tapraid5/zfn-xhp/zfn-xhp/zfn00217/zfn3548d17z | xppws | S=1 | 12/22/16 | 11:21 | Art: 2016-1106

APA NLM

<Journal of Experimental Psychology: Human Perception and Performance

0096-1523/17/$12.00 http://dx.doi.org/10.1037/xhp0000343

AQ: 1

Objective Effects of Knowledge on Visual Perception

AQ: au
AQ: 2

Gary Lupyan
University of Wisconsin–Madison

To what extent is what we perceive influenced by what we know? Although a large literature purports to show effects of knowledge, expectations, and other cognitive states on various aspects of perception, strong counterarguments have been advanced that these demonstrations are confounded by nonperceptual factors. For example, although letters are easier to recognize in meaningful words than meaningless letter strings, skepticism remains that such effects of knowledge on visual recognition mean that knowledge literally helps people see. In Experiment 1, a perceptual matching task is used to show that meaningful words look sharper than meaningless letter strings. In Experiments 2 through 4, it is shown that people are more accurate in detecting subtle changes in blur when they occur in meaningful words compared with meaningless letter strings. In Experiment 5, it is shown that this improvement in performance cannot be explained solely by differences in visual familiarity, but is predicted by semantic factors such as word imageability. These results provide a strong empirical rejoinder to claims that perception is encapsulated from knowledge.

Public Significance Statement

Does having some prior knowledge about what you are looking at help you see it? This article presents several experiments arguing that it does. The first study shows that meaningful words such as "much" appear sharper to people than meaningless letter combinations (such as "mcuh"). The subsequent 4 studies show that people are much better able to detect subtle changes to meaningful words than to meaningless letters or to less meaningful words. This work supports the idea that what we see depends not just on what we are looking at, but also on what we know and expect.

Keywords: change detection, knowledge, reading, recognition, top-down effects

Supplemental materials: http://dx.doi.org/10.1037/xhp0000343.supp

Suppose an optometrist asks you to look at a line on a Snellen chart while testing your eyesight. The letters look a bit blurry, but you can make out the first as "R." To what extent does your knowledge of what an "R" looks like actually help you perceive it? Beyond single letters, to what extent does recognizing a combination of letters or object features as comprising a meaningful word or object help in perceiving it?

Much hinges on this seemingly simple question. According to some theories, perception provides the material from which higher level representations (such as words and objects) are built (e.g., Riesenhuber & Poggio, 2000; Serre, Oliva, & Poggio, 2007; Treisman & Gelade, 1980). On these views, perception is an important *source* of knowledge, but is itself encapsulated from its effects. Because the mapping from visual representations to meaning happens later, meaningfulness cannot affect perception per se (Firestone & Scholl, 2015; Pylyshyn, 1999).[1] On an alternative formulation, what we perceive depends both on the input and on the cognitive state of the viewer (see Kahan, 2016; Lupyan, 2015; O'Callaghan et al., 2016 for recent reviews). If true, theories of perception need to incorporate mechanisms by which perception is constrained and augmented by a viewer's knowledge and expectations. But to what extent is what we see influenced by what we know? Supporting evidence seems to abound. Numerous studies show perception to be influenced by the viewer's expectations (e.g., Summerfield & Egner, 2009 for review). But showing that our perception is influenced by knowledge faces a hurdle: we need to consider that demonstrations of knowledge influencing perception may instead be cases of knowledge influencing how we remember or interpret what we perceive. For example, it is well-established that people's ability to recognize letters is substantially improved when the letters are part of a real, meaningful word (Balota, Yap, & Cortese, 2006 ; McClelland & Rumelhart, 1981;

Fn1

The study was funded by NSF PAC 1331293 to the author. Special thanks to Jonathan Lang for intellectual input and help with stimulus creation, and to Sarah Kraemer for help with data collection.

Correspondence concerning this article should be addressed to Gary Lupyan, 1202 West Johnson Street, Madison, WI 53706. E-mail: lupyan@wisc.edu

[1] On virtually all theories of perception, perception is viewed as an intrinsically interactive and inferential process (Barlow, 1997; Gregory, 1980), but the "inferences" in question are sometimes thought to operate within a putative perception module, as when one kind of perceptual domain (e.g., shading) informs another (e.g., 3D structure). Therefore, the claim that perception is inferential does not mean that what we see is in any way influenced by what we know.

Reicher, 1969 for review). If it were the case that the improved recognition performance stemmed from people being better at remembering letters within a meaningful word, or people being better at guessing what those letters might be, such evidence would not constitute an effect of knowledge on perception per se.

Carefully controlled studies have convincingly ruled out such memory and guessing based accounts of the word superiority effect (e.g., Johnston, 1978; Jordan & Thomas, 2002; Prinzmetal & Lyon, 1996; see McClelland, Mirman, & Holt, 2006 for discussion of analogous arguments in auditory speech perception). Yet, skepticism remains that such findings still do not speak to knowledge affecting *perception* (Firestone & Scholl, 2015; Lupyan, 2015; Raftopoulos, 2015). For example, in a recent critique of top-down effects on perception, Firestone and Scholl write that while "top-down effects on perception are meant to be effects on what we see," "many such studies report effects on how we *recognize* various stimuli" (2015, sect. 4.6, italics in original). It can be argued that distinguishing perception from recognition in this way is problematic (Lupyan, in press). Nevertheless, it is true that many studies purporting to show effects of knowledge on perception rely on recognition as the primary measure. Although guessing and memory-based accounts of the word-superiority effect have been ruled out to most researcher's satisfaction (see Balota et al., 2006 for discussion), recognition measures have been argued to tap processes downstream of "perception proper" (see Firestone & Scholl, 2015 for discussion). By relying on measures of perceptual processing that lie closer to what is commonly studied in vision science, I hope to provide more direct evidence that what we know affects what we see.

In Experiment 1, I ask whether meaningful words look sharper than visually similar, but meaningless letter strings. I do this by using a perceptual matching task in which people adjust the blurriness of one letter string to match another letter string. If people see meaningful words sharper/more clearly than meaningless pseudowords, then a word that is at the same objective level of blur as a pseudoword should look sharper and the pseudoword would need to be made sharper to match (see Levin & Banaji, 2006 for similar logic). Why might meaningful words look sharper than pseudowords? There are several possibilities. The first is that perception of meaningful words more so than of pseudowords is mediated by representations that bind together commonly occurring letter features and letter sequences into a visual lexical representation that abstracts away much of the noisiness and idiosyncrasy of the current visual input (Humphreys, Evett, & Quinlan, 1990; Wiley, Wilson, & Rapp, 2016). Insofar as our perception of real words is mediated by these representations to a greater degree than our perception of meaningless letter sequences, we would expect letter strings that map onto such higher level representations to be perceived more clearly. An alternative formulation is that our perception is mediated by the very same visual representations regardless of whether the word is meaningful or not, but the higher-level representations (partly) activated when viewing a low-quality stimulus helps to "clean it up" (this interactive account is similar to that originally proposed by the connectionist model of the word superiority effect, McClelland & Rumelhart, 1981; Rumelhart & McClelland, 1982; see Jordan, Thomas, & Scott-Brown, 1999 for an especially vivid example of such a clean up process).

An effect of knowledge on apparent sharpness is likely not an end in itself, but rather a *consequence* of the perceptual system performing some computation more effectively. In Experiments 2 through 5, I tested the effects of prior knowledge on one such computation—people's ability to detect subtle changes to sharpness. People viewed strings of words or letters that became sharper or blurrier right before their eyes and had to indicate whether the letter string changed, and how. In addition to investigating a further consequence of meaningfulness on perception, these experiments help distinguish between the two accounts mentioned above. If meaningful words appear sharper than pseudowords because the perception of the former but not the latter is mediated by lexical representations, then people should be *worse* at seeing subtle changes in blur to a meaningful word because a partially recognized word and slightly more recognized word would both map onto the very same lexical representation. Alternatively, within frameworks that stress prediction as a key aspect of perception (e.g., Clark, 2013; Kveraga, Ghuman, & Bar, 2007; Noppeney et al., 2008; Rao & Ballard, 1999), we can think of meaningful words as being processed within a stronger set of expectations (priors) compared with pseudowords, which helps in detecting deviations between what is expected and what is observed (see Lupyan, 2013 for a similar argument in the domain of unexpected colors and color afterimages). On this account, people should be *better* able to detect subtle changes in blur to meaningful words.

Blur detection may appear to be an odd and even contrived task. However, our visual system appears to regularly compute and compare relative blur in the service of depth perception (Held, Cooper, & Banks, 2012). Because we can only focus on a single plane detecting blur in an object signals that the object is closer or farther away than the plane of focus. The computational mechanisms of blur detection are fairly well understood (Burge & Geisler, 2011; Wang & Simoncelli, 2003; Watson & Ahumada, 2011; Webster, Georgeson, & Webster, 2002), though less so than those underlying detection of other perceptual changes, such as motion.

But perhaps the best argument for what makes blur detection a good domain in which to investigate effects of knowledge on perception is that it is closely related to our intuitive notion of what it means to see well. Detecting differences in blur is essentially what we do when we are being fit for eyeglasses and the optometrist asks if option (prescription) A or option B looks clearer.

Experiments 1A and 1B

Participants were asked to adjust the level of blur in meaningless letter strings (pseudowords) to match the blur of visually similar meaningful letter strings, or vice versa. If a meaningful word (e.g., "much") appears sharper than a pseudoword (e.g., "mcuh"), then when presented at the same objective level of blur, the meaningful word would look sharper and participants would need to make the pseudoword sharper to match.

Method

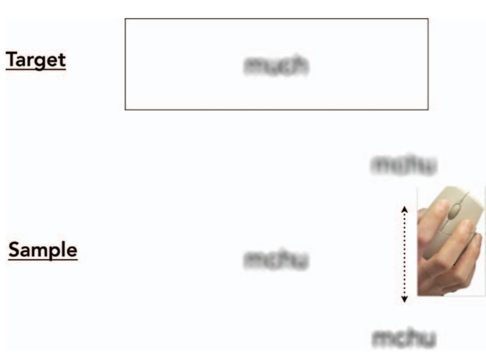
Participants. A total of 35 undergraduate students were tested in exchange for course credit: $n = 15$ in Experiment 1A; final $n = 16$ in Experiment 1B after removing 2 participants for responding at random and 1 for primarily responding with extreme values (0s and 1s). There was no formal procedure for determining sample size for Experiment 1A apart from the author's experience with running similar within-subject designs. The sample size for Experiment 1B was set to approximately match that of Experiment 1A.

Materials. The target stimuli were 15 different four-letter words (e.g., "seem," "worn," "much"; see supplementary materials for full list and OSF repository at osf.io/h6m) printed in an Arial font. For each word a matched pseudoword was created by scrambling letter order (e.g., much→mcuh). The words were then blurred using Adobe Photoshop's version CS6 "Field Blur" filter which is designed to closely mimic an out of focus lens. The exact algorithm is proprietary, and therefore not available to the author. An almost identical effect can be obtained by convolving the original image with a circular disk using ImageMagick with the command:

```
convert originalImage.png -define convolve:scale = ! -morphology Convolve Disk:6 blurredImage.png
```

Blurring selectively removes high spatial frequencies which are relatively more important to reading than low spatial frequencies, though reading words with high spatial frequencies shows all the signatures of normal reading (Jordan, Dixon, McGowan, Kurtev, & Paterson, 2016a, 2016b). Two versions of each letter string were pre-generated, one at a field blur level of 6 and one at 8 (which closely corresponds to the same parameter value in the ImageMagick command above). A blur continuum was then created by varying the opacity of the sharper stimulus placed on top of the blurrier one. The supplementary movies (osf.io/h6mqf) shows the blur in action.

Procedure. On each trial, participants saw a blurred word—the F1 *target*—inside a centrally presented box (see Figure 1). Below the target was the *sample* which participants were instructed to adjust to the same level of blurriness as the target by using the mouse-wheel. The targets were presented at four levels of blur—.3, .4, .6, and .7—where the values indicate the location of the stimulus on the blur continuum. A value of .3 means the stimulus was 30% of the way between the sharpest and the blurriest endpoint. At the start of a trial the sample was set to the midpoint (.5) such that participants had to blur it further on half of the trials and sharpen it on the other half. Participants completed 10 practice trials followed by 120 testing trials with each unique letter string shown four times during the study. The target and sample words were the same size, subtending approximately 0.9° (height) × 2.5° (width) of visual angle. A walkthrough of the task can be viewed online at osf.io/h6mqf.



ONLINE COLOR

Figure 1. A sample trial from Experiment 1 (meaningful trial shown). Participants adjusted the sharpness of the sample to match the sharpness of the target. Both the sample and target remained visible the entire time. See AQ: 4 the online article for the color version of this figure.

Experiment 1B was identical to Experiment 1A except in half the trials (the *same* trials), the target and sample were the same exact letter string. The purpose of Experiment 1B was to replicate the results of Experiment 1A and to further rule out effects of response bias (see below). To keep the length of the experiment the same as Experiment 1A, Experiment 1B used only two starting levels of blur: .3 and .7.

Analytic approach. Analyses for all studies were conducted using mixed-effects linear models (R version 3.2.2: lme4 version 1.1.10). Responses were modeled using linear regression on the raw data. p values were computed using Satterthwaite approximations (lmerTest package). Predictors were centered. All models included subject and item random intercepts and random slopes unless the slopes prevented model convergence or if the inclusion of a random intercepts substantially inflated the correlation between fixed effects.

Results

Participants' final responses for various levels of blur are shown in Figure 2A. In all conditions there was a tendency to underesti- F2 mate the amount of blur, $b = .74$, 95% CI = [.67, .80]. Importantly, for all levels of blur, to obtain a suitable match between a meaningful target and a meaningless sample, participants needed to make the sample sharper than the meaningful target. Adjusting them to the same objective level of blur, would, evidently, make a meaningful sample appear too sharp. Conversely, to obtain a suitable match between a meaningless target and a meaningful sample, participants needed to make the sample blurrier, $b = -.11$, $t = -5.36$, $p = .0001$, 95% CI = [−0.15, −0.07]. This pattern of results is consistent with the prediction that meaningful words appear sharper than meaningless letter strings.

Might participants' responses reflect something other than the perceived level of blur? For example, perhaps participants just made the samples as sharp as possible if they could read the target and as blurry as possible if they could not read it. We can address this possibility in two ways: First, we can eliminate all responses that were at floor or ceiling (7.5% of all responses). Repeating the analysis above with these extreme responses removed yielded an equally strong effect of meaningfulness, $b = -0.09$, $t = -4.87$ $p = .0003$. Second, we can control for individuals' likelihood of responding at floor or ceiling (which ranged from 0% to 18.3%) and include this rate as a covariate in the model. Doing so leaves the effect essentially unchanged, $b = -.11$, $t = -5.36$, $p = .0001$.

Also speaking against the use of such a discrete response strategy is the linearity of the slopes shown in Figure 2A. It is possible, however, that this average linearity masks severe nonlinearities at the level of individual participants. Examining the individual data revealed that blurriness of the target was a positive, highly significant predictor in a linear model ($t > 2.75$; $p < .01$) for 13 of the 15 participants. Removing the two participants whose responses did not track the objective blur in as linear a way only strengthened the effect of meaningfulness, $b = -0.12$, $t = -5.83$, $p < .0001$.

Finally, perhaps participants responded more diligently for some trials than others and it is this difference in diligence is responsible for the effect. We can test this possibility in three ways. First, we can include absolute error as a covariate. Doing so yields essentially unchanged results $b = -0.10$, $t = -5.24$, $p = .0001$. Second, it can be observed that on the majority of the trials

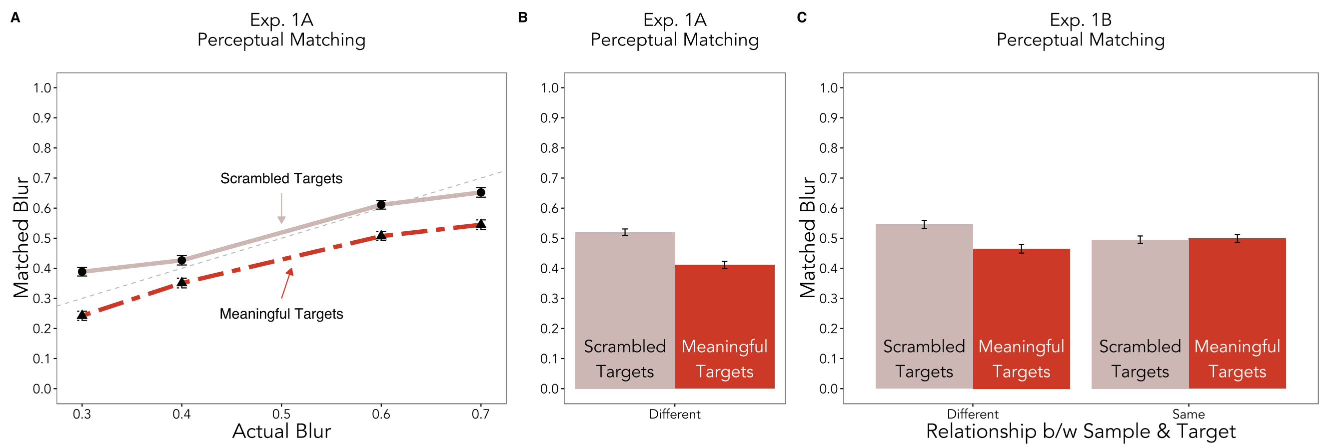


Figure 2. (A) Results of Experiment 1A. Greater values indicate greater level of blur. Solid line: participant matched blur for scrambled targets (meaningless samples); Dashed line: performance for meaningful targets (scrambled samples). Light dashed line indicates perfect performance for contrast. (B) Collapsed results showing that, for the same level of objective target blur, participants adjusted samples to be blurrier when matching them to scrambled targets than when matching them meaningful targets. (C) Results of Experiment 1B showing a replication of Experiment 1A and that the main effect of meaningfulness disappears when targets and samples are identical. See the online article for the color version of this figure.

(~75%), participants only moved the mousewheel in one direction. On the remaining trials, participants moved it back and forth, presumably to find as close a perceptual match as possible. Making adjustments in only one direction was quicker than making back-and-forth adjustments as evidenced by a negative relationship between the rate of single-direction turns:

$$|num_{up} - num_{down}| / (num_{up} + num_{-down}),$$

and RTs, $b = -4.28$, $t = -19.88$, $p < .0001$.

Making finer adjustments did yield better performance (smaller absolute error), $t = .07$, $t = 6.92$, $p < .0001$ suggesting that responses incorporating back and forth adjustments were more diligent. Including the wheel-turn measure above as a covariate made no difference to the effect of meaningfulness, $b = -.11$, $t = -5.36$, $p = .0001$. In fact, the difference in perceived blur between meaningful and scrambled words remained significant even after removing the 75% of trials on which participants only made adjustments in one direction, $b = -.09$, $t = -5.54$, $p = .0002$.

Experiment 1B

As further evidence that responses in Experiment 1A are tracking the level of perceived sharpness, Experiment 1B tested whether the results could be explained by the so-called "El Greco Fallacy" (Firestone & Scholl, 2014). To illustrate the potential problem, consider Levin and Banaji's (2006) examination of whether African American faces actually appear to be darker than Caucasian faces of ostensibly equal complexion. The authors used a perceptual matching task very similar to the one used here. When adjusting an African American sample face to a Caucasian target, they made it too light, as would be expected if they saw the African American sample face as being darker than the Caucasian target face when the two were objectively matched. But participants showed a similar pattern when *the two faces were identical*. Firestone and Scholl (2014, 2015) argued that this was clear evidence of a response bias because if people's knowledge truly affected their perception of lightness of African American and Caucasian faces, then when the target and sample were the same face, both stimuli should be equally affected by this knowledge and the effect of race (in Levin & Banaji, 2006) or, by analogy, of meaningfulness (in the present studies) should disappear. Experiment 1B sought to check whether making the sample and target identical reduced or eliminated the observed difference between the meaningless and scrambled conditions.

Results

Participants again tended to underestimate the amount of blur, $b = .84$, 95% CI = [.69, .99]. As shown in Figure 2B, there was a highly significant trial-type (same vs. different) by meaningfulness interaction, $b = .08$, 95% CI = [.05, .12], $t = 5.21$, $p < .0001$. There was an effect of meaningfulness on *different* trials, replicating Experiment 1A, $b = -.08$, $t = -5.21$, 95% CI = [−.11, −.05], $p = .0001$. With the target and sample the *same*, there was now no main effect of meaningfulness, $b = .004$, 95% CI = [−.03, .03], $t = .27$, $p = .79$.

A direct comparison of Experiments 1A and 1B showed that the effect of meaningfulness on *different* trials was not statistically different between the two studies, $b = .04$, 95% CI = [−.01, .09], $t = 1.48$, $p = .15$. Contrasting the *different* trials of Experiment 1A with the *same* trials of Experiment 1B revealed a highly significant interaction, $b = .12$, 95% CI = [.07, .16], $t = 4.70$, $p < .0001$.

Discussion of Experiments 1A and 1B

Participants viewed two letter strings, a constant *target* and an adjustable *sample* with the goal of adjusting the blur of the sample to match that of the target. Participants' adjustments closely paralleled the objective level of blur, indicating their comfort with the task. I predicted that meaningful words would appear to be sharper than scrambled letter strings. This means that at the same objective level of blur, the meaningful word would appear too sharp and would need to be made blurrier to match, leading to an overall greater blur on scrambled-word target trials than meaningful-word target trials (Figure 2A and 2B).

To further rule out the possibility that participants were just making the sample sharper or blurrier depending on how well they could read the target, I conducted Experiment 1B which was identical to 1A except including trials in which the target and sample were identical letter strings. If the difference in blur between meaningful and scrambled words reflected a response bias, it should have also been obtained on the identical trials, yet it disappeared entirely (Figure 2B).

A critic may contend that the present results should not be taken as evidence that participants saw meaningful words as sharper than meaningless words and that perceptual matching behavior could *still* be ascribed to biased responding based on how well people can make out the letters in the target string. I will come back to this point in the General Discussion, but to foreshadow: consider that such a response bias account could *also* be invoked to explain how a person being fit for eyeglasses goes about deciding whether they see something sharper with prescription A or with prescription B. Perhaps people are "just" responding that A is sharper when they think that option A helps them read the letters before them. But if one believes that eye-exams tend to measure *something* perceptual, then one should be ready to accept that the present experiment is as well.

Experiment 2

Experiment 1 showed that meaningful letter strings appeared sharper than visually similar but meaningless letter strings. The next set of studies examined what happens when letter strings *actually* become blurrier or sharper right before the observer's eyes. Answering this question sheds light on the mechanisms underlying the observed increase in perceived sharpness for meaningful words in Experiment 1 and in understanding the functional consequences accompanying these changes in appearance.

These experiments were inspired by predictive-coding frameworks of perception, according to which perceptual processes draw on whatever knowledge can lower overall prediction error (Lupyan, 2015; Lupyan & Clark, 2015; O'Callaghan et al., 2016). Knowledge, on such frameworks, is a source of top-down priors within which bottom-up incoming input can be processed more effectively. One consequence of such top-down support may be enhanced detection of a change that can be better predicted, such as the sharpening of a blurry meaningful visual stimulus. This prediction contrasts with an account on which perception of meaningful words is mediated by higher-level lexical representations (i.e., it is the lexical representation that we 'see' when viewing a meaningful word). This account predicts *lower* performance for meaningful words because meaningful words displayed at slightly different levels of blur would be expected to activate the very same lexical representation.

Method

Participants. Twenty-four undergraduate students participated in exchange for course credit. The final sample was $n = 23$ after removing one participant for chance-level performance. The sample size for this study and the studies that follow were based on pilot testing that helped gauge the typical variability of the effect, though no formal procedure for determining sample size was used.

Materials. The stimuli were 15 short sentences, for example, "The wine was too sweet" obtained from Johnson and Hamm (2000), blurred in the same manner as the words in Experiment 1. Pseudowords were created by scrambling letter order within each sentence, maintaining the number of "words" and letters per word the same, for example, "The wine was too sweet"→"Aen eoet ews wio stwht." Each meaningful sentence had 4 pseudoword sentence counterparts, one of which was reserved for the practice session. See supplementary materials for a full listing of the materials. The sentences subtended approximately 0.9° (height) × 10° to 15° (width) of visual angle and were rendered in the Arial font.

Procedure. The basic procedure is shown in Figure 3. Each F3 trial began with a 1-s presentation of a black-outlined rectangle in which appeared a blurred sentence. After a variable delay (1 s to 1.9 s), the letters further blurred or sharpened (with equal probability) over the course of 830 ms. These changes were subtle (see the supplementary videos for demonstrations). With the letters still present on the screen, participants were asked if they had blurred, sharpened, or if there was no change. Because the changes were difficult to perceive, it was important to allow participants to respond "no-change" if they actually failed to see a change, though this was always an error. This design was used to help rule out the possibility that participants were biased to judge meaningful or pseudowords as changing or not-changing. Note that any bias participants may have had to respond "blur" or "sharpen" would cancel itself out: a bias to respond "blur" would improve accuracy on "blur" trials to the same extent as it would decrease accuracy on "sharpen" trials. After making their response, participants were, on some trials, asked to make a legibility judgment on a 7-point Likert scale: "How clearly do you see the letters/words above?" (1 = *can't make them out at all*; 7 = *can easily identify every letter*).

The experiment began with 30 practice trials containing only random letter strings (one of the four variants of the scrambled strings used in the main experiment) undergoing more obvious blurring/sharpening than the main experiment session to reduce any ambiguity as to what constituted blurring and sharpening. For the first 10 practice trials participants received auditory accuracy feedback. Following the practice session, participants were informed that for the subsequent trials the letters would sometimes spell out meaningful sentences (Experiments 2–3) or words (Experiments 4–5) and that the changes would be more difficult to detect. Participants completed 120 trials with each sentence appearing 4 times. A walkthrough of the task is available at osf.io/h6mq.

Analytic approach for experiments 2–5. The data were analyzed using mixed-effects linear models (R version 3.2.2: lme4 version 1.1.10). Accuracy was modeled using logistic regression (glmer, family = binomial) on the raw data. I used mixed effects logistic regression in lieu of signal detection theory (SDT) because the procedure does not allow participants to make false alarms,[2] Fn2 and so nothing would be gained by a signal detection theory

[2] It is possible, of course, to arbitrarily treat one trial-type as a False Alarm and another as a Hit and apply a correcting factor (the 2AFC version of SDT, Green & Swets, 1966), but this analysis likewise offers no clear benefit over the mixed-effects logistic regression approach.

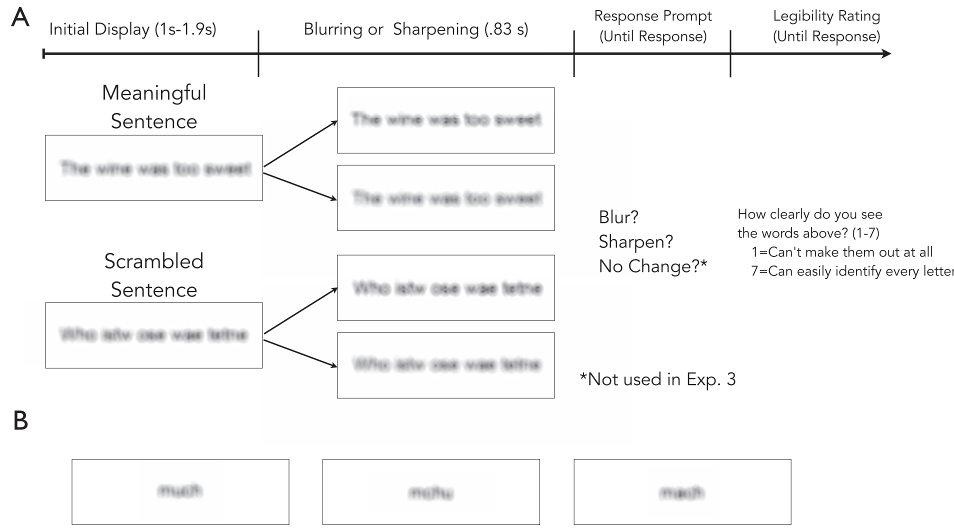


Figure 3. Design of Experiments 2 through 5. Participants had to detect whether letter strings became blurrier or sharper. (A) Experiments 2 and 3 contrasted meaningful sentences (e.g., "The wine was too sweet") to scrambled letter strings (e.g., "Who istw ose wae tetne") containing the same letters. (B) Experiment 4 contrasted meaningful words (e.g., "much") to scrambled words (e.g., "mchu") containing the same letters; Experiment 5 contrasted relatively high frequency words (e.g., "much") to lower frequency words containing the same first and last letters (e.g., "mach"). In all studies, the sentences/words remained until a response was made.

analysis. What would be lost in using standard SDT is flexible incorporation of multiple sources of error that mixed-effects modeling allows. In all the analyses that follow I report descriptive statistics for the different error types, leaving no ambiguity as to the kinds of error people make in the various conditions.

Continuous dependent variables (RTs and legibility ratings) were modeled using linear regression applied to the raw (unaggregated) data. The remaining details were the same as in Experiment 1 except that trials with unusually long responses (>5 s from the end of the blurring or sharpening) were excluded (~2.5% of trials). Analysis of response times included correct trials that were within 3.5 SDs of each participant's mean RTs.

Results

F4 The basic results are shown in Figure 4A. Overall accuracy was 67.5% to detect blurring (M_{RT} = 995 ms) and 75.3% to detect sharpening (M_{RT-} = 892 ms). Of the errors, 88% were failures to detect the change (i.e., responding 'no-change'). Only rarely did participants mistake sharpening for blurring (5.6% of the errors) or vice versa (6.4% of the errors).

Performance on the meaningful sentences (M = 75.9%) was considerably higher than on scrambled sentences (M = 69.2%), b = .50, 95% CI = [.25, .70], z = 3.91, p = .0001. This meaningfulness advantage interacted significantly with change-type, b = .69, 95% CI = [.30, 1.08], z = 3.48, p = .0005. Meaningfulness significantly increased detection of sharpening from 71.4% to 81.6%, b = 1.07, 95% CI = [0.50, 1.65], z = 3.65, p = .0002; detection of blurring was also improved (from 67.1% to 70.0%) but not significantly so (b = .24, 95% CI = [−0.05, 0.53], z = 1.6, p = .10). There were no effects of meaningfulness on RTs (t = .19) and no interactions between change-type and meaningfulness for RTs (t = 0.0).

The objectively higher accuracy for the meaningful sentences was mirrored by people's legibility ratings. Participants reported that they could see the individual letters in the meaningful sentences (M = 4.6) more than the very same letters in the scrambled sentences (M = 2.2), b = 2.36, 95% CI = [2.03, 2.70], t = 13.9, $p < .0001$. As with accuracy, the effect of meaningfulness interacted with the change type, b = .82, 95% CI [.49, 1.15], t = 4.87, $p < .0001$: meaningfulness yielded an increase in legibility from 3.0 to 5.7 when the sentence sharpened, b = 2.80, 95% CI = [2.35, 3.25], t = 12.23, $p < .00001$, and from 1.5 to 3.3 when they blurred, b = 1.94, 95% CI = [1.51, 2.39], t = 8.61, $p < .00001$.

Experiment 3

Most of the errors in Experiment 2 were failures to see blurring, rather than confusing blurring with sharpening, with participants being more likely to respond 'no-change' on blur trials when the stimuli were meaningless strings compared with meaningful sentences. But perhaps participants simply had a higher threshold for responding 'no-change' when viewing meaningless strings? Such a greater bias to respond 'no-change' (for whatever reason) would then lead to lower accuracy. If the apparent effect of meaningfulness on accuracy in Experiment 2 were just a matter of a differences in the threshold to respond no-change, then the effect of meaningfulness should disappear when the no-change option was not available. If, however, people were genuinely better at seeing changes to blur in meaningful sentences, the accuracy advantage should persist (though overall accuracy should increase because there are fewer opportunities for errors).

Method

Participants. Twenty-five undergraduate students participated in exchange for course credit.

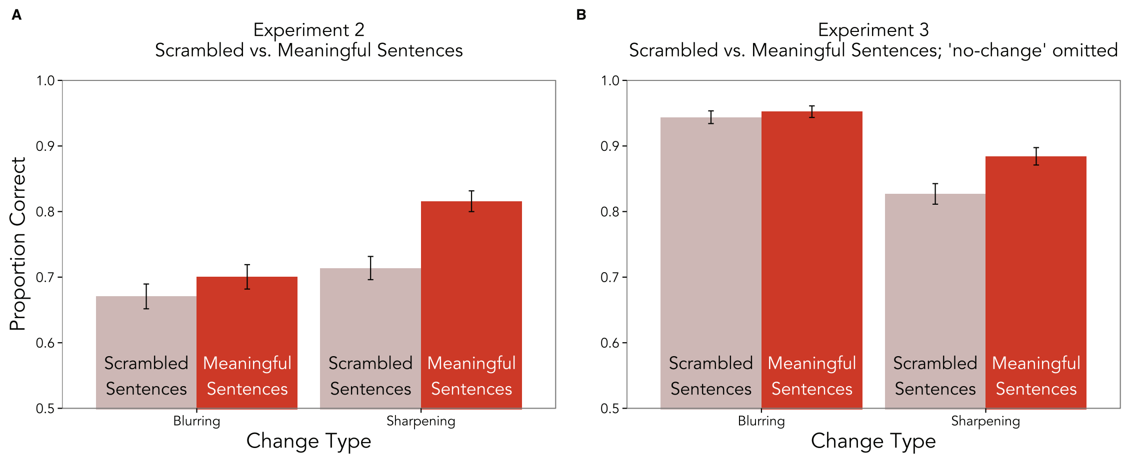


Figure 4. Results of Experiments 2 through 3. (A) Participants were more accurate in detecting visual changes, particularly sharpening, when they occurred in meaningful sentences compared with meaningless letter strings. (B) A similar pattern was obtained when the option to respond no-change was omitted. See the online article for the color version of this figure.

Materials and procedure. The materials were identical to Experiment 2. The only change to the procedure was omitting the "no-change" response option.

Results

The basic results are shown in Figure 4B. Removing the no-change response option meant that participants had to guess whether the stimulus blurred or sharpened even if they failed to register any change. These guesses were of course sometimes successful, as evident by the higher mean accuracy in the present experiment compared with Experiment 2. Overall accuracy on the blurring trials was 95.8% (M_{RT-} = 854 ms); on sharpening trials it was 85.6% (M_{RT-} = 861 ms). These accuracy rates translate to a 14% likelihood of misperceiving sharpening as blurring, and 5.2% likelihood of misperceiving blurring as sharpening.

If the accuracy advantage in Experiment 2 were due to participants being more likely to respond 'no-change' when they had a hard time reading the letter string, then removing the option to respond 'no-change' should also remove the meaningfulness accuracy advantage. And yet, performance on the meaningful sentences (M = 91.8%) remained significantly higher than on scrambled sentences (M = 88.6%), b = .52, 95% CI = [.17, .87], z = 2.94, p = .003. This meaningfulness advantage again interacted significantly with change-type, b = .69, 95% CI = [.03, 1.22], z = 2.05, p = .04. As in Experiment 2, meaningfulness significantly increased detection of sharpening from 82.7% to 88.4%, b = 1.03, 95% CI = [0.55, 1.49], z = 4.32, $p < .0001$; detection of blurring was also improved (from 94.4.1% to 95.2%) but not significantly so (b = .24, 95% CI = [−.31, 0.65], $z < 1$). There were no effects of meaningfulness on RTs ($t < 1$) and no interactions between change-type and meaningfulness for RTs ($t < 1$).

As in Experiment 2, the objectively higher accuracy for the meaningful sentences were mirrored by people's legibility ratings. Participants reported that they could see the individual letters in the meaningful sentences (M = 4.75) more than the very same letters in the scrambled sentences (M = 2.39), b = 2.37, 95% CI = [1.86, 2.88], t = 9.1, $p < .0001$. As with accuracy, the effect of meaningfulness interacted with the change type, b = .70, 95% CI [.37, 1.03], t = 4.17, $p < .0001$: meaningfulness yielded an increase in legibility from 3.24 to 5.91 when the sentence sharpened, b = 2.78, 95% CI = [2.30, 3.26], t = 11.35, $p < .00001$, and from 1.63 to 3.45 when they blurred, b = 2.07, 95% CI = [1.44, 2.71], t = 6.43, $p < .00001$.

Experiment 4

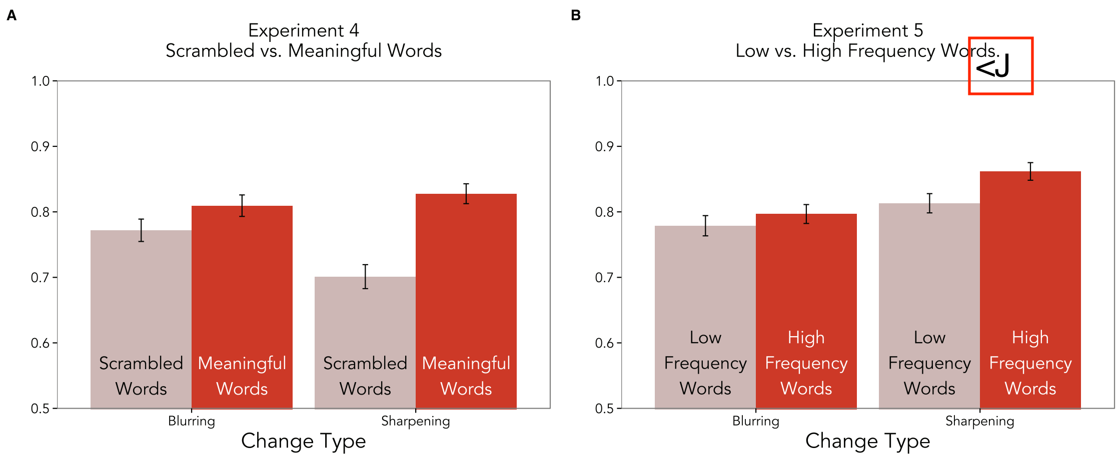
Could the observed meaningfulness advantage arise from differences in how people visually examined meaningful sentences compared with the pseudowords? Experiment 4 was identical to Experiment 2, but contrasted single meaningful words, for example, "much," to pseudowords containing the same letters, for example, "mcuh"—the same strings used in Experiment 1. Viewing these single words did not require any eye movements or, arguably, even covert attentional shifts. Virtually all participants reported that they performed the task by focusing on one or two of the letters and monitoring for changes.

Method

Participants. Twenty-four undergraduate students participated in exchange for course credit. The final sample was n = 22 after removing two participants for chance-level performance.

Materials. The stimuli were identical to those used in Experiment 1: 15 four-letter words and 15 matched pseudowords containing the same letters in a different order. Unlike Experiments 2 and 3, only one scrambled letter string was used for each word ensuring that each unique string was seen an equal number of times by each participant. The words subtended approximately 0.9° (height) × 2.5° (width) of visual angle.

Procedure. The procedure was identical to Experiment 2.



ONLINE COLOR

Figure 5. Results of Experiments 4 through 5. (A) Participants were more accurate in detecting visual changes, particularly sharpening, when they occurred in meaningful words compared with meaningless letter strings (Experiment 4), (B) and in words of relatively high compared with low frequency (Experiment 5). See the online article for the color version of this figure.

Results

F5 Basic results are shown in Figure 5A. Overall accuracy to detect blurring was 79.1% (M_{RT-} = 1015 ms), to detect sharpening, 76.5% (M_{RT-} = 1104 ms). As in Experiment 2, the majority of the errors (75%) tended to be failures to detect the change rather than misperceiving sharpening for blurring, or vice versa. Meaningfulness significantly improved overall performance, from M = 73.7% to M = 81.9%, b = .62, 95% CI = [.34, .91], z = 4.255, p < .0001. The difference in accuracy remained significant when no-change responses were excluded, b = .70, 95% CI = [.33, 1.07], z = 3.71, p = .0002.

As in Experiments 2 and 3, the accuracy advantage on meaningful trials interacted with the type of change, b = .64, 95% CI = [.21, 1.06], z = 2.95, p = .003. People were more accurate in detecting sharpening of meaningful compared with scrambled words (from 70.1% to 82.8%), b = .99, 95% CI = [.61, 1.37], z = 5.13, p < .0001. Detecting blurring increased from 77.2% to 80.9%, a marginal difference, b = .34, 95% CI = [−.04, .71], z = 1.77, p = .08. An analysis of RTs showed a nonsignificant reduction of RTs for meaningful (M = 1040 ms) compared with scrambled (M = 1078 ms) words, b = −49, t = −1.39, p = .17, and a marginal interaction between meaningfulness and change-type in the same direction as the accuracy results, b = −119, t = −1.77, p = .08. Correctly detecting sharpening in meaningful words was accomplished somewhat more quickly (M = 1042 ms.) than in scrambled words (M = 1176 ms.), b = −115, 95% CI = [−228, −1], t = −1.98, p = .06, with no corresponding change to RTs for detecting blurring, t < 1.

As in Experiments 2 and 3, the effect of meaningfulness on accuracy was mirrored by people's judgments of the legibility of the individual letters. People reported that they could make out the letters of real words (M = 4.36) more than of scrambled words (M = 2.63), b = 1.67, 90% CI = [1.38, 1.95], t = 11.36, p < .0001. The increase in legibility was larger when the word became sharper (an increase from 3.24 to 5.25) than when it became blurrier (from 2.01 to 3.46), b = .64, t = 3.92, p < .0001, a highly significant interaction, b = .63, 95% CI = [.32, .96], t = 3.98, p = .0001.

Experiment 5

Experiments 2 through 4 showed that people are better at detecting perceptual changes in meaningful sentences and words than in visually similar pseudowords. However, in Experiments 2 through 4 the pseudowords differed from meaningful words not only in meaning but also in being unpronounceable and containing orthographically rare or illegal letter combinations (e.g., "mcuh," "stte"). Experiment 5 was identical to Experiment 4, but contrasted high-frequency words (e.g., "seem") with visually similar lower-frequency words (e.g., "seam"). This allowed for investigating whether the previously reported change-detection advantage for meaningful words stemmed strictly from differences in visual familiarity with certain letter combinations (e.g., it is conceivable that for English-speakers, the "tt" in the pseudoword "stte" is an unusual and hence difficult-to-process visual stimulus) or also depended on certain *semantic* characteristics of the words.

Method

Participants. Thirty undergraduate students participated in exchange for course credit. Four participants were removed for chance-level performance and 1 for failure to complete the study, leaving a sample size of n = 25. Twenty additional participants were recruited from Amazon Mechanical Turk to provide imageability ratings for each word. AQ: 3

Materials and procedure. In Experiment 5, 32 meaningful words were used; aside from the stimuli, the experimental procedure was identical to Experiment 4. Half of these were relatively high-frequency four-letter words, for example, "much." Each word was paired with a lower-frequency counterpart having at least the same initial and final letters, for example, seem/seam, worn/wren, much/mach (see supplementary materials or osf.io/h6mqf for a full listing). The mean frequencies of the high and low frequency set were, respectively 478 and 3.18 per million, based on SUBTLEX-US (Brysbaert & New, 2009). I next compared the two sets of words on their bigram frequencies and orthographic neighborhoods (Balota et al., 2007). The high- and low- frequency sets did not differ in mean summed letter bigram frequencies (log-transformed), $p > .8$ showing that there was nothing unusual about the letter combinations used in the low-frequency words. The words in the higher-frequency set did have slightly more orthographic neighbors, $M_{\text{high}} = 12.5$, $M_{\text{low}} = 10.0$, $t = 2.4$, $p = .02$). The main criteria for creating the word pairs was ensuring a large frequency discrepancy between the words and that the words within each pair were visually similar (e.g., "dead"/"deed," "born"/"boon"). As a consequence of these selection criterion, many of the low-frequency words were missing from published norms such as the MRC Psycholinguistic database (Wilson, 1988). T1 Table 1 shows Pearson correlations between word-frequency (SUBTLEX-US), meaningfulness (Colorado norms from the MRC database), concreteness (also from the MRC), and imageability (from Cortese & Fugett, 2004) and a new imageability measure I collected by having 20 participants recruited from Amazon Mechanical Turk rate each of the 64 words, displayed one at a time, on a 1- to 7-point imageability scale using the same procedure as Cortese and Fugett (2004).

The negative relationship between word-frequency and concreteness (a measure available primarily for the high-frequency words) is due to the presence of high-frequency/low concreteness words like "much," "been" and "from." The newly collected imageability ratings are very highly correlated with previously published norms. The lack of a relationship between imageability and word-frequency hides an interaction. Within low-frequency words, imageability was positively correlated with frequency ($r = .38$, $p = .03$). Rarer words (e.g., "wean") tended to be less imageable than slightly more frequent words (e.g., "lung"). Within the high-frequency words, the imageability-frequency correlation was marginally negative, ($r = -.30$, $p = .09$) for the same reason as the negative relationship between concreteness and frequency: the highest frequency words such as "much" and "from" are not very imageable.

Results

Basic results are shown in Figure 5B. Overall accuracy to detect blurring was 78.2% ($M_{RT-} = 1237$ ms) and to detect sharpening, 85.0% ($M_{RT-} = 1295$ ms); 75% of all errors were failures to detect the change. Compared with the meaningfulness manipulations of Experiments 2 through 4, manipulating word frequency led to a smaller, but still reliable effect on detection accuracy: accuracy was 82.8% for high-frequency words and 79.6% for their low-frequency counterparts, $b = .39$, 95%CI = [.12, .65], $z = 2.85$, $p = .004$. Repeating this analysis using a continuous measure of word-frequency (Log SUBTLEX-US frequency) likewise showed an advantage for the high-frequency words, $b = .13$, 95% CI = [.04, .23], $t = 2.80$, $p = .005$.

The high-frequency accuracy advantage remained significant when no-change responses were omitted, $b = .37$, 95% CI = [.03, .72], $t = 2.10$, $p = .04$. The high-frequency advantage interacted marginally with change-type, $b = .40$, 95% CI = [−.02, .82], $z = 1.86$, $p = .06$ in the same direction as Experiments 2 through 4. Higher word frequency aided detection of sharpening more (from 81.3% to 86.1%) compared with detection of blurring (from 77.8% to 79.7%).

Letter legibility judgments were higher for high-frequency words ($M = 3.73$) compared with their low-frequency counterparts ($M = 3.44$), $b = .33$, 95% CI = [.22, .44], $t = 5.9$, $p < .0001$, despite the words sharing many of the same visual properties, and the same initial and final letters. Legibility judgments did not interact significantly with change-type, but the effect of frequency was numerically greater after sharpening than after blurring.

An analysis of RTs showed that participants were numerically faster to respond to high-frequency ($M = 1256$ ms) relative to low-frequency words ($M = 1279$ ms), but this difference was not significant ($t < 1$), and did not significantly interact with change-type ($t < 1$). RTs were also not significantly affected by the meaningfulness and imageability measures (see below).

Is Change Detection Predicted by Characteristics Beyond Frequency?

By definition, more frequent words are those that are likely to have been seen more frequently by our participants. Is the described advantage for high frequency words attributable solely to differences in accumulated experiences with the word-forms, or does it also depend on factors related to the word's meaning? Before reporting these analysis, it is important to note that words are not randomly assigned to frequency. Differences in frequency

Table 1
Lexical Characteristics of the Words Used in Experiment 5 Showing Pearson Correlations (Top-Entry) and Corresponding p Values (Bottom Entry)

Variable	Log freq.	Meaningfulness	Concreteness	Imageability
Meaningfulness (29/32 high; 5/32 low)	−.031 .861			
Concreteness (30/32 high; 9/32 low)	−.475 .002	.698 <.0005		
Imageability (Cortese & Fugett) (30/32 high; 24/32 low)	.029 .837	.759 <.0005	.832 <.0005	
New Imageability (All)	−.046 .721	.742 <.0005	.923 <.0005	.921 <.0005

Note. All measures except the newly collected imageability ratings were missing for some of the words as indicated in the first column (e.g., only 5 of the 32 low-frequency words had available meaningfulness norms).

will therefore always relate to differences in some aspects of meaning: there is a good reason why "back" and "look" are more frequent words than "bunk" and "lurk."

To examine the contributions of semantic factors (meaningfulness, concreteness, and imageability) to change detection, I conducted a series of mixed effect logistic regression models with frequency as a continuous predictor and examining further contributions of meaningfulness, concreteness, and imageability. Because most low-frequency words did not have meaningfulness and concreteness ratings (see Table 1), these two analyses was limited to just the 32 high-frequency words.

Controlling for word frequency, performance was positively predicted by both meaningfulness, $b = 0.004$, 95% CI = [0.0004, 0.007], $t = 2.22$, $p = .03$, and concreteness, $b = .03$, 95% CI = [.01, .04], $t = 3.39$, $p = .0007$. An analysis of all items using the newly collected measure of imageability revealed that imageability was not a significant predictor of accuracy overall, $t < 1$, but entered into a highly significant interaction with word frequency, $b = .13$, 95% CI = [.06, .21], $t = 3.64$, $p = .0003$. Imageability was a strong predictor of performance for the higher-frequency words, $b = .24$, 95% CI = [.10, .38], $t = 3.45$, $p = .0006$, but not for the lower-frequency words, $b = -.11$, 95% CI = [−.31, .09], $z = 1.2$, $p = .26$. For example, participants were 61% correct for "want," a very high frequency but low-imageability word, and at 84% correct for "roof," a considerably lower-frequency but more imageable word. At least when dealing with words having appreciable frequencies, it is imageability—a semantic attribute—rather than simply word frequency that is the better predictor of visual change detection.

General Discussion

To what extent is what we see influenced by what we know? Despite numerous demonstrations of knowledge affecting visual recognition, claims that more basic perception is likewise influenced by knowledge have remained contentious because perception has often been measured in ways removed from what some consider to be perception proper (Firestone & Scholl, 2015; Pylyshyn, 1999). The present work provides a compelling demonstration of knowledge affecting an aspect of visual appearance (Experiment 1) and the ability to detect changes happening right before people's eyes (Experiments 2–5).

In Experiment 1, people were asked to adjust the sharpness of meaningless pseudowords and meaningful words to make them look identical, the pseudowords were adjusted to be sharper than meaningful words. This means that when shown a meaningless letter string and a meaningful word at the same level of actual blur, meaningful words look sharper than pseudowords and the pseudowords needed to be made sharper to match.

If recognizing a letter string makes it appear sharper what would happen if it *actually* became sharper? There are several possibilities. First, if the greater perceived sharpness of meaningful words is due to the perceptual experience being mediated by a higher level representation—that is, we see "much" as sharper than "mcuh" because only the former experience is mediated by a stored lexical representation—then change detection should be *more* difficult for meaningful words because a blurry "much" and a slightly sharper "much" would be mediated by the same higher level representation.[3] A second possibility is that the greater expectations generated by partially recognizing a word can produce changes to sensitivity (Bar et al., 2006; Lupyan & Clark, 2015; O'Callaghan et al., 2016): meaningful words generate stronger predictions allowing for more effective comparisons with a changing input. On this account, change-detection performance should be better for meaningful words.

The results of Experiments 2 through 5 showed that people were objectively better in detecting low-level visual changes (particularly sharpening) in meaningful words and sentences compared with the same changes occurring in meaningless or less meaningful words.[4] We can contrast this result with an account heretofore not mentioned: that the better ability to recognize meaningful words leads us to better predict what the word is which in turn leads us to *see what we expect to see* (see Balcetis & Dunning, 2006; Fodor, 1984; Siegel, 2012 for related discussion). Although in cases of greatly reduced or nonexistent perceptual input, we might see what we expect to see (Jordan, Sheen, Abedipour, & Paterson, 2015; Lupyan, 2013 for discussion), the results do not support this account for the present case. On the *see what you expect to see* account, participants should have been more likely to respond to real words as becoming sharper regardless of whether they did or not (a type of response criterion shift). This did not occur.

To get a stronger intuition of the difference between the expectations-aiding-perception account and perceive-what-you-expect account, imagine expecting to taste milk, but taking a sip of orange juice instead. The resultant experience is *not* of tasting milk. Rather, it is of tasting orange juice *within* a prior expectation of milk—a phenomenologically distinct experience.

Alternative Explanations of the Observed Findings

Critics have argued that claimed effects of knowledge on perception can be variously ascribed to response bias, memory, recognition, or attention (Firestone & Scholl, 2015; Pylyshyn, 1999; Raftopoulos, 2015; see Lupyan, in press, for an appraisal of these critiques). The next section addresses how such alternatives might apply to the present results.

Can the Effects of Meaningfulness Be Accounted by Response Bias?

Distinguishing between sensitivity and response bias is important and has played a key role in advancing the study of perception in the form of signal detection theory (SDT; Green & Swets, 1966; see Witt, Taylor, Sugovic, & Wixted, 2015 for recent discussion).

[3] As pointed out by a reviewer, it is possible to make the opposite prediction by positing that change detection is subserved by monitoring the activation changes of an abstract lexical representation which exist for the meaningful words, but not the pseudowords. This account, in addition to having trouble dealing with Experiment 5 (why would it be easier to monitor the lexical representation of a more frequent or more imageable word?), is a logical possibility, but conflicts with the *raison d'etre* of lexical representations: to provide a word representation that is insensitive to perceptual changes.

[4] These results are compatible with findings that domain expertise decreases change-blindness (Werner & Thies, 2000). But improved performance on these more cognitively demanding change-detection tasks may derive from experts strategically attending to different parts of the scene compared with domain novices.

tapraid5/zfn-xhp/zfn-xhp/zfn00217/zfn3548d17z | xppws | S=1 | 12/22/16 | 11:21 | Art: 2016-1106

The results of Experiment 1 can be categorized as a kind of bias: a bias to perceive meaningful words as sharper than meaningless letter strings, but, I maintain, it is a *perceptual* bias, much as the Müller-Lyer illusion reflects a perceptual bias (Witt et al., 2015).

I argue that neither Experiment 1 nor the results of the change-detection task of Experiments 2 through 5 can be explained as a simple *response bias* in the classic SDT sense on which the differences in blur adjustments in Experiments 1a and 1b and change detection in Experiments 2 through 5 reflect solely changes in the likelihood of making one response versus another without a concomitant change to some perceptual process. Although the results of Experiment 1A left open the possibility that participants may have been biased to make the sample sharper anytime they perceived a legible target word (a form of the "El Greco Fallacy," Firestone & Scholl, 2014), this interpretation was ruled out by Experiment 1B. In Experiments 2 through 5, it is conceivable that participants were biased to respond "sharpen" whenever they viewed meaningful words/sentences insofar as these letter strings were easier to read. This bias would translate to a higher accuracy on "sharpen" trials. But such a bias would also translate to lower accuracy on "blur" trials whereas performance on "blur" trials was equivalent or better for meaningful compared with meaningless trials. Finally, it is conceivable that participants had a stronger bias to respond "no change" when viewing meaningless sentences/words; however, Experiment 3 showed that the meaningfulness advantage was obtained when "no-change" response was omitted as a response option. Moreover, in Experiments 4 and 5, the meaningfulness advantage remained significant after excluding all no-change responses. These results further rule out the possibility
Fn5 that the change in performance can be ascribed to a response bias.[5]

Does Meaningfulness Affect Recognition, Memory, or Perception?

Might the results of Experiments 2 through 5 reflect effects of knowledge on recognition or memory rather than on perception? Such an argument implies that perception can be cleanly separated from memory and recognition, an assumption contradicted by considerable evidence (see Lupyan, in press). For example, (Kahan & Enns, 2014) present evidence that memory (analogous to the notion of meaningfulness as discussed here) is involved in the earliest stages of object processing, even prior to the assignment of edges, a result that clearly undermines the idea that memory is downstream of perception (Firestone & Scholl, 2015).

Nevertheless, let us consider what a memory-based account would look like. Perhaps participants were more likely to detect the visual change if they could not read the string before the change, but could read it after—a situation more likely to arise with meaningful words. Several participants did report using such a strategy (they showed all the same effects of meaningfulness, but had somewhat lower overall accuracy, suggesting that if nothing else, this strategy was suboptimal). More than 90% of the participants reported using a different strategy: monitoring for moving contours or a change in contrast. I invite readers to try the task for themselves (see the movie at osf.io/h6mqf) to gain a better appreciation for the usefulness of such a monitoring strategy. If we were to assume that such monitoring cannot be done without recognizing the image and comparing one's recognition estimates before and after the change, then detecting even such simple visual changes is necessarily influenced by recognition and memory, ruling out the argument that these processes are in any sense postperceptual (i.e., downstream of perception). To paraphrase Gleitman and Papafragou's reflection on an analogous debate—the relationship between language and cognition—why should it be so hard to pry perception and knowledge apart if they are so separate? (Gleitman & Papafragou, 2005, p. 653). It may well turn out that what we call recognition and memory are the mechanisms *by which* knowledge augments perception just as attention is one of the mechanisms by which knowledge flexibly augments perceptual processing.

Do Effects of Meaningfulness Reflect Perceptual Familiarity or Knowledge?

Throughout this paper I have repeatedly used the terms "meaningful" and "meaningless" to refer to letter strings that not only differ in meaningfulness (e.g., "mcuh" vs. "much") but also differ in perceptual familiarity. People have seen the visual object that is "much" more often than the visual object that is the string "mcuh." To what extent can the results be explained by such differences in visual familiarity? First, it is by no means obvious that a visual stimulus with which an observer has had more experience should appear sharper than one with which the observer has had less experience. Admitting this to be the case is *prima facie* evidence that what something looks like is partly a function of past experiences rather than simply in-the-moment perceptual processing. Second, Experiment 5 shows that visual experience alone (as measured by bigram and word frequency) cannot explain the change-detection advantage that some letter strings have over others. Bigram frequency was completely nonpredictive of performance. Word frequency is predictive of performance, but semantic variables such as concreteness, meaningfulness, and imageability predict performance when word frequency is partialed out. These results are sufficient to rule out the possibility that the perceptual advantage for higher-frequency words is driven solely by visual familiarity, but further work is clearly needed to understand why these semantic factors contribute to change-detection performance.

Conclusion

Taken together, the results show that meaningful letter strings look sharper than unfamiliar/meaningless ones, and that word knowledge improves accuracy in seeing simple perceptual changes happening right before one's eyes. Much work remains to elucidate the precise mechanisms that underlie these effects and to relate them in more detail to models of predictive coding that inspired the present work. Perception is not encapsulated from knowledge, but enriched by it.

[5] Ruling out biases is a laudable goal, but when invoked carelessly, it risks becoming a distraction. In discussing the critiques of the New Look Movement, Erdelyi (1974) argued that in their calls for distinguishing between response bias and sensitivity, these critiques risked removing everything of substance from the study of perception: "From the logical standpoint, it should be clear that no internal neural event imaginable could exist outside the all-enveloping sweep of the term "response" . . . Would it be reasonable to suggest [that] because the firing of a retinal rod is unarguably a "response" to some stimulus, that the process should be removed "from the field of perception" and placed "back with response variables?" (Erdelyi, 1974, p. 7).

References

Balcetis, E., & Dunning, D. (2006). See what you want to see: Motivational influences on visual perception. *Journal of Personality and Social Psychology, 91,* 612–625. http://dx.doi.org/10.1037/0022-3514.91.4.612

Balota, D. A., Yap, M. J., & Cortese, M. J. (2006). Visual word recognition: The journey from features to meaning (a travel update). *Handbook of Psycholinguistics, 2,* 285–375.

Balota, D. A., Yap, M. J., Hutchison, K. A., Cortese, M. J., Kessler, B., Loftis, B., . . . Treiman, R. (2007). The English lexicon project. *Behavior Research Methods, 39,* 445–459. http://dx.doi.org/10.3758/BF03193014

Bar, M., Kassam, K. S., Ghuman, A. S., Boshyan, J., Schmidt, A. M., Dale, A. M., . . . Halgren, E. (2006). Top-down facilitation of visual recognition. *Proceedings of the National Academy of Sciences of the United States of America, 103,* 449–454. http://dx.doi.org/10.1073/pnas.0507062103

Barlow, H. B. (1997). The knowledge used in vision and where it comes from. *Philosophical Transactions of the Royal Society of London Series B, Biological Sciences, 352,* 1141–1147. http://dx.doi.org/10.1098/rstb.1997.0097

Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods, Instruments & Computers, 41,* 977–990. https://doi.org/10.3758/BRM.41.4.977

Burge, J., & Geisler, W. S. (2011). *Optimal image-based defocus estimates from individual natural images* (p. IMC2). Washington, DC: OSA. https://doi.org/10.1364/ISA.20~~11.IMC2~~

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences, 36,* 181–204. http://dx.doi.org/10.1017/S0140525X12000477

Cortese, M. J., & Fugett, A. (2004). Imageability ratings for 3,000 monosyllabic words. *Behavior Research Methods, Instruments, & Computers, 36,* 384–387. http://dx.doi.org/10.3758/BF03195585

Erdelyi, M. H. (1974). A new look at the new look: Perceptual defense and vigilance. *Psychological Review, 81,* 1–25. http://dx.doi.org/10.1037/h0035852

Firestone, C., & Scholl, B. J. (2014). "Top-down" effects where none should be found: The El Greco fallacy in perception research. *Psychological Science, 25,* 38–46. http://dx.doi.org/10.1177/0956797613485092

Firestone, C., & Scholl, B. J. (2015). Cognition does not affect perception: Evaluating the evidence for 'top-down' effects. *Behavioral and Brain Sciences*. Advance online publication. http://dx.doi.org/10.1017/S0140525X15000965

Fodor, J. A. (1984). Observation reconsidered. *Philosophy of Science, 51,* 23–43. http://dx.doi.org/10.1086/289162

Gleitman, L., & Papafragou, A. (2005). Language and thought. In K. Holyoak, & B. Morrison (Eds.), *Cambridge handbook of thinking and reasoning* (pp. 633–661). Cambridge, UK: Cambridge University Press.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York, NY: Wiley.

Gregory, R. L. (1980). Perceptions as hypotheses. *Philosophical Transactions of the Royal Society of London Series B, Biological Sciences, 290,* 181–197. http://dx.doi.org/10.1098/rstb.1980.0090

Held, R. T., Cooper, E. A., & Banks, M. S. (2012). Blur and disparity are complementary cues to depth. *Current Biology, 22,* 426–431. http://dx.doi.org/10.1016/j.cub.2012.01.033

Humphreys, G. W., Evett, L. J., & Quinlan, P. T. (1990). Orthographic processing in visual word identification. *Cognitive Psychology, 22,* 517–560. http://dx.doi.org/10.1016/0010-0285(90)90012-S

Johnson, B. W., & Hamm, J. P. (2000). High-density mapping in an N400 paradigm: Evidence for bilateral temporal lobe generators. *Clinical Neurophysiology: Official Journal of the International Federation of Clinical Neurophysiology, 111,* 532–545. http://dx.doi.org/10.1016/S1388-2457(99)00270-9

Johnston, J. C. (1978). A test of the Sophisticated Guessing Theory of word perception. *Cognitive Psychology, 10,* 123–153. http://dx.doi.org/10.1016/0010-0285(78)90011-7

Jordan, T. R., Dixon, J., McGowan, V. A., Kurtev, S., & Paterson, K. B. (2016a). Effects of spatial frequencies on word identification by fast and slow readers: Evidence from eye movements. *Frontiers in Psychology, 7,* 1433. http://dx.doi.org/10.3389/fpsyg.2016.01433

Jordan, T. R., Dixon, J., McGowan, V. A., Kurtev, S., & Paterson, K. B. (2016b). Fast and slow readers and the effectiveness of the spatial frequency content of text: Evidence from reading times and eye movements. *Journal of Experimental Psychology: Human Perception and Performance, 42,* 1066–1071. http://dx.doi.org/10.1037/xhp0000234

Jordan, T. R., Sheen, M., Abedipour, L., & Paterson, K. B. (2015). Seeing inscriptions on the Shroud of Turin: The role of psychological influences in the perception of writing. *PLoS ONE, 10,* e0136860. http://dx.doi.org/10.1371/journal.pone.0136860

Jordan, T. R., & Thomas, S. M. (2002). In search of perceptual influences of sentence context on word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28,* 34–45. http://dx.doi.org/10.1037/0278-7393.28.1.34

Jordan, T. R., Thomas, S. M., & Scott-Brown, K. C. (1999). The illusory-letters phenomenon: An illustration of graphemic restoration in visual word recognition. *Perception, 28,* 1413–1416. http://dx.doi.org/10.1068/p2919

Kahan, T. A. (2016). What dot-based masking effects can tell us about visual cognition: A selective review of masking effects at the whole-object and edge-based levels. In B. H. Ross (Ed.), *Psychology of learning and motivation* (Vol. 64, pp. 247–284). San Diego, CA: Academic Press. http://dx.doi.org/10.1016/bs.plm.2015.09.007

Kahan, T. A., & Enns, J. T. (2014). Long-term memory representations influence perception before edges are assigned to objects. *Journal of Experimental Psychology: General, 143,* 566–574. http://dx.doi.org/10.1037/a0033723

Kveraga, K., Ghuman, A. S., & Bar, M. (2007). Top-down predictions in the cognitive brain. *Brain and Cognition, 65,* 145–168. http://dx.doi.org/10.1016/j.bandc.2007.06.007

Levin, D. T., & Banaji, M. R. (2006). Distortions in the perceived lightness of faces: The role of race categories. *Journal of Experimental Psychology: General, 135,* 501–512. http://dx.doi.org/10.1037/0096-3445.135.4.501

Lupyan, G. (2013). Semantic effects on color afterimages. In *Meeting of the Vision Science Society,* Naples, FL. http://dx.doi.org/10.1167/13.9.466

Lupyan, G. (2015). Cognitive penetrability of perception in the age of prediction: Predictive systems are penetrable systems. *Review of Philosophy and Psychology, 6,* 547–569. http://dx.doi.org/10.1007/s13164-015-0253-4

Lupyan, G. (in press). Not even wrong. The "it's just X" fallacy. Commentary on Firestone & Scholl (2015). *Behavioral and Brain Sciences*.

Lupyan, G., & Clark, A. (2015). Words and the world: Predictive coding and the language-perception-cognition interface. *Current Directions in Psychological Science, 24,* 279–284. http://dx.doi.org/10.1177/0963721415570732

McClelland, J. L., Mirman, D., & Holt, L. L. (2006). Are there interactive processes in speech perception? *Trends in Cognitive Sciences, 10,* 363–369. http://dx.doi.org/10.1016/j.tics.2006.06.007

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: 1. An account of basic findings. *Psychological Review, 88,* 375–407. http://dx.doi.org/10.1037/0033-295X.88.5.375

Noppeney, U., Josephs, O., Hocking, J., Price, C. J., & Friston, K. J. (2008). The effect of prior visual information on recognition of speech and sounds. *Cerebral Cortex, 18,* 598–609. http://dx.doi.org/10.1093/cercor/bhm091

O'Callaghan, C., Kveraga, K., Shine, J. M., Adams, R. B., Jr., & Bar, M. (2016). Predictions penetrate perception: Converging insights from brain, behaviour and disorder. *Consciousness and Cognition*. http://dx.doi.org/10.1016/j.concog.2016.05.003

Prinzmetal, W., & Lyon, C. E. (1996). The word-detection effect: Sophisticated guessing or perceptual enhancement? *Memory & Cognition, 24,* 331–341. http://dx.doi.org/10.3758/BF03213297

Pylyshyn, Z. (1999). Is vision continuous with cognition? The case for cognitive impenetrability of visual perception. *Behavioral and Brain Sciences, 22,* 341–365. http://dx.doi.org/10.1017/S0140525X99002022

Raftopoulos, A. (2015). Reframing the problem of cognitive penetrability. In L. Magnani, P. Li, & W. Park (Eds.), *Philosophy and cognitive science II* (pp. 3–20). New York, NY: Springer International Publishing. http://dx.doi.org/10.1007/978-3-319-18479-1_1

Rao, R. P., & Ballard, D. H. (1999). Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience, 2,* 79–87. http://dx.doi.org/10.1038/4580

Reicher, G. M. (1969). Perceptual recognition as a function of meaningfulness of stimulus material. *Journal of Experimental Psychology, 81,* 275–280. http://dx.doi.org/10.1037/h0027768

Riesenhuber, M., & Poggio, T. (2000). Models of object recognition. *Nature Neuroscience, 3,* 1199–1204. http://dx.doi.org/10.1038/81479

Rumelhart, D. E., & McClelland, J. L. (1982). An interactive activation model of context effects in letter perception: II. The contextual enhancement effect and some tests and extensions of the model. *Psychological Review, 89,* 60–94. http://dx.doi.org/10.1037/0033-295X.89.1.60

Serre, T., Oliva, A., & Poggio, T. (2007). A feedforward architecture accounts for rapid categorization. *Proceedings of the National Academy of Sciences of the United States of America, 104,* 6424–6429. http://dx.doi.org/10.1073/pnas.0700622104

Siegel, S. (2012). Cognitive penetrability and perceptual justification. *Noûs (Detroit, Mich.), 46,* 201–222. http://dx.doi.org/10.1111/j.1468-0068.2010.00786.x

Summerfield, C., & Egner, T. (2009). Expectation (and attention) in visual cognition. *Trends in Cognitive Sciences, 13,* 403–409. http://dx.doi.org/10.1016/j.tics.2009.06.003

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology, 12,* 97–136. http://dx.doi.org/10.1016/0010-0285(80)90005-5

Wang, Z., & Simoncelli, E. P. (2003). Local phase coherence and the perception of blur. *Advances in Neural Information Processing Systems*. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.9.3565

Watson, A. B., & Ahumada, A. J. (2011). Blur clarified: A review and synthesis of blur discrimination. *Journal of Vision, 11,* 10. http://dx.doi.org/10.1167/11.5.10

Webster, M. A., Georgeson, M. A., & Webster, S. M. (2002). Neural adjustments to image blur. *Nature Neuroscience, 5,* 839–840. http://dx.doi.org/10.1038/nn906

Werner, S., & Thies, B. (2000). Is "change blindness" attenuated by domain-specific expertise? An expert-novices comparison of change detection in football images. *Visual Cognition, 7,* 163–173. http://dx.doi.org/10.1080/135062800394748

Wiley, R. W., Wilson, C., & Rapp, B. (2016). The effects of alphabet and expertise on letter perception. *Journal of Experimental Psychology: Human Perception and Performance, 42,* 1186–1203. http://dx.doi.org/10.1037/xhp0000213

Wilson, M. D. (1988). MRC psycholinguistic database: Machine-usable dictionary, version 2.00. *Behavior Research Methods, Instruments, & Computers, 20,* 6–10. http://dx.doi.org/10.3758/BF03202594

Witt, J. K., Taylor, J. E. T., Sugovic, M., & Wixted, J. T. (2015). Signal detection measures cannot distinguish perceptual biases from response biases. *Perception, 44,* 289–300. http://dx.doi.org/10.1068/p7908

Received March 2, 2016
Revision received October 20, 2016
Accepted October 20, 2016 ■