Not all perceptual difficulties lower memory predictions: Testing the perceptual fluency hypothesis with rotated and inverted object images

Miri Besken¹ · Elif Cemre Solmaz¹ · Meltem Karaca^{1,2} · Nilsu Atılğan^{1,3}

Published online: 21 February 2019
© The Psychonomic Society, Inc. 2019

Abstract

Studies typically show that perceptual difficulties at the time of encoding lower memory predictions. One potential exception to this is the inverted-word manipulation, in which participants produce equivalent memory predictions for upright and inverted words, despite higher free-recall performance for the inverted words (Sungkhasettee, Friedman, & Castel in *Psychonomic Bulletin & Review*, 18, 973–978, 2011). In the present set of experiments, we aimed to investigate the contributions of online perceptual difficulties versus a priori beliefs through two disfluency manipulations conceptually similar to the inverted-word manipulation: inversion and canonicity. The *inversion* manipulation involved presentation of upright and inverted object images, whereas the *canonicity* manipulation involved presentation of objects to participants from frequent (canonical) or infrequent (noncanonical) viewing perspectives. Memory predictions were made either on an item-by-item basis or aggregately. In all studies, the perceptual identification latencies for inverted and noncanonical items were slower than those for upright and canonical items, respectively. In experiments conducted with item-by-item memory predictions, predictions were not significantly different from each other across encoding conditions. In contrast, in experiments using aggregate memory predictions, fluent items produced higher memory predictions than did disfluent items. These results show that in certain cases, participants may not consider online objective perceptual difficulties. Moreover, item-by-item and aggregate memory predictions produce different patterns, evidence of a dissociation between the two types of predictions. The results are discussed in light of theories that rely on objective perceptual fluency differences across encoding conditions versus theories that rely on participants' a priori beliefs about fluency.

Keywords Metamemory · Perceptual fluency · Judgments of learning (JOLs) · Image rotation · Image inversion

Perceptual difficulties produced by experimental manipulations typically lead to lower judgments of learning (JOLs), one's confidence in their subsequent memory performance at the time of encoding (Rhodes, 2016). In the visual modality, blurred words or backward-masked words produce lower JOLs than do clear or intact words (Besken & Mulligan, 2013; Yue, Castel, & Bjork, 2013). Likewise, fragmented

object images produce lower JOLs than do intact object images (Besken, 2016). Similarly, in the auditory modality, words presented at a low volume (Frank & Kuhlmann, 2017; Rhodes & Castel, 2009) or words interspersed with silences (Besken & Mulligan, 2014; Susser, Mulligan, & Besken, 2013) lead to lower JOLs than do words presented at a high volume or intact words. Yet perceptual difficulties do not usually affect later retrieval. Rather, participants mistakenly attribute ease of encoding to ease of retrieval at a subsequent memory test, creating a metacognitive illusion (Rhodes & Castel, 2008).

The lower JOLs for the perceptually more difficult than for the easy items is usually explained through two competing accounts: the perceptual fluency hypothesis (Rhodes & Castel, 2008) and analytic processing theory (Dunlosky, Mueller, & Tauber, 2015; Mueller & Dunlosky, 2017; Mueller, Dunlosky, Tauber, & Rhodes, 2014). The *perceptual*

✉ Miri Besken
mbesken@bilkent.edu.tr

¹ Department of Psychology, Bilkent University, Ankara, Turkey

² Department of Psychology, University of Massachusetts, Lowell, MA, USA

³ Department of Psychology, University of Minnesota, Minneapolis, MN, USA

fluency hypothesis contends that disfluent items create more subjective difficulties at the time of encoding, specific to each episode than do fluent items (Rhodes & Castel, 2008). This hypothesis relies on *experience-based processes* specific to each episode, such as accuracy and processing latency at the time of encoding (Koriat, Bjork, Sheffer, & Bar, 2004). *Analytic processing theory*, on the other hand, argues that the differences between perceptually fluent and disfluent conditions can be explained by participants' general knowledge about learning and memory processes. According to this view, participants look for cues that reduce uncertainty at the time of encoding in order to make memory predictions (Mueller & Dunlosky, 2017). For example, they may have a general belief that fluent items will be remembered more than disfluent items and apply this belief to their JOLs. Analytic processing theory relies on *theory-based processes*, or participants' general knowledge about learning and memory without reference to specific episodes, and can usually be tested through questionnaires (Koriat et al., 2004). Both experience-based processes (Rhodes & Castel, 2008, 2009; Undorf, Zimdahl, & Bernstein, 2017; Yang, Huang, & Shanks, 2018) and theory-based processes (Mueller & Dunlosky, 2017; Mueller et al., 2014; Susser, Jin, & Mulligan, 2016) have gained ground in explaining the perceptual fluency findings, and the controversy is not yet resolved.

Perceptually fluent items typically receive higher JOLs than perceptually disfluent items, but there are exceptions to this finding. One important exception is a study by Sungkhasettee, Friedman, and Castel (2011). In this study, participants were presented with upright or inverted words, followed by JOLs after each item. Even though participants' JOLs were not significantly different across encoding conditions, participants' free-recall performance was higher for inverted than for upright words. Even though the theoretical account that the researchers used for explaining this finding is different, the findings of the inversion manipulation may also be interpreted as a manipulation that violates the typical perceptual fluency hypothesis. As the classic mental rotation task with two-dimensional (2-D) and three-dimensional (3-D) objects (Cooper, 1975; Shepard & Metzler, 1971) and its application to words (Koriat & Norman, 1984) show, the further an object or a word is rotated from its original or upright position on a 2- or 3-D plane, the longer participants take to identify the item through same/different judgments. Thus, if the perceptual fluency hypothesis is valid, one would assume that a word that is rotated 180 deg on a 2-D plane would take longer to identify than the upright item (Koriat & Norman, 1984), and it should cause objective online perceptual identification difficulties. If JOLs are driven by these difficulties, participants should produce lower JOLs for inverted than for upright words. However, this was not the case in Sungkhasettee et al.'s findings: They did not find JOL differences across encoding conditions in their experiment. Yet, they also did not take

objective measures of perception. Thus, it is difficult to make this claim without employing objective measures. Perhaps the inversion manipulation might be a good candidate to show whether or not objective perceptual difficulties are always reflected in JOLs. The first goal of the present set of experiments was to investigate whether manipulations that are pertinent to viewing perspective produce objective difficulties for participants, and whether these difficulties are reflected in JOLs. If the increased identification latency for more disfluent items were not reflected in the JOLs, this might constitute proof against the direct effect of experience-based cues on JOLs.

A second issue is that there may be differences between measures depending on whether memory predictions are taken aggregately or on an item-by-item basis. Most of the time, aggregate JOLs and item-by-item JOLs produce similar patterns (e.g., Besken & Mulligan, 2013, 2014), but experience-based versus belief-based cues may not always affect item-by-item and aggregate JOLs to the same extent. In a recent study, Frank and Kuhlmann (2017) presented participants with high- and low-volume words, followed by immediate item-by-item JOLs. In addition, they asked participants to indicate which type of words they predicted they would remember better either before or after the studying phase through aggregate JOLs. The results revealed that even when participants did not predict remembering words differently across encoding conditions on an aggregate basis, they gave higher JOLs to high-volume than to low-volume words on an item-by-item basis, producing a dissociation between the two types of memory predictions. Thus, Frank and Kuhlmann contended that experience-based cues may override beliefs even for participants with no a priori beliefs about a manipulation. However, they did not assess whether it was specifically fluency driving these decisions. If theory-based versus experience-based cues affect global and item-by-item memory predictions to differing extents, it is important to assess how experience-based cues such as fluency contribute to these two types of judgments. In the present study, we systematically investigated how aggregate and item-by-item memory predictions are affected by viewing perspective manipulations, which create a high potential of finding objective perception differences across fluent and disfluent encoding conditions. Dissociations between these measures might provide us with more insight about how theories and experience are employed in making memory predictions on the item and aggregate levels.

Another goal of the present set of studies was to investigate how inversion and canonicity manipulations might affect actual memory performance. Research typically reveals that self-generation of an item produces higher memory performance than does passive processing of the item (Slamecka & Graf, 1978), including such perceptual manipulations as deleted-letter generation (Begg, Vinski, Frankovich, &

Holgate, 1991), letter-transposition generation (Kinoshita, 1989; Mulligan, 2002; Nairne & Widner, 1988), auditory generation (Besken & Mulligan, 2014; Susser et al., 2013), and picture generation (Besken, 2016; Kinjo & Snodgrass, 2000). When participants put more effort into identifying the items, this can sometimes boost performance for the hard-to-identify images (Besken, 2016; Besken & Mulligan, 2013, 2014; Sungkhasettee et al., 2011). In contrast, certain manipulations also show that this might not always be the case (e.g., Yue et al., 2013). Both inversion and canonicity are manipulations that require more effort in terms of identification, and it is not warranted that these manipulations would always enhance memory performance.

In the present set of experiments, we used two pictorial manipulations that produced perceptual difficulties in order to investigate the effects of perceptual fluency on identification latencies, JOLs, and free-recall performance. One manipulation involved the extension of Sungkhasettee et al.'s (2011) inversion manipulation to images. Participants were presented with upright or inverted images of objects that usually have a proper upright direction (e.g., a tree, bottle, or piano), followed by either immediate item-by-item JOLs (Exp. 1) or aggregate JOLs (Exp. 2) at the end of the encoding phase. Inverting an object corresponds to a 180-deg rotation of the object image in 2-D space. The second manipulation was a canonicity manipulation, in which participants viewed images of objects from a canonical (an item viewed at its most commonly viewed perspective, such as a guitar from the front) or a non-canonical (an item that is viewed from an uncommon perspective, such as guitar from the side) perspective, again followed by item-by-item (Exps. 3, 6, and 7) or aggregate (Exps. 4a, 4b, 6, and 7) JOLs and a free-recall test. The canonicity manipulation corresponds to rotation of the object in 3-D space in various degrees from its most viewed, upright position. For both of these manipulations, the more disfluent items (inverted or noncanonical items) should be identified more slowly than upright or canonical items. Moreover, if the perceptual fluency hypothesis is valid, these objective online difficulties should lead to higher JOLs for the more-fluent upright or canonical items than for the disfluent inverted or non-canonical items. However, if Sungkhasettee et al.'s findings generalize to object images, JOLs might not differ significantly from each other across encoding conditions, specifically in terms of item-by-item JOLs. Free-recall performance might then be higher for perceptually disfluent than for perceptually fluent items, since disfluent items require more effort during encoding, leading to increased performance, as in Sungkhasettee et al.'s study. However, encoding condition might also not produce an advantage for either encoding manipulation. The main reason for this prediction is the JOL reactivity finding (Mitchum, Kelley, & Fox, 2016; Soderstrom, Clark, Halamish, & Bjork, 2015), which refers to a reduction or elimination of encoding manipulation

differences across conditions when participants are asked to make item-by-item JOLs, because JOLs sometimes cause participants to pay attention to factors they otherwise would not.

Experiment 1

In Experiment 1, participants were presented with a mixed list of upright and inverted images, followed by immediate item-by-item JOLs. Participants had to identify the items before they entered their JOLs, and then they had to remember the items' names for a later memory test. We hypothesized that the identification latency should be slower for inverted than for upright images. If the perceptual fluency hypothesis is valid, easily perceived fluent upright images should also produce higher JOLs than disfluent inverted images. However, if Sungkhasettee et al. (2011) findings generalize to images, no differences should be seen for JOLs across conditions. Finally, free-recall performance could be similar across conditions, due to reactivity, or higher for inverted than for upright images, due to desirable difficulties.

Method

Participants A priori power analyses showed that 26 participants were required in order to detect of a moderate size effect of $d = 0.5$ with a power of .80 and an alpha of .05. To keep the participants tested in each condition equal, 28 participants between the ages of 18 and 30 from Bilkent University participated in exchange for course credit. They were all native speakers of Turkish.

Material and design The encoding condition was manipulated within subjects with two levels, consisting of upright images and inverted images. These images were common items that had one specific direction (e.g., horse, bottle), selected from two image databases: a pool of pairs of related objects (POPORO; Kovalenko, Chaumon, & Busch, 2012) and the Bank of Standardized Stimuli (BOSS; Brodeur, Dionne-Dostie, Montreuil, & Lepage, 2010). These items were piloted with native Turkish speakers to ensure that the chosen images were identified and named identically by a majority of participants. The final image list consisted of 40 critical images and ten other images for the practice (two), primacy (four), and recency (four) portions of the list. Two separate lists were created, randomly assigning half of the critical images to each encoding condition. These lists were counterbalanced across participants for encoding condition and presented to equal numbers of participants. This and all subsequent experiments were approved by the Bilkent University Ethics Committee.

Procedure The experiment was conducted individually in a well-lit room. The experiment had three phases: encoding,

distractor, and testing phases. Before encoding, participants were told that they would see upright and inverted images, and that they would have to identify these items as quickly as possible, remember the items for a later memory test, and indicate their confidence that they would remember each item on the later memory test, on a scale from 0 (*I am not sure that I will remember this item at all*) to 100 (*I am quite sure that I will remember this item*), by typing the number on the screen.

Each trial started with a blank screen for 100 ms, followed by an upright or inverted image for 6,000 ms. For each trial, participants were asked to identify the image and type in its name as quickly as possible and press Enter when they had finished typing. As soon as they pressed Enter, the background turned gray and the response could not be modified, even though the participants could still view the objects in their entirety during the allotted time. The program automatically recorded the first-keypress latency (the time that elapsed from the onset of the image until participants pressed a key) and the total typing duration (the time from the first keypress until the participants pressed Enter) for each image. Six seconds after the onset of the image, the program proceeded to the JOL screen, regardless of whether or not participants had completed typing. Participants typed in a number between 0 and 100 and pressed Enter to proceed to the next trial. JOLs were self-paced and on a continuous scale. The order of critical images was randomized anew for each participant, with the prerequisite that no more than two items from the same encoding condition were shown consecutively.

For the distractor phase, participants were asked to solve arithmetic problems (e.g., $27 \times 4 = ?$) presented to them on the screen one at a time for a total of 3 min. For the testing phase, participants were asked to write down the names of all upright and inverted images that they could remember from the first part of the experiment within 5 min. Participants could terminate this part earlier by pressing the ESC key.

Results and discussion

The descriptive statistics are presented in Table 1. The alpha level was set at .05. All the effect sizes for paired-samples t tests were calculated using Cohen's d_z , which takes into account within-subjects correlations, for this and subsequent experiments. During encoding, that identification rates for upright (97.50%) and inverted (96.85%) images were high and not significantly different across encoding conditions, by a Wilcoxon signed-rank test ($p = .574$).

For encoding, the median first keypress response latency and typing duration were calculated separately for upright and inverted images for each participant, excluding trials in which participants misidentified the image or did not type in at all (exclusion rate = 2.85%), and the means of the medians were submitted to paired-samples t test. Participants' first-keypress latency was significantly faster for upright than for inverted

pictures, $t(27) = 3.29$, $p = .003$, $r = .80$, Cohen's $d = 0.62$. Typing durations did not differ significantly across encoding conditions, $t(27) = 0.85$, $p = .403$, $r = .93$, $d = 0.16$.

For JOLs, all trials in which the participants misidentified the items or failed to enter a value between 0 and 100 for JOLs were excluded from the analyses (exclusion rate = 3.57%). Participants' JOLs across encoding conditions were not significantly different from each other, $t(27) = 1.11$, $p = .278$, $r = .91$, $d = 0.21$. Counts and percentages of participants who produced different patterns of JOLs (higher JOLs for upright than for inverted items, higher JOLs for inverted than for upright items, or equal JOLs for inverted and upright items) and effect sizes for JOL differences across encoding conditions are presented in Table 2 for this and all subsequent experiments.

High identification rates at study make it possible to analyze free-recall data without conditionalizing on correct study identification. Since the unconditional and conditionalized data produced the same results across all experiments, only unconditional free-recall data are reported for this and all subsequent experiments. Participants' free-recall performance was not significantly different across encoding conditions, $t(27) = 0.15$, $p = .879$, $r = .42$, $d = 0.03$.

To determine how accurate participants were in estimating their actual memory performance, mean Goodman–Kruskal gamma correlations were calculated separately for each participant for both upright and inverted images. Both encoding conditions yielded mean gamma correlations significantly higher than 0— $t(27) = 3.01$, $p = .006$, $d = 0.57$, for upright images; $t(27) = 3.84$, $p < .001$, $d = 0.73$, for inverted images—but they were not significantly different from each other, $t(27) = 0.07$, $p = .945$, $r = -.18$, $d = 0.01$.

As we hypothesized, participants were slower to identify inverted than upright images, showing objective experiential difficulties in perception. Yet this objective perceptual difficulty was reflected in neither JOLs nor free-recall performance. In other words, we found no evidence that objective online difficulties were associated with item-by-item JOLs. In terms of JOLs, this finding is consistent with Sungkhasettee et al.'s (2011) finding that the inversion manipulation produces results for images similar to its results for simple word stimuli. One possibility is that the inversion manipulation is more readily driven by *beliefs about the effects of the manipulation* rather than by *objective online difficulties*. This is tested further in the following experiments.

In terms of actual memory performance, we did not find any advantage for inverted images, as compared with upright images. Obviously, our methodology was slightly different from that of Sungkhasettee et al. (2011), in that we required participants to type the names of the object images, whereas Sungkhasettee et al. only asked participants to say the word aloud. Processing and identifying

Table 1 Means and standard errors of the means (in parentheses) for first-keypress latencies, total typing durations, item-by-item JOLs, aggregate JOLs, proportions of correct recall, and Goodman–Kruskal Gamma correlations for each encoding condition of Experiments 1–7

Exp #	Manipulation	First-keypress latency (ms)			Total typing duration (ms)			Item-by-Item JOL (Out of 100)			Aggregate JOL (Out of 100)			Proportion correct recall			Goodman–Kruskal Gamma correlations		
		Upright/ Canonical	Inverted/ Noncanonical		Upright/ Canonical	Inverted/ Noncanonical		Upright/ Canonical	Inverted/ Noncanonical		Upright/ Canonical	Inverted/ Noncanonical		Upright/ Canonical	Inverted/ Noncanonical		Upright/ Canonical	Inverted/ Noncanonical	
1	Inversion	1,337.39 (41.26)	1,420.61 (36.94)		1,489.23 (104.30)	1,456.48 (97.24)		60.99 (3.37)	59.28 (3.75)		–	–		0.38 (0.02)	0.38 (0.02)		0.21 (0.07)	0.21 (0.06)	
2	Inversion	1,178.15 (23.30)	1,305.17 (36.39)		1,102.85 (35.24)	1,108.27 (38.31)		–	–		62.96 (3.82)	57.00 (4.44)		0.38 (0.02)	0.37 (0.02)		–	–	
3	Canonicity	1,447.22 (40.59)	1,564.22 (47.41)		1,621.09 (108.88)	1,575.00 (103.15)		71.79 (3.02)	70.42 (2.79)		–	–		0.43 (0.02)	0.46 (0.03)		0.20 (0.06)	0.15 (0.06)	
4a	Canonicity	–	–		–	–		–	–		63.16 (4.30)	55.47 (4.74)		0.38 (0.02)	0.43 (0.02)		–	–	
4b	Canonicity	1,375.33 (35.98)	1,465.06 (39.92)		1,311.55 (71.33)	1,319.81 (57.69)		–	–		59.41 (3.81)	50.94 (3.84)		0.38 (0.02)	0.42 (0.02)		–	–	
5	Canonicity	–	–		–	–		–	–		73.10 (2.27)	59.44 (2.85)		–	–		–	–	
6	Canonicity	1,486.33 (35.16)	1,644.58 (39.54)		1,694.77 (145.84)	1,701.88 (124.47)		69.43 (3.70)	67.28 (3.51)		67.91 (3.45)	53.44 (4.05)		0.42 (0.02)	0.43 (0.02)		0.26 (0.06)	0.27 (0.07)	
7	Canonicity	1,454.94 (41.14)	1,548.61 (40.03)		1,607.36 (96.18)	1,615.44 (100.82)		76.42 (3.57)	72.18 (3.63)		69.00 (2.78)	48.28 (3.08)		0.43 (0.03)	0.49 (0.02)		0.11 (0.09)	0.01 (0.10)	

Table 2 Breakdown of participants who gave higher judgments of learning (JOLs) to fluent (upright/canonical) than to disfluent (inverted/noncanonical) items, higher JOLs to disfluent than to fluent items, and

equal JOLs for fluent and disfluent items, in terms of both counts and percentages, indicated separately for item-by-item and aggregate JOLs, plus the effect size for each effect in Experiments 1–7

Exp #	Item-by-Item JOLs								Aggregate JOLs							
	Fluent > Disfluent		Fluent = Disfluent		Fluent < Disfluent		Cohen's <i>d</i>	Fluent > Disfluent		Fluent = Disfluent		Fluent < Disfluent		Cohen's <i>d</i>		
	Count	Percentage	Count	Percentage	Count	Percentage		Count	Percentage	Count	Percentage	Count	Percentage			
1	17	60.71%	0	0.00%	11	39.29%	0.21	–	–	–	–	–	–	–		
2	–	–	–	–	–	–	–	13	50.00%	9	35.00%	4	12.50%	0.59		
3	18	56.25%	1	3.13%	13	40.63%	0.25	–	–	–	–	–	–	–		
4a	–	–	–	–	–	–	–	16	50.00%	10	31.25%	6	18.75%	0.44		
4b	–	–	–	–	–	–	–	20	62.50%	8	25.00%	4	12.50%	0.68		
5	–	–	–	–	–	–	–	42	67.74%	9	14.52%	11	17.74%	1.43		
6	20	62.50%	1	3.13%	11	34.38%	0.29	26	81.25%	1	3.13%	5	15.63%	0.83		
7	25	78.13%	1	3.13%	6	18.75%	0.53	30	93.75%	0	0.00%	2	6.25%	1.19		
Avg		64.40%		2.34%		33.26%	0.32		67.54%		18.15%		13.89%	0.86		

image names might possibly be a more laborious process, regardless of whether the image is upright or inverted, because this requires processing of the item both visually and verbally (Paivio, 1975, 1986), leading to no significant difference across encoding conditions. When words are used, participants do not have to process the meaning of the items; thus, both encoding conditions might have been encoded relatively more superficially with the word stimuli in Sungkhasettee et al.'s study than with the images in the present experiment.

Experiment 2

Participants' JOLs were not affected from the objective perceptual difficulties associated with the inversion manipulation for item-by-item JOLs. Typically, item-by-item JOLs are hypothesized to reflect objective online difficulties more concretely, in that they are made immediately after each item (Kelley & Jacoby, 1996). Another method to measure memory predictions is aggregate JOLs. These are memory predictions that are obtained from participants at the end of the encoding phase, also sometimes referred to as *poststudy global differentiated predictions* (e.g., Dunlosky & Hertzog, 2000; Frank & Kuhlmann, 2017). Aggregate JOLs, in contrast to item-by-item JOLs, may depend more on beliefs than on objective online difficulties (Frank & Kuhlmann, 2017). If the inversion manipulation affects participants' beliefs, using aggregate JOLs might lead to higher memory predictions for upright than for inverted items, because when participants are making predictions globally, the framing might be more obvious to participants and might activate beliefs, as compared with item-by-item JOLs. Free-recall

performance might be higher for inverted than for upright images, due to desirable difficulties, or it might be equivalent if the inversion manipulation is similar to the blurring manipulation, which also does not influence actual memory performance (Yue et al., 2013).

Method

Participants Twenty-six participants who did not participate in Experiment 1 were recruited from the same participant pool in return for course credit.

Design, materials, and procedure The design and the material were identical to those of Experiment 1. The only difference was that the item-by-item JOLs were replaced with aggregate JOLs at the end of the encoding phase. Once participants had studied all the images, they were told that they had been presented with equal numbers of upright and inverted images. They were asked to indicate separately for both upright and inverted images what percentage of each image type they thought they would remember in a subsequent memory test. The nature of the subsequent test was not specified further. The order of aggregate JOLs was counterbalanced across participants. The distractor and the recall phases were identical to those in Experiment 1.

Results

Descriptive statistics for Experiment 2 are presented in Table 1. Participants identified 96.35% of the upright and 94.80% of the inverted images correctly. The identification rates did not differ significantly across encoding conditions ($p = .313$). As in Experiment 1, participants' first-keypress latency was faster for upright than for inverted images, $t(25)$

$= 3.99$, $p = .001$, $r = .50$, $d = 0.78$, and typing durations were not significantly different across encoding conditions, $t(25) = 0.21$, $r = .76$, $p = .834$, $d = 0.04$ (misidentification exclusion rate = 4.42%). Unlike in Experiment 1, participants produced higher JOLs for upright than for inverted images, $t(25) = 3.01$, $p = .006$, $r = .90$, $d = 0.59$. As in Experiment 1, participants' free-recall performance was not significantly different for upright and inverted images, $t(25) = 0.41$, $p = .688$, $r = .44$, $d = 0.08$.

As in Experiment 1, the identification latency was slower for inverted than for upright items, with a difference of 127 ms. In a similar experimental setup, Jolicoeur (1985) found that 180-deg rotations of natural objects in the 2-D plane take 129 ms. Since inversion can be considered a 180-deg rotation of the object in the 2-D plane, this finding is quite consistent with the previous literature. Free recall did not change significantly across encoding conditions. However, unlike in Experiment 1, participants predicted they would have higher memory performance for upright than for inverted items, so the aggregate JOLs produced predictions that were in line with the perceptual identification speed. Yet, only aggregate JOLs showed the superiority for fluent over disfluent items. This finding might be a consequence of beliefs formed about the manipulation rather than a direct effect of perceptual online difficulties. Aggregate JOLs label the manipulation more clearly than does the mere presentation of upright and inverted object images. If participants are forming beliefs about the manipulation, this might be more obvious to the participants when it is clearly framed for them.

Experiment 3

The inversion manipulation produced identification difficulties for both experiments, but these identification difficulties were reflected only in memory predictions for aggregate JOLs, not in item-by-item JOLs. Even though JOLs are tested regularly on an item-by-item basis, aggregate JOLs are not used so frequently, and most of the time they reveal biases similar to those using item-by-item JOLs (Besken, 2016; Besken & Mulligan, 2013, 2014). However, in certain cases aggregate and item-by-item memory predictions may not reveal similar results (Frank & Kuhlmann, 2017), and the inversion manipulation seems to be one in which aggregate and item-by-item JOLs do not agree with each other. It is essential to show that this difference between item-by-item and aggregate JOLs did not occur by chance, but may represent a specific bias among participants given manipulations involving rotation. To see whether this was the case, we used another manipulation, canonicity, that could be considered conceptually similar to the inversion manipulation. Whereas inversion refers to 180-deg rotation of an image in 2-D space, in the canonicity manipulation, the image is rotated in 3-D space.

Even though both manipulations involve rotations, some differences between the manipulations exist, as well. For example, rotation in 3-D space may be more detrimental to object recognition than is rotation in the 2-D plane, because distinctive features and surfaces of the object might be partially or completely occluded when the object is rotated in 3-D space, or the global outline of the object might change relative to the canonical view (e.g., Humphrey & Jolicoeur, 1993; Lawson & Humphreys, 1998; Lawson, Humphreys, & Jolicoeur, 2000; Newell & Findlay, 1997).

In line with previous research relevant to the canonicity manipulation, perceptual identification latencies should be slower for noncanonical than for canonical items. If the perceptual fluency hypothesis holds, participants should produce lower JOLs for noncanonical than for canonical items. In contrast, if the canonicity manipulation is similar to the inversion manipulation, JOLs should not differ significantly across encoding conditions on an item-by-item basis.

Finally, regarding actual memory performance, to our knowledge only one other study has investigated the effects of canonicity on memory performance. Using the canonicity manipulation, Gomez, Shutter, and Rouder (2008) found that participants recognized more noncanonical than canonical objects. In contrast, participants' free-recall performance was higher for canonical than for noncanonical objects. Yet, Gomez et al.'s method for assessing long-term memory was somewhat different methodologically, with a low number of words on each list followed by immediate testing in various blocks, so that participants might have employed working memory rather than long-term memory. If we use the methods employed in typical memory long-term memory experiments, noncanonical items might be relatively more difficult to recognize and process, and they might require participants to generate parts of the items that were not seen in the image; thus, the manipulation might lead to better memory for noncanonical than for canonical items. Alternatively, memory performance might also not be significantly different across conditions, because of JOL reactivity (Mitchum et al., 2016; Soderstrom et al., 2015).

Method

Participants Thirty-two participants from a general education course at Bilkent University participated in the experiment for course credit. The sample size was determined to be 32 participants for Experiments 3, 4a, 4b, 6, and 7 because at least 26 participants are required to detect a moderate-sized effect of $d = 0.5$ with a power of .80 and an alpha of .05. The experiments conceptually closest to the present experiment (e.g., Besken, 2016, for picture fragmentation, and Sungkhasettee et al., 2011, for an inverted-word manipulation) also used sample sizes that varied from 20 to 32.

Design, materials, and procedure The experiment was conducted with encoding condition as the within-subjects variable, with two levels: canonical and noncanonical images. All materials were created by the researchers, by taking photos of 100 common items from various perspectives, and were presented to a group of pilot participants who identified the names of the objects and evaluated the materials for the frequency of viewing that item from that perspective in real life (ratings ranging from 1, *I never see this object from this angle*, to 5, *I almost always see this item from this angle*) through an online survey. A total of 54 items were chosen from this pool, consisting of 44 critical items and two practice, four primacy, and four recency items, which were identified with the same name from both canonical and noncanonical views and were given high points for canonical or low points for noncanonical items, in terms of the viewing frequency from that perspective. Two lists were created, with half of the critical items randomly assigned to each encoding condition, counterbalancing the items for encoding condition across participants. The order of presentation was randomized anew for each participant. The procedure was very similar in nature to that of Experiment 1. Participants were told that they would study pictures of objects from different perspectives. Some of the objects would be presented from perspectives they were used to (i.e., canonical), and some of the objects would be presented from

perspectives from which they might not be used to viewing the objects (i.e., noncanonical). Participants were asked to identify and type the names of the objects on the screen within 6 s and to remember these items for a subsequent memory test. The experimental program recorded the first-keypress latency and total typing duration. Each item preceded a self-paced, immediate, item-by-item JOL, identical to that in Experiment 1. An example of an object in its canonical and noncanonical formats is shown in Fig. 1, along with the procedure for the encoding phase. The distractor phase consisted of a pattern-matching exercise for 3 min, followed by a 5-min-long free-recall phase in which participants were asked to write down the names of all objects they could remember from the first phase of the experiment. Participants could self-terminate the recall phase earlier, if they wanted to.

Results

Descriptive statistics for Experiment 3 are presented in Table 1. Participants identified 94.59% of canonical and 95.31% of noncanonical items at encoding. The identification performance was not significantly different across encoding conditions ($p = .638$). As in Experiments 1 and 2, misidentified items were excluded from the response latency analyses (exclusion rate = 5.04%). Participants' first keypress was significantly faster for canonical than for noncanonical

Fig. 1 Example of an object image in its canonical and noncanonical formats and of the procedure for Experiments 3, 4a, 4b, 6, and 7. Experiments 4a and 4b did not include the self-paced item-by-item JOL display

items, $t(31) = 3.87$, $p < .001$, $r = .78$, $d = 0.69$, and total typing durations did not differ by encoding condition, $t(31) = 1.05$, $p = .301$, $r = .92$, $d = 0.19$.

Despite the significant difference in objective measures of identification latencies, participants' JOLs were not significantly different for canonical and noncanonical items, $t(31) = 1.44$, $p = .161$, $r = .95$, $d = 0.25$ (exclusion rate for misidentified items or JOLs not entered = 6.11%). The unconditional free-recall performance did not differ by encoding condition, $t(31) = 0.84$, $p = .405$, $r = .22$, $d = 0.15$.

Two one-sample t tests yielded mean Goodman–Kruskal gamma correlations for both canonical and noncanonical images that were significantly higher than 0— $t(30) = 3.34$, $p = .002$, $d = 0.60$, for canonical; $t(31) = 2.55$, $p = .016$, $d = 0.45$, for noncanonical—but the correlations were not significantly different from each other, $t(30) = 0.39$, $p = .700$, $r = -.05$, $d = 0.07$.

Experiments 4a and 4b

As with the inversion manipulation, item-by-item JOLs were not significantly different for fluent canonical images and disfluent noncanonical items on an item-by-item basis. It is important to show that the canonicity manipulation, which is conceptually similar to the inversion manipulation, also produces similar results for aggregate JOLs. Noncanonical stimuli should therefore produce slower identification latencies and lower aggregate JOLs than canonical items. Free-recall performance should be higher for noncanonical than for canonical items, however, due to the desirable-difficulties principle, or produce equivalent memory across encoding conditions, if participants' recognition are not more effortful for either condition.

Method

Participants Thirty-two participants for Experiments 4a and 4b each participated in the experiment for course credit for a general education course at Bilkent University.

Design, materials and procedure Experiments 4a and 4b were similar to Experiment 3, except for the replacement of item-by-item JOLs with aggregate JOLs. Before the experiment began, participants were told before the experiment that they would be presented with canonical and noncanonical items. Since participants were asked to make aggregate JOLs, it was important for them to distinguish between canonical and noncanonical items. To help them identify more easily which items were canonical and noncanonical, the pictures were marked with a thin blue or yellow rectangular frame around the object. The canonical images were framed with a yellow frame,

and the noncanonical images were framed with a blue frame, and participants were informed about the meaning of the color scheme. The frame color was counterbalanced across participants per encoding condition. At the end of the encoding phase, participants were asked to make two predictions separately for canonical and noncanonical items. In Experiment 4a, the first-keypress latency and total typing duration were not recorded due to a programming error. The experiment was conducted again after the mistake was corrected as Experiment 4b. The results for both experiments are reported.

Results

Descriptive statistics for Experiments 4a and 4b are presented in Table 1. For Experiment 4a, the identification rates were 95.04% and 97.72%, respectively, for canonical and noncanonical items, and differed significantly by a sign test ($p = .036$). Participants predicted that their memory performance would be significantly higher for canonical than for noncanonical items, $t(31) = 2.50$, $p = .018$, $r = .77$, $d = 0.44$, despite their superior actual free-recall performance for noncanonical relative to canonical items, $t(31) = 2.20$, $p = .035$, $r = .33$, $d = 0.39$.

For Experiment 4b, the identification rates were 95.31% and 95.04%, respectively, for canonical and noncanonical items. Participants' first keypress was significantly faster for canonical than for noncanonical items, $t(31) = 3.15$, $p = .004$, $r = .72$, $d = 0.56$, and typing duration did not differ by encoding condition, $t(31) = 0.15$, $p = .885$, $r = .63$, $d = 0.03$ (exclusion rate = 4.83). Participants predicted they would remember a higher percentage of canonical than of noncanonical items, $t(31) = 3.83$, $p < .001$, $r = .83$, $d = 0.68$, despite the null effect for actual free-recall performance, $t(31) = 1.71$, $p = .096$, $r = .31$, $d = 0.30$.

The aim of Experiments 3, 4a, and 4b was to investigate whether the canonicity manipulation, which in theory might be similar to the inversion manipulation (Humphrey & Jolicoeur, 1993; Lawson & Humphreys, 1998; Newell & Findlay, 1997), would produce similar results in terms of objective identification latencies and of predicted and actual memory performance. As with the inversion manipulation, it took longer for participants to identify noncanonical than to identify canonical items, and these differences were not reflected in item-by-item JOLs, but they were reflected in aggregate JOLs. The free-recall performance was slightly higher for noncanonical than for canonical items, but this difference did not always reach significance. At the end of four experiments, one might claim that both inversion and canonicity manipulations affect participants' predicted and actual free-recall performance similarly. These manipulations of viewing

perspective, through both inversion and rotation, affect only aggregate JOLs and not item-by-item JOLs. Thus, perceptual difficulties are not always manifested in memory predictions. JOLs may sometimes need to be framed and presented aggregately in order to produce an effect, which provides evidence against experience-based processes. However, the lack of evidence for experience-based processes in determining memory predictions does not automatically provide evidence for theory-based processes.

Experiment 5

The aim of Experiment 5 was to investigate whether manipulations such as canonicity or inversion may be influenced by theory-based processes. In other words, we investigated whether people have a priori beliefs about the effects of a perspective-pertinent manipulation even before they are exposed to the manipulation. One common method to assess a priori beliefs about the effects of a manipulation is to present participants with the scenario of the actual experiment and ask them to make predictions about the scenario (Besken, 2016; Koriati et al., 2004; Kornell, Rhodes, Castel, & Tauber, 2011; Mueller et al., 2014; Susser & Mulligan, 2015). To test this, a group of online participants who had not taken part in any of the previous experiments were given the scenario of the canonicity manipulation and asked to make predictions about their memory performance for canonical and noncanonical items. If participants have a priori beliefs about the canonicity manipulation, this should be in line with the findings from Experiments 4a and 4b, with participants giving higher memory predictions to canonical than to noncanonical items. The presence or absence of an example in the description of the scenario was used as an additional independent variable in the present experiment, because the presence of an example might be a specific instance of experience-based processes and might affect memory predictions differentially. If presenting a single instance of both a canonical and a noncanonical item produces a main effect or interaction, this could constitute evidence for experience-based processes.

Method

Participants The experiment was advertised on various social network pages, and a total of 66 people participated. One group was shown an exemplar of one canonical and one noncanonical item, whereas the other group was not. As in Experiments 3, 4a, and 4b, we aimed to have 32 participants per condition, summing up to 64 participants for our total sample size. Two more participants participated online before we were able to close the online survey. Four of the participants were excluded from the analyses, because they reported

that they had already participated in Experiment 3, 4a, or 4b at Bilkent University.

Design, materials, and procedure Participants were presented with the canonicity manipulation scenario. The program assigned participants randomly to a version of the scenario. In one version of the scenario, participants were presented with one instance of a canonical and a noncanonical item ($N = 31$). In the other version, no examples were presented ($N = 31$). Participants were asked to indicate their predicted percentages of free-recall performance for both encoding conditions, on a sliding scale.

Results

When the scenario was given along with an instance of a canonical and a noncanonical image, the mean prediction percentages were 74.65 for canonical images ($SD = 18.29$) and 58.39 for noncanonical images ($SD = 24.18$). When no instance of canonical and noncanonical items was given along with the scenario, the mean prediction percentages were 71.55 ($SD = 17.64$) for canonical and 60.48 for noncanonical images ($SD = 20.84$). These means were submitted to a repeated measures analysis of variance (ANOVA) with encoding condition as the repeated measure and the presence of an example as the between-subjects variable. The main effect of encoding condition was significant, $F(1, 60) = 30.46$, $MS = 189.90$, $p < .001$, $\eta_p^2 = .34$, with higher memory predictions for canonical than for noncanonical images. The main effect of the presence of an example was not significant, $F(1, 60) = 0.01$, $MS = 642.20$, $p = .913$, $\eta_p^2 < .01$. The interaction was also not significant, $F(1, 60) = 1.10$, $MS = 189.90$, $p = .298$, $\eta_p^2 = .02$.

The results of Experiment 5 revealed that even if participants did not experience the experimental material personally, they formed beliefs that are in line with typical beliefs about fluency: They mistakenly believed that the canonical items would be remembered better than noncanonical items, similar to those participants who actually went through the process in Experiments 4a and 4b. Yet, there were no differences in the JOLs across encoding conditions in Experiment 3, in which participants were asked to make item-by-item judgments as they went through the process. Hence, the beliefs did not seem to affect JOLs in Experiment 3, despite the objective perceptual difficulties associated with processing noncanonical items.

Experiment 6

One might wonder whether aggregate JOLs are in line with the experience of the canonicity manipulation. On the one hand, one might experience a certain type of subjective feeling (e.g., that canonical items were easier to process than

noncanonical items), but the feeling might not be correlated with one's aggregate JOLs, because item-by-item and aggregate JOLs might not depend on the same bases of information (Frank & Kuhlmann, 2017; Kelley & Jacoby, 1996). On the other hand, the experience of going through the canonicity manipulation might also change one's general beliefs about the effects of the manipulation. Thus, the first aim of Experiment 6 was to see whether aggregate JOLs and item-by-item JOLs were correlated or independent of each other. To investigate this, we used a method that had previously been used by Frank and Kuhlmann. Frank and Kuhlmann asked participants to make both item-level and aggregate JOLs for a volume manipulation. The results revealed that even when participants did not believe that items presented in a loud volume would be remembered better than items presented in a quiet volume while making their aggregate judgments, they still gave higher item-by-item JOLs to high-volume than to low-volume items. It is important to see whether such a dissociation is also present for the canonicity manipulation. If one could show that the pattern is different for "believers" than for "nonbelievers," this could partially disentangle the relationship between beliefs and experience for item-by-item JOLs. The free-recall performance should be similar to the findings in Experiment 3, because the general design was quite similar, except that additional aggregate JOLs were solicited at the end of the encoding phase.

Method

Participants Thirty-two participants from Bilkent University participated in the experiment for either course credit or ten Turkish liras (~\$2 at the time of participation).

Design, materials, and procedure The design, materials, and procedure were identical to the canonicity manipulation with immediate item-by-item JOLs in Experiment 3. The only modification was the addition of the aggregate JOLs to the end of the encoding phase, as in Experiments 4a and 4b. Once the encoding phase was completed, participants were asked to indicate what percentages of the items they predicted they would remember in both canonical and noncanonical points of view. Unlike Experiments 4a and 4b, there was no frame around the images that showed participants' encoding condition, in order to control for the confound of framing between Experiments 3 and 4.

Results

Descriptive statistics for Experiment 6 are presented in Table 1. Participants identified 97.45% of canonical and 97.45% of noncanonical items at encoding. The identification performance was not significantly different across encoding conditions ($p = 1.00$). As in the previous experiments,

misidentified items were excluded from the analyses (exclusion rate = 2.56%). Participants' first-keypress latencies were significantly faster for canonical than for noncanonical items, $t(31) = 5.83, p < .001, r = .74, d = 1.03$. There was no difference in total typing times across encoding conditions, $t(31) = 0.11, p = .916, r = .89, d = 0.02$. Participants' item-by-item JOLs were not significantly different from each other, $t(31) = 1.62, p = .116, r = .93, d = 0.29$. However, they predicted that they would remember a higher percentage of canonical than of noncanonical items in the aggregate JOLs, $t(31) = 4.69, p < .001, r = .67, d = 0.83$. Participants' free-recall performance was not significantly different across encoding conditions, $t(31) = 0.43, p = .671, r = -.02, d = 0.08$.

One way to disentangle the effects of beliefs versus experience is to separate participants into a group of believers (aggregate JOL for canonical > aggregate JOL for noncanonical) versus nonbelievers (aggregate JOLs for canonical < aggregate JOL for noncanonical, or aggregate JOLs for canonical = aggregate JOL for noncanonical). In all, 81% of the participants produced higher aggregate JOLs for canonical than for noncanonical items. A t test with believers as a variable revealed that they also produced significantly higher item-by-item JOLs for canonical ($M = 71.17, SE = 4.04$) than for noncanonical ($M = 68.26, SE = 4.00$) items, $t(25) = 2.37, p = .026, r = .95, d = 0.46$. Also, 6% of the participants produced equal aggregate JOLs across encoding conditions, and another 13% of the participants produced higher aggregate JOLs for noncanonical than for canonical items. Even though only six of the 32 participants qualified as nonbelievers in this sample, a paired-sample t test for this group revealed no significant difference across encoding conditions for item-by-item JOLs: $t(5) = 0.26, p = .808, r = .86, d = 0.11; M = 61.86, SE = 9.31$ for canonical items; $M = 63.08, SE = 7.61$ for noncanonical items. Obviously, the sample consisted of only six participants; thus, the results should be approached with caution.

Two one-sample t tests revealed that the Goodman–Kruskal gamma correlations for both canonical and noncanonical items were significantly above 0: $t(30) = 4.03, p < .001, d = 0.72$, for canonical items; $t(31) = 3.75, p < .001, d = 0.66$, for noncanonical items. The resolution did not differ significantly across conditions by a paired samples t test, $t(30) = 0.11, p = .917, r = .20, d = 0.02$.

As in the previous experiments, participants typically produced higher aggregate JOLs for canonical than for noncanonical items, despite the lack of significant differences across encoding conditions for item-by-item JOLs, providing further proof that aggregate JOLs and item-by-item JOLs may not always reflect the same processes. A more thorough partitioning of the data revealed that participants whose item-by-item JOLs were higher for canonical than for noncanonical items also produced higher aggregate JOLs for canonical than for noncanonical items. Yet the present data do not inform us as to whether the aggregate data pattern was formed

through the subjective online experience of perceiving canonical faster than noncanonical items, or whether this was a preconceived notion about the canonicity manipulation.

Experiment 7

Experiment 6 revealed that if one rates canonical items higher than noncanonical items at the list level, item-level JOLs tend to agree with this pattern. Yet this finding might have been a consequence of experience with the manipulation, moderating the aggregate JOLs. If this is a preconceived notion, we should also see the same pattern when participants make their aggregate JOLs *ahead* of the encoding phase. For Experiment 7, participants first made their aggregate JOLs, followed by the encoding phase with item-by-item JOLs. This was the first experiment in which participants made their JOLs ahead of the learning phase. If participants are asked to pay attention to the manipulation before the experiment starts, this type of a manipulation might potentially also change metacognitive control processes, producing higher free-recall performance for noncanonical than for canonical items. Alternatively, it might not have an influence on memory performance, as in the previous experiments.

Method

Participants Thirty-two participants from Bilkent University participated in the experiment for course credit or ten Turkish liras.

Design, materials, and procedure The procedure was identical to that of Experiment 6. The only difference was that aggregate JOLs were collected at the beginning rather than at the end of the encoding phase.

Results

Descriptive statistics for Experiment 7 are presented in Table 1. Participants identified 96.86% of the canonical and 95.45% of the noncanonical items at encoding. Identification rates did not differ significantly across encoding conditions by a sign test ($p = .188$). As in the previous experiments, misidentified items were excluded from the response latency analyses (exclusion rate = 3.8%). First-key response latencies were faster for canonical than for noncanonical items, $t(31) = 3.85, p < .001, r = .82, d = 0.68$, and there was no difference in total typing times across encoding conditions, $t(31) = 0.17, p = .869, r = .88, d = 0.03$. Item-by-item JOLs were significantly higher for canonical than for noncanonical items, $t(31) = 3.00, p = .005, r = .92, d = 0.53$. Similarly, participants produced significantly higher aggregate JOL ratings for canonical than for noncanonical items, $t(31) = 6.71, p < .001, r = .45, d =$

1.19. However, participants' free-recall performance was higher for noncanonical than for canonical items, $t(31) = 2.10, p = .044, r = .29, d = 0.37$.

Two one-sample t tests yielded mean Goodman–Kruskal gamma correlations for both canonical and noncanonical image that were not significantly higher than 0— $t(27) = 1.14, p = .263, d = 0.22$, for canonical, $t(31) = 0.12, p = .96, d = 0.02$, for noncanonical—and that also were not significantly different from each other, $t(27) = 0.39, p = .698, r = -.18, d = 0.07$.

Across seven experiments, this was the first time that item-by-item JOLs were significantly higher for canonical than for noncanonical items. However, a closer investigation of the dataset revealed that 30 of the 32 participants (93.75%) predicted ahead of the study phase that their memory would be higher for canonical than for noncanonical items through their aggregate JOLs. The proportion of believers to nonbelievers was quite high for this specific experiment, and this preconceived notion that canonical items would produce higher memory performance than noncanonical items might have affected the item-level JOLs in the same direction, leading to a significant difference across encoding conditions on an item-level basis. When the aggregate JOLs were acquired ahead of the study phase, it may have forced participants to pay closer attention to the experimental manipulation, which in turn might have been reflected in item-level JOLs. This is consistent with Mueller and Dunlosky (2017), who showed that experimenters can induce certain beliefs in participants before the experiment starts that are reflected in item-by-item JOLs—for instance, by suggesting that one color is more memorable than the other, without changing the actual fluency of the materials.

General discussion

Perceptual difficulties, which can be measured through identification latencies, typically lower memory predictions (Besken, 2016; Besken & Mulligan, 2013, 2014; Undorf et al., 2017; Yang et al., 2018), but there are exceptions to this finding. For example, the word inversion manipulation does not produce JOL differences across encoding conditions (upright and inverted words), despite the increased memory performance for inverted as compared to upright words (Sungkasettee et al., 2011). In the present set of experiments, we first aimed to investigate the contributions of experience-based processes (i.e., subjective difficulties) versus theory-based processes (i.e., general knowledge about learning and memory) to two disfluency manipulations conceptually similar to the inverted-word manipulation: inversion and canonicity. Second, we investigated how experience-based versus theory-based processes may be used differentially when memory predictions are obtained either on an item-by-item basis or aggregately. Third, we investigated the effects of these

manipulations on actual memory performance. This section will first present a brief summary of the findings about JOLs, followed by a discussion about the potential contributions of experience-based versus theory-based processes to item-by-item and aggregate JOLs, along with directions for potential future research. Finally, actual memory performance findings will be discussed.

First, in all the experiments, participants started identifying items more quickly in the fluent (i.e., canonical, upright) than in the disfluent (i.e., noncanonical, inverted) encoding condition, showing objective fluency differences between the encoding conditions. Analyses conducted for the experiments employing item-by-item JOLs revealed a negative correlation between the first-keypress latency and JOLs across encoding conditions, indicating that slower response latencies were associated with lower JOLs, as is shown in Table 3. This is in line with the typical perceptual fluency finding, but the mean correlations were not significantly different for fluent and disfluent items. Second, despite the objective identification difficulties, this difference was not reflected across encoding conditions for item-by-item JOLs (Exps. 1, 3, and 6, except for Exp. 7). Third, even if the objective differences in perception were not reflected in item-by-item JOLs, they were always reflected in aggregate JOLs, regardless of whether this prediction was made without (Exp. 5), before (Exp. 7), or after exposure to the experimental manipulation (Exps. 2, 4a, 4b, and 6). Fourth, the effect sizes for item-by-item JOLs differences across encoding conditions were lower than the effect sizes for aggregate JOLs, even within the same experiment, as can also be seen in Table 2. Finally, when participants made their aggregate JOLs without or before exposure to the material (Exps. 5 and 7), the mean effect size for their aggregate JOLs was numerically higher ($M = 1.31$) than when they made the JOL after exposure to the material ($M = 0.64$ for Exps. 2, 4a, 4b, and 6).

In summary, objective fluency differences across encoding conditions were not always considered to produce JOL changes at the item level. Moreover, aggregate JOLs did not always correlate with item-by-item JOLs. Frank and Kuhlmann (2017) showed that in certain situations, participants may not hold any specific beliefs about the effects of a manipulation (i.e., volume intensity) on memory, but they may still show a bias to predict that more-fluent, high-volume items will be remembered more than less-fluent, low-volume items when the judgments are made on the item level. The present manipulation shows, in contrast, that even when people hold beliefs about a manipulation, these beliefs may not always be applied to specific cases, producing a different type of dissociation between item- and aggregate-level JOLs. In a recent study, Kornell and Hausman (2017) provided evidence that participants sometimes have certain beliefs about their future performance, which they may not always apply to their memory predictions. These dissociations across item-by-item versus aggregate JOLs are important to demonstrate, because they may also provide us with more insight about the extents to which theory-based and experience-based processes are used in making judgments at the item and list levels.

One question that needs to be answered is the nature of the underlying mechanisms that lead to dissociations between aggregate and item-by-item JOLs. One possibility is that all findings about perceptual fluency can be accommodated by theory-based processes. A similar argument has been made by Susser et al. (2016). Susser et al. (2016) reported that when participants are not aware of the objective perceptual fluency differences across conditions, they fail to modify their item-by-item JOLs accordingly (see Besken, 2016, Exp. 3, for a similar finding). However, in Susser et al.'s (2016) identity-priming manipulation, participants' JOLs were modified when they were aware of the fluency differences. In the present manipulation, participants were explicitly aware of the

Table 3 Mean Pearson correlation coefficients and their standard errors (in parentheses) between response latency and item-by-item judgments of learning (JOL) for first-keypress latency and total typing duration for Experiments 1, 3, 6, and 7

Exp #	First keypress latency			Total typing duration		
	Upright / Canonical	Inverted/ Noncanonical	Average	Upright / Canonical	Inverted/ Noncanonical	Average
1	-.20** (.04)	-.25** (.05)	-.25** (.03)	-.10 (.05)	-.09 (.06)	-.09 (.04)
3	-.15** (.04)	-.12* (.05)	-.15** (.03)	-.03 (.04)	.03 (.03)	.01 (.02)
6	-.15** (.04)	-.28** (.05)	-.22** (.04)	-.05 (.05)	-.02 (.05)	-.03 (.04)
7	-.23** (.06)	-.22** (.05)	-.22** (.04)	-.06 (.04)	-.07 (.04)	-.05 (.03)

* Mean correlation is significantly different from 0 at $p < .05$ through a one-sample t test. ** Mean correlation is significantly different from 0 at $p < .01$ through a one-sample t test

manipulation; however, despite the explicit differences across conditions, participants mostly did not modify their item-by-item JOLs. Only in Experiment 7, when the aggregate JOLs were made before the encoding phase, did the pattern of item-by-item JOLs change as well. Thus, when participants are asked about these beliefs before the encoding phase, this might sometimes also affect their judgments at an item level, because participants may pay more attention to the manipulation. However, most of the time, even being aware of the manipulation does not necessarily translate to changes in item-by-item JOLs. For both inversion and canonicity manipulations, items are always intact. When items are intact, orientation and perspective may not be considered as cues that affect memory predictions. Even though both inversion and canonicity manipulations provide cues that are sufficiently discernable to be noticed by participants, such cues may still be subtler than cues in which the image is obstructed with a checkerboard pattern; thus, the manipulation fails to induce a belief on the item level. It also needs to be indicated that theory-based processes do not explain why the effect sizes for aggregate JOLs get smaller when participants are exposed to the experimental material.

A second explanation that dissociates item-by-item and aggregate JOLs is related to experience-based processes, in that participants' objective and subjective online difficulties might moderate JOLs, but the role of experience-based processes in explaining the present set of results appears to have been minimal. Participants' item-by-item JOLs and first-keypress latency correlations showed that slower response rates were typically associated with lower JOLs, but this was not different across encoding conditions, and paired-sample *t* tests failed to show an effect for either inversion or canonicity. This is consistent with the findings of Sungkhasettee et al. (2011). These objective differences are reflected in aggregate JOLs, but since this result is obtained even before exposure to the material, it is difficult to attribute this finding to experience-based processes.

A third possibility is that participants may hold beliefs before the encoding phase that are then modified by the experience-based processes during the experiment, but not sufficiently to affect item-by-item JOLs or to eliminate differences across encoding conditions for aggregate JOLs made at the end of the encoding phase. There seems to be some evidence for this explanation. When aggregate JOLs are made without or before exposure to the experimental material, this depends completely on theory-based processes, since participants have no experience, producing the highest effect sizes across encoding conditions. When participants are exposed to the paradigm, experience-based processes should have more of an effect; yet, despite objective response latency differences, the effect sizes for item-by-item JOL differences across encoding conditions are always smaller than those for aggregate JOL differences. Finally, aggregate JOLs at the end of the

encoding phase, which should be a combination of both theory-based and experience-based processes, yield a significant difference again, but they have lower effect sizes than when the aggregate JOLs that are made without exposure to the procedure.

Perhaps, for fluency to modify beliefs on an online basis, it needs to go over a threshold. For example, with manipulations such as picture fragmentation and auditory generation, the threshold between the identification of two encoding conditions is around 200 ms on average. Yet, with the present manipulations, the identification difference across encoding conditions was much lower, around 111.5 ms on average. Thus, participants might not have been aware that these manipulations produced online objective difficulties, or they might have found the difficulty difference to be minimal, such that they adjusted their expectations on both item-by-item and aggregate bases. The discrepancy in decision-making across item-by-item and aggregate JOLs might be similar to decision-making processes in recognition: Previous research has consistently shown that there are conditions under which fluency is not the only criterion used for making recognition decisions (e.g., Whittlesea & Williams, 2000, 2001a, 2001b). For example, the attribution-discrepancy hypothesis contends that the fluency of an item at recognition per se is not what affects familiarity decisions. Participants evaluate both the expected and experienced fluency of processing at recognition. When there is a discrepancy between expected and experienced fluency, such that the items are perceived more fluently than the participants expected, they adjust their recognition decisions accordingly, and may attribute the increased fluency to familiarity (Whittlesea & Williams, 2000). Perhaps a similar case can be made for JOLs, as well. Participants may expect the rotation manipulations to produce strong fluency differences before the experiment begins, but the subjective difficulties that the participants experience during the procedure are lower than they expect, leading to no differences for item-by-item JOLs and lower effect sizes for aggregate JOLs made at the end of the encoding phase, because experience-based processes modify the belief-based expectations that were present at the beginning of the experiment. Obviously, this theory needs to be tested further. One way to realize it might be to make participants aware of their online processing difficulties or to set up an expectation to manipulate the attribution discrepancy during encoding.

For completeness, we also reported resolution for all experiments that contained item-by-item JOL ratings. In terms of resolution, Goodman–Kruskal gamma correlations revealed that the resolutions across encoding conditions were similar to each other (Exps. 1, 3, 6, and 7) and typically significantly above 0 for one-sample *t* tests (except for Exp. 7). Thus, participants were able to distinguish between items that they would and would not be able to recall above chance level, separately for each encoding condition, and this did not

change, regardless of whether the items were presented in an upright/canonical versus an inverted/noncanonical format. It is reasonable that rotation manipulations would not affect resolution differentially across encoding conditions, because objects' orientations in ordinary life change constantly as we move around in the environment. We are used to seeing objects from various perspectives, even though some points of view may be identified more quickly and efficiently than others, and subsequent recall resolution thus might not be differentially affected by orientation.

A secondary aim of the present study was to investigate whether orientation and perspective may modify actual memory performance. Our results failed to reveal actual free-recall differences across conditions for the inversion manipulation, unlike Sungkhasettee et al. (2011). Obviously, the methodology used in Sungkhasettee et al. differed from ours in that participants were asked to identify the words only by saying them aloud, not by typing them onto the screen. Moreover, the processing of pictures may be different in nature from the processing of words, leading to differences across the retrieval of word and picture material. For example, encoding pictures may require both visual and verbal codes, whereas words are typically encoded verbally (Paivio, 1975, 1986). A similar point of view asserts that pictures may be coded more distinctively and elaborately than simple word material (e.g., Mintzer & Snodgrass, 1999; Nelson, Reed, & Walling, 1976). Pictures may require more lexical coding than words, whereas the meaning does not need to be processed for word stimuli (Nelson et al., 1976). Either way, the difference between two levels of a manipulation such as inversion or canonicity may not lead to as much processing difference for pictures as it does for words. For the canonicity manipulation, we found slightly increased recall performance for noncanonical as compared to canonical items. A cross-investigation of the canonicity experiments (Exps. 3, 4a, 4b, 6, and 7) revealed that the average proportion of recall for noncanonical items is 3.8% higher than that for canonical items, and the average effect size for these five experiments was $d = 0.26$. The difference across encoding conditions sometimes reached significance, but the effect sizes were typically small to medium in size. This increased performance is consistent with effortful processing of items. As a manipulation of fluency, noncanonical images provide a slight advantage over canonical items, in line with predictions based on desirable difficulties (Bjork, 1994; McDaniel & Butler, 2010).

To sum up, the present set of experiments revealed that memory predictions that are made via aggregate and item-by-item JOLs can be dissociated through certain manipulations, such as inversion and canonicity. Moreover, aggregate and item-by-item memory predictions might be influenced to differing extents by a priori beliefs and online subjective difficulties. The present set of experiments showed that it is important to assess participants' a priori beliefs when memory

predictions are being assessed, because online experience-based cues can only explain a portion of perceptual fluency effects.

Author note This work was conducted for partial fulfillment of senior thesis project requirements for E.C.S., M.K., and N.A. This work was partially funded by the Scientific and Technological Research Council of Turkey (Türkiye Bilimsel ve Teknolojik Araştırma Kurumu) Program Code 2209A—Undergraduate Student Research Support, Grant number 1919B0111601407. Portions of this work were presented at the Psychonomic Society's 58th Annual Meeting, the 19th Turkish National Psychology Congress, the 4th International Symposium on Brain and Cognitive Science, and the International Conference on Memory 2016.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

- Begg, I., Vinski, E., Frankovich, L., & Holgate, B. (1991). Generating makes words memorable, but so does effective reading. *Memory & Cognition*, 19, 487–497. <https://doi.org/10.3758/BF03199571>
- Besken, M. (2016). Picture-perfect is not perfect for metamemory: Testing the perceptual fluency hypothesis with degraded images. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 42, 1417–1433. <https://doi.org/10.1037/xlm0000246>
- Besken, M., & Mulligan, N. W. (2013). Easily perceived, easily remembered? Perceptual interference produces a double dissociation between metamemory and memory performance. *Memory & Cognition*, 41, 897–903. <https://doi.org/10.3758/s13421-013-0307-8>
- Besken, M., & Mulligan, N. W. (2014). Perceptual fluency, auditory generation, and metamemory: Analyzing the perceptual fluency hypothesis in the auditory modality. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 40, 429–440. <https://doi.org/10.1037/a0034407>
- Bjork, R. A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe & A. P. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 185–205). Cambridge, MA: MIT Press.
- Brodeur, M. B., Dionne-Dostie, E., Montreuil, T., & Lepage, M. (2010). The Bank of Standardized Stimuli (BOSS), a new set of 480 normative photos of objects to be used as visual stimuli in cognitive research. *PLoS ONE*, 5, 10773. <https://doi.org/10.1371/journal.pone.0010773>
- Cooper, L. A. (1975). Mental rotation of random two-dimensional shapes. *Cognitive Psychology*, 7, 20–43. [https://doi.org/10.1016/0010-0285\(75\)90003-1](https://doi.org/10.1016/0010-0285(75)90003-1)
- Dunlosky, J., & Hertzog, C. (2000). Updating knowledge about encoding strategies: A componential analysis of learning about strategy effectiveness from task experience. *Psychology and Aging*, 15, 462–474. <https://doi.org/10.1037/0882-7974.15.3.462>
- Dunlosky, J., Mueller, M. K., & Tauber, S. K. (2015). The contribution of subjective fluency and theories of memory to people's judgments of memory. In D. S. Lindsay, C. M. Kelley, A. P. Yonelinas, & H. L. Roediger III (Eds.), *Remembering: Attributions, processes, and control in human memory: Papers in honour of Larry L. Jacoby* (pp. 46–63). New York, NY: Psychology Press.
- Frank, D. J., & Kuhlmann, B. G. (2017). More than just beliefs: Experience and beliefs jointly contribute to volume effects on metacognitive judgments. *Journal of Experimental Psychology:*

- Learning, Memory, and Cognition*, 43, 680–693. <https://doi.org/10.1037/xlm0000332>
- Gomez, P., Shutter, J., & Rouder, J. N. (2008). Memory for objects in canonical and noncanonical viewpoints. *Psychonomic Bulletin & Review*, 15, 940–944. <https://doi.org/10.3758/pbr.15.5.940>
- Humphrey, G. K., & Joliceur, P. (1993). An examination of the effects of axis foreshortening, monocular depth cues, and visual field on object identification. *Quarterly Journal of Experimental Psychology*, 46, 137–159. <https://doi.org/10.1080/14640749308401070>
- Joliceur, P. (1985). The time to name disoriented natural objects. *Memory & Cognition*, 13, 289–303. <https://doi.org/10.3758/BF03202498>
- Kelley, C. M., & Jacoby, L. L. (1996). Adult egocentrism: Subjective experience versus analytic bases for judgment. *Journal of Memory and Language*, 35, 157–175. <https://doi.org/10.1006/jmla.1996.0009>
- Kinjo, H., & Snodgrass, J. G. (2000). Does the generation effect occur for pictures? *American Journal of Psychology*, 113, 95–121. <https://doi.org/10.2307/1423462>
- Kinoshita, S. (1989). Generation enhances semantic processing? The role of distinctiveness in the generation effect. *Memory & Cognition*, 17, 563–571. <https://doi.org/10.3758/BF03197079>
- Koriat, A., Bjork, R. A., Sheffer, L., & Bar, S. K. (2004). Predicting one's own forgetting: The role of experience-based and theory-based processes. *Journal of Experimental Psychology: General*, 133, 643–656. <https://doi.org/10.1037/0096-3445.133.4.643>
- Koriat, A., & Norman, J. (1984). What is rotated in mental rotation? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 10, 421–434. <https://doi.org/10.1037/0278-7393.10.3.421>
- Kornell, N., & Hausman, H. (2017). Performance bias: Why judgments of learning are not affected by learning. *Memory & Cognition*, 45, 1270–1280. <https://doi.org/10.3758/s13421-017-0740-1>
- Kornell, N., Rhodes, M. G., Castel, A. D., & Tauber, S. K. (2011). The ease-of-processing heuristic and the stability bias dissociating memory, memory beliefs, and memory judgments. *Psychological Science*, 22, 787–794. <https://doi.org/10.1177/0956797611407929>
- Kovalenko, L. Y., Chaumon, M., & Busch, N. A. (2012). A pool of pairs of related objects (POPORO) for investigating visual semantic integration: Behavioral and electrophysiological validation. *Brain Topography*, 25, 272–284. <https://doi.org/10.1007/s10548-011-0216-8>
- Lawson, R., & Humphreys, G. W. (1998). View-specific effects of depth rotation and foreshortening on the initial recognition and priming of familiar objects. *Attention, Perception, & Psychophysics*, 60, 1052–1066. <https://doi.org/10.3758/bf03211939>
- Lawson, R., Humphreys, G. W., & Joliceur, P. (2000). The combined effects of plane disorientation and foreshortening on picture naming: One manipulation or two? *Journal of Experimental Psychology: Human Perception and Performance*, 26(2):568–581.
- McDaniel, M. A., & Butler, A. C. (2010). A contextual framework for understanding when difficulties are desirable. In A. S. Benjamin (Ed.), *Successful remembering and successful forgetting: Essays in honor of Robert A. Bjork* (pp. 175–199). New York, NY: Psychology Press.
- Mintzer, M. Z., & Snodgrass, J. G. (1999). The picture superiority effect: Support for the distinctiveness model. *American Journal of Psychology*, 112, 113–146. <https://doi.org/10.2307/1423627>
- Mitchum, A. L., Kelley, C. M., & Fox, M. C. (2016). When asking the question changes the ultimate answer: Metamemory judgments change memory. *Journal of Experimental Psychology: General*, 145, 200–219. <https://doi.org/10.1016/j.jml.2016.10.008>
- Mueller, M. L., & Dunlosky, J. (2017). How beliefs can impact judgments of learning: Evaluating analytic processing theory with beliefs about fluency. *Journal of Memory and Language*, 93, 245–258. <https://doi.org/10.1016/j.jml.2016.10.008>
- Mueller, M. L., Dunlosky, J., Tauber, S. K., & Rhodes, M. G. (2014). The font-size effect on judgments of learning: Does it exemplify fluency effects or reflect people's beliefs about memory? *Journal of Memory and Language*, 70, 1–12. <https://doi.org/10.1016/j.jml.2013.09.007>
- Mulligan, N. W. (2002). The emergent generation effect and hypermnesia: Influences of semantic and nonsemantic generation tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28, 541–554. <https://doi.org/10.1037/0278-7393.28.3.541>
- Naime, J. S., & Widner, R. L. (1988). Familiarity and lexicality as determinants of the generation effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 14, 694–699. <https://doi.org/10.1037/0278-7393.14.4.694>
- Nelson, D. L., Reed, V. S., & Walling, J. R. (1976). Pictorial superiority effect. *Journal of Experimental Psychology: Human Learning and Memory*, 25, 523–528. <https://doi.org/10.1037/0278-7393.2.5.523>
- Newell, F. N., & Findlay, J. M. (1997). The effect of depth rotation on object identification. *Perception*, 26, 1231–1257. <https://doi.org/10.1068/p261231>
- Paivio, A. (1975). Perceptual comparisons through the mind's eye. *Memory & Cognition*, 3, 635–647. <https://doi.org/10.3758/BF03198229>
- Paivio, A. (1986). *Mental representation: A dual coding approach*. Hillsdale, NJ: Erlbaum.
- Rhodes, M. G. (2016). Judgments of learning: Methods, data, and theory. In J. Dunlosky & S. K. Tauber (Eds.), *The Oxford handbook of metamemory* (pp. 65–80). New York, NY: Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199336746.013.4>
- Rhodes, M. G., & Castel, A. D. (2008). Memory predictions are influenced by perceptual information: Evidence for metacognitive illusions. *Journal of Experimental Psychology: General*, 137, 615–625. <https://doi.org/10.1037/a0013684>
- Rhodes, M. G., & Castel, A. D. (2009). Metacognitive illusions for auditory information: Effects on monitoring and control. *Psychonomic Bulletin & Review*, 16, 550–554. <https://doi.org/10.3758/PBR.16.3.550>
- Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science*, 171, 701–703. <https://doi.org/10.1126/science.171.3972.701>
- Slamecka, N. J., & Graf, P. (1978). The generation effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Human Learning and Memory*, 4, 592–604. <https://doi.org/10.1037/0278-7393.4.6.592>
- Soderstrom, N. C., Clark, C. T., Halamish, V., & Bjork, E. L. (2015). Judgments of learning as memory modifiers. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 41, 553–558. <https://doi.org/10.1037/a0038388>
- Sungkhasettee, V. W., Friedman, M. C., & Castel, A. D. (2011). Memory and metamemory for inverted words: Illusions of competency and desirable difficulties. *Psychonomic Bulletin & Review*, 18, 973–978. <https://doi.org/10.3758/s13423-011-0114-9>
- Susser, J. A., Jin, A., & Mulligan, N. W. (2016). Identity priming consistently affects perceptual fluency but only affects metamemory when primes are obvious. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 42, 657–662. <https://doi.org/10.1037/xlm0000189>
- Susser, J. A., & Mulligan, N. W. (2015). The effect of motoric fluency on metamemory. *Psychonomic Bulletin & Review*, 22, 1014–1019. <https://doi.org/10.3758/s13423-014-0768-1>
- Susser, J. A., Mulligan, N. W., & Besken, M. (2013). The effects of list composition and perceptual fluency on judgments of learning (JOLs). *Memory & Cognition*, 41, 1000–1011. <https://doi.org/10.3758/s13421-013-0323-8>
- Undorf, M., Zimdahl, M. F., & Bernstein, D. M. (2017). Perceptual fluency contributes to effects of stimulus size on judgments of learning.

- Journal of Memory and Language*, 92, 293–304. <https://doi.org/10.1016/j.jml.2016.07.003>
- Whittlesea, B. W., & Williams, L. D. (2000). The source of feelings of familiarity: The discrepancy-attribution hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26, 547–565. <https://doi.org/10.1037/0278-7393.26.3.547>
- Whittlesea, B. W. A., & Williams, L. D. (2001a). The discrepancy-attribution hypothesis: I. The heuristic basis of feelings and familiarity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 27, 3–13. <https://doi.org/10.1037/0278-7393.27.1.3>
- Whittlesea, B. W. A., & Williams, L. D. (2001b). The discrepancy-attribution hypothesis: II. Expectation, uncertainty, surprise, and feelings of familiarity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 27, 14–33. <https://doi.org/10.1037/0278-7393.27.1.14>
- Yang, C., Huang, T. S. T., & Shanks, D. R. (2018). Perceptual fluency affects judgments of learning: The font size effect. *Journal of Memory and Language*, 99, 99–110. <https://doi.org/10.1016/j.jml.2017.11.005>
- Yue, C. L., Castel, A. D., & Bjork, R. A. (2013). When disfluency is—and is not—a desirable difficulty: The influence of typeface clarity on metacognitive judgments and memory. *Memory & Cognition*, 41, 229–241. <https://doi.org/10.3758/s13421-012-0255-8>