size they were written in, they were judged to be less typical. Since people often have difficulty determining the source of fluency (Jacoby & Dallas, 1981; Schwarz, 2004), their fluency experience is misinterpreted as being due to the conceptual link between an exemplar and a category; as such, the differential fluency is used as a metacognitive cue and influences category judgments. In Experiment 2, we demonstrated the robustness of this account by having participants make judgments of features instead of exemplars in relation to categories. Again, participants judged the stimuli whose fluency had been lowered as less typical, despite the fact that they were identical in content to the control stimuli. In order to rule out affective explanations, we tested the hypothesis that, like other heuristic cues used in judgments, fluency is subject to causal discounting. When presented with a salient alternative explanation for the lowered fluency of one stimulus set (namely, that the laser printer on which they had been printed was running out of toner), participants overcompensated for the now-salient lack of fluency by rating category exemplars as more typical rather than less typical, as they had before. Once participants became aware that the fluency experience was not due to a conceptual link between an exemplar and a category, they attempted to disregard fluency as a cue, but overcompensated and erred in the other direction (as in Oppenheimer, 2004a, 2005; Schwarz, 2004). In Experiment 4, we demonstrated the effect of fluency using a within-subject design.

7.1. Implications for theories of categorization

The results presented here, though small, are not typically predicted or accounted for in current models of categorization. Exemplar and prototype models, which rely on shared semantic features of category exemplars to make decisions about category membership, generally do not take into account experiential information in making categorization decisions. Although a feature-based exemplar or prototype model could account for fluency effects by including fluency as a feature, this sort of ad-hoc addition fails to acknowledge that the effect of fluency may be heuristic and meta-cognitive – as our evidence suggests, since it is subject to discounting – rather than being a product of a feature based general categorization process.

Further, as categories and their exemplars are conceptually related for all categories, high fluency would have to be a universal feature across all categories; a form of representation that seems fairly unparsimonious. One possible alternative would be a theory in which heuristic information modifies traditional feature-based judgments. In other words, the model would judge the typicality of "robin" before adjusting that judgment according to heuristic input. However, this solution may beg the question: why add heuristic information after a categorization judgment has already been arrived at? Why not leave it out altogether and avoid the errors that heuristic cues may sometimes introduce?

Likewise, causal and knowledge-based models of categorization are bound to a presentation-independent view of categorization unless they explicitly acknowledge presentation as part of a category's causal model. Barring this modification, the particular stimulus presentation which engages conceptual knowledge should not affect the overall outcome in categorization decisions, thus rendering them unable to account for heuristic categorization phenomena. One possible solution to this would be a causal based framework in which features unrelated to category membership could be represented (e.g., Oppenheimer, Tenenbaum, & Krynski, in preparation). In such a model, metacognitive factors could be used as causal explanation of features, or even as features themselves. Such models are still in the preliminary stages of development, but show great promise for accounting for heuristic judgment in categorization.

Another class of promising models includes Gluck and Bower's (1988) account of categorization based on the informational value of different cues. This model, an extension of Rescorla and Wagner's (1972) model of animal learning, focuses on their participants' use of the validity of a particular cue relative to other competitor cues rather than the objective probability that a particular cue is associated with a stimulus. In general, models which are agnostic as to what constitutes a cue will be able to integrate fluency information with semantic features to account for the results reported here. While a standard exemplar model might require the inclusion of fluency as a feature for every stimulus to be categorized, a more flexible model should identify fluency as just one of many competing cues that may offer information about typicality or feature likelihood. Even though fluency may be robust as a cue for typicality, the superior validity of semantic cues would predict that fluency would play a relatively small role in any given decision. Indeed, the effects of fluency in our experiments, while reliable, are quite small in magnitude. Further, our studies asked participants to make their judgments based on a word or a description, not a feature-laden representation – perhaps this lack of immediate link to the features of a particular stimulus increased participants' reliance on other perceptual cues, such as fluency.

8. Conclusion

The data presented here are indicative of a general cognitive trend. The pattern observed in our data – that of using fluency for making judgments – is hardly unique to categorization. The work of Tversky and Kahneman and the extensive literature in the heuristics and biases tradition indicates that a wide variety of judgment tasks are carried out through a process of learning multiple, easily-identified metacognitive patterns that substitute for more computationally intensive evaluation tasks. Indeed, taken in this context, categorization judgments can be seen simply as one type of judgment, in which human beings use a variety of fast and easy shortcuts whenever possible, only resorting to slow and effortful processing when the task requires it. While category learning and category judgments in novel situations may rely primarily on featural and conceptual information, as we become more familiar with exemplars, features, and categories, we gain access to a new set of cues – heuristic cues – that can simplify the categorization process.

References

Alter, A.L. & Oppenheimer, D.M. (submitted for publication). Uniting the tribes of fluency into a metacognitive nation. Quarterly Journal of Experimental Psychology.

Benjamin, A. S., Bjork, R. A., & Hirshman, E. (1998). Predicting the future and reconstructing the past: A Bayesian characterization of the utility of subjective fluency. *Acta Psychologica, 98*, 267–290.

Buchner, A. (1994). Indirect effects of synthetic grammar learning in an identification task. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*(3), 550–556.

Einhorn, H. J., & Hogarth, R. M. (1986). Decision making under ambiguity. *Journal of Business, 59*(4), 225–250.

Epley, N. & Norwick, R.J. (submitted for publication). Confidence as Inference from Incidental Experience.

Gluck, M. A., & Bower, G. H. (1988). From conditioning to category learning: An adaptive network model. *Journal of Experimental Psychology: General*.

Goldstone, R. L., & Kersten, A. (2002). Concepts and categorization. In A. F. Healy & R. W. Proctor (Eds.), *Comprehensive handbook of psychology, Volume 4: Experimental psychology*. New Jersey: Wiley.

Jacoby, L. L., & Dallas, M. (1981). On the relationship between autobiographical memory and perceptual learning. *Journal of Experimental Psychology: General, 110*, 306–340.

Keil, F. C. (1989). *Concepts, kinds, and development*. Cambridge, MA: Bradford Books/MIT Press.

Kelley, H. H. (1973). The process of causal attribution. *American Psychologist, 28*, 107–128.

Kinder, A., Shanks, D. R., Cock, J., & Tunney, R. J. (2003). Recollection, fluency and the explicit/implicit distinction in artificial grammar learning. *Journal of Experimental Psychology: General, 132*(4), 551–565.

Kleider, H. M., & Goldinger, S. D. (2004). Illusions of face memory: Clarity breeds familiarity. *Journal of Memory and Language, 50*, 196–211.

Lindsay, D. S., & Kelley, C. M. (1996). Creating illusions of familiarity in a cued recall remember/know paradigm. *Journal of Memory and Language, 35*, 197–211.

Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review, 85*, 207–238.

Meyer, D. E., & Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology, 90*, 227–234.

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

Monin, B., & Oppenheimer, D. M. (2005). Correlated averages vs. averaged correlations: Demonstrating the warm glow heuristic beyond aggregation. *Social Cognition, 23*, 257–278.

Murphy, G. L., & Medin, D. L. (1985). The role of theories in conceptual coherence. *Psychological Review, 92*, 289–316.

Newell, B. R., & Bright, J. E. H. (2001). The relationship between the structural mere exposure effect and the implicit learning process. *Quarterly Journal of Experimental Psychology, 54*(4), 1087–1104.

Nickerson, C. A. E. (1995). Does willingness to pay reflect the purchase of moral satisfaction? A reconsideration of Kahneman and Knetsch. *Journal of Environmental Economics and Management, 28*, 126–133.

Norwick, R.J., & Epley, N. (2003). Experiential determinants of confidence. Poster presented at the Society for Personality and Social Psychology, Los Angeles, CA.

Oppenheimer, D. M. (2004a). Spontaneous discounting of availability in frequency judgment tasks. *Psychological Science, 15*(2), 100–105.

Oppenheimer, D. M. (2005). Consequences of erudite vernacular utilized irrespective of necessity: Problems with using long words needlessly. *Journal of Applied Cognitive Psychology, 20*(2), 139–156.

Oppenheimer, D.M., Tenenbaum, J.B., & Krynski, T.R. (in preparation). Categorization as causal explanation: Discounting and augmenting in a Bayesian framework.

Reber, R., Winkielman, P., & Schwarz, N. (1998). Effects of perceptual fluency on affective judgments. *Psychological Science, 9*(1), 45–48.

Reber, R., Wurtz, P., & Zimmerman, T. (2004). Exploring "fringe" consciousness: The subjective experience of perceptual fluency and its objective bases. *Consciousness and Cognition, 13*, 47–60.

Rehder, B., & Hastie, R. (2001). Causal knowledge and categories: The effects of causal beliefs on categorization, induction, and similarity. *Journal of Experimental Psychology: General, 130*, 323–360.

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and non-reinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning II: Current research and theory*. New York: Appleton-Century-Crofts.

Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology, 7*, 573–605.

Rothman, A. J., & Schwarz, N. (1998). Constructing perceptions of vulnerability: Personal relevance and the use of experiential information in health judgments. *Personality and Social Psychology Bulletin, 24*, 1053–1064.

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

Schwarz, N. (2004). Meta-cognitive experiences in consumer judgment and decision making. *Journal of Consumer Psychology, 14*(4), 332–348.

Schwarz, N., & Clore, G. L. (1983). Mood, misattribution, and judgments of well-being: Informative and directive functions of affective state. *Journal of Personality and Social Psychology, 45*(3), 513–523.

Skurnik, I., Schwarz, N., & Winkielman, P. (2000). Drawing inferences from feelings: The role of naive beliefs. In H. Bless & J. P. Forgas (Eds.), *The message within: The role of subjective experience in social cognition and behavior* (pp. 162–175). Philadelphia, PA: Taylor & Francis.

Stark, C. E. L., & McClelland, J. L. (2000). Repetition priming of words, pseudowords, and nonwords. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*(4), 945–972.

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology, 5*(2), 207–232.

Westerman, D. L., Miller, J. K., & Lloyd, M. E. (2003). Change in perceptual form attenuates the use of the fluency heuristic in recognition. *Memory and Cognition, 31*(4), 619–629.

Whittlesea, B. W. A. (1993). Illusions of familiarity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*(6), 1235–1253.

Whittlesea, B. W. A., & Dorken, M. D. (1993). Incidentally, things in general are particularly determined: An episodic-processing account of implicit learning. *Journal of Experimental Psychology: General, 122*, 227–248.

Whittlesea, B. W. A., & Leboe, J. P. (2002). The heuristic basis of remembering and classification: Fluency, generation, and resemblance. *Journal of Experimental Psychology: General, 129*, 84–106.

Whittlesea, B. W. A., Jacoby, L. L., & Girard, K. (1989). Illusions of immediate memory: Evidence of an attributional basis for feelings of familiarity and perceptual quality. *Journal of Memory and Language, 29*, 716–732.

Whittlesea, B. W. A., & Williams, L. D. (2000). The source of feelings of familiarity: The discrepancy-attribution hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*(3), 547–565.

Winkielman, P., Schwarz, N., Fazendeiro, T. A., & 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.

Witherspoon, D., & Allan, L. G. (1985). The effect of a prior presentation on temporal judgments in a perceptual identification task. *Memory and Cognition, 13*(2), 101–111.

Please cite this article in press as: Oppenheimer, D. M., & Frank, M. C., A rose in any other font would not smell as sweet: ..., *Cognition* (2007), doi:10.1016/j.cognition.2007.05.010