*Psychonomic Bulletin & Review*
*2007, 14 (4), 620-628*

# A visual short-term memory advantage for faces

**Kim M. Curby**
*Temple University, Philadelphia, Pennsylvania*

**and**

**Isabel Gauthier**
*Vanderbilt University, Nashville, Tennessee*

What determines how much can be stored in visual short-term memory (VSTM)? Studies of VSTM have focused largely on stimulus-based properties such as the number or complexity of the items stored. Recent work also suggests that capacity is severely reduced for items within the same category. However, the importance for VSTM capacity of more qualitative differences in processing for different categories has not been investigated. For example, faces are processed more holistically than other objects. In Experiments 1 and 2, we show that the processing of faces, objects that are crucial socially and for which we possess considerable expertise, overcomes these limitations. More faces can be stored in VSTM than objects from other complex nonface categories. As in prior studies, at short encoding durations we found that capacity for faces was less than that for other categories. However, at longer encoding durations, capacity for faces exceeded that for nonface objects, and this advantage was specific to upright faces. Because inversion reduces holistic processing, the interaction of orientation with VSTM capacity—which occurred for faces but not objects in Experiment 3—suggests that it is holistic processing that confers an advantage for face VSTM when sufficient encoding time is allowed.

Shelves are stocked with books offering advice on improving one's memory. Even a mundane task like attaching new electronic equipment to a television forces us to look continuously back and forth between the diagram and installation, revealing the limited amount of visual information we can keep in memory at any one time. Surprisingly, what constrains visual short-term memory (VSTM) is relatively unknown. Some suggest that VSTM has a fixed number of "slots," each capable of temporarily storing one object (Vogel, Woodman, & Luck, 2001). Others argue that VSTM capacity is influenced by the complexity of the items stored (Alvarez & Cavanagh, 2004). But is VSTM fixed solely by stimulus factors such as perceptual complexity or object number, or can it be influenced by the processing strategy used to encode objects?

VSTM capacity is typically estimated at three to four objects, but observers can retain information about many more features distributed across four objects (Vogel et al., 2001). Accordingly, VSTM has been argued to be object- rather than feature-based, its capacity indifferent to the number of features per object.

Recent studies question strong versions of the object-based theory of VSTM capacity (Eng, Chen, & Jiang, 2006; Wheeler & Treisman, 2002). For example, VSTM capacity appears to decrease as stimulus complexity (or information load) increases, so that people can maintain more colored squares than complex line drawings in VSTM (Alvarez & Cavanagh, 2004). Information load was operationalized as the rate of visual search among items of a category. A strong positive correlation between capacity and search rate was found, suggesting that objects of greater complexity require more "space" in VSTM. This relationship is particularly strong when perceptual encoding time is limited (Eng et al., 2006), as complex objects require more time to be encoded into VSTM than do simpler objects like colored squares: The relationship between complexity and capacity is reduced at longer encoding durations when perceptual processing is no longer a bottleneck.

Despite object-based theories suggesting hard-wired VSTM capacity limits, there is evidence that VSTM for faces may be affected by processing strategy. Faces are processed more holistically than objects or inverted faces (Tanaka & Sengco, 1997). Holistic processing results in the individual features and the relations between these features being relatively inseparable (Farah, Wilson, Drain, & Tanaka, 1997). Notably, visual search rate is faster for upright than inverted faces (Tong & Nakayama, 1999). This suggests that VSTM capacity may be higher for upright than inverted faces despite their containing the same low-level visual information and thus being equivalent in stimulus-based complexity. However, if the search rate difference between upright and inverted faces reflects a difference analogous to that reported by Alvarez and Cavanagh (2004), it should predominantly affect VSTM when encoding time is limited. However, we postulated that holistic processing confers an advantage to upright

**K. M. Curby, kim.curby@temple.edu**

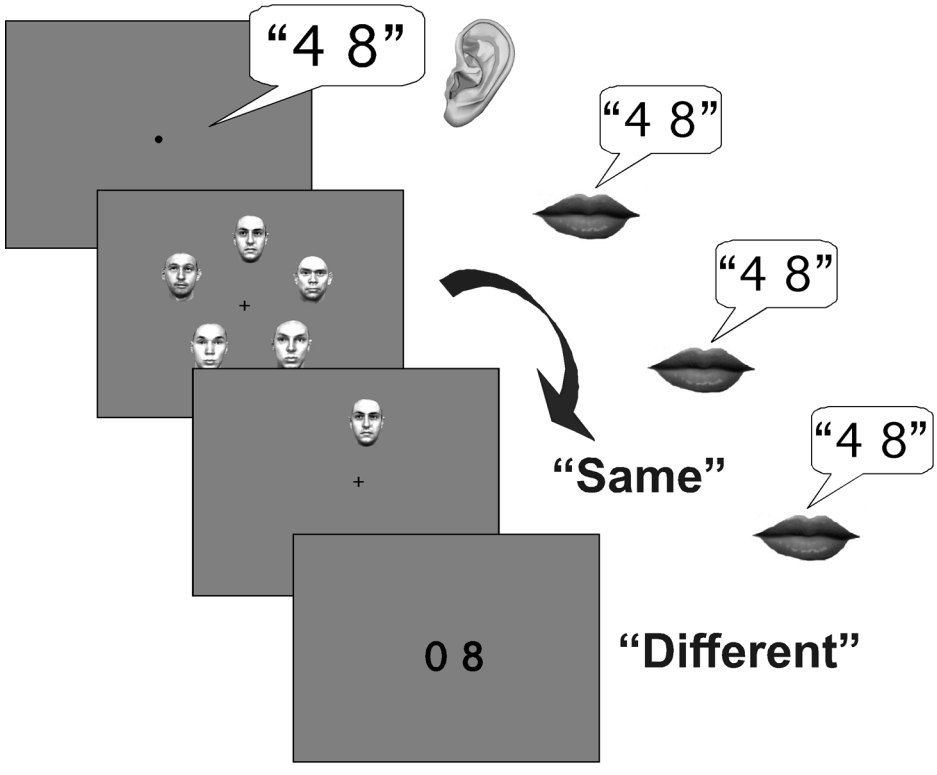


**Figure 1. The sequence of events that occurred in each trial: Participants first were presented with an auditory stimulus that consisted of two digits and a mask, which they overtly rehearsed throughout the trial to prevent verbal rehearsal. The study array, consisting of 1–5 faces evenly spaced in a circle (6.1º diameter) (either all upright or all inverted), then appeared for 500, 1,200, or 2,500 msec. After a 1,200-msec delay a face-probe was presented in one of the locations from the study array. The probe remained until participants indicated with a keypress whether the face was the same as (50% of trials) or different from the one that appeared in that location in the study array. To minimize confusion, within each trial, the probe was never a face that had appeared at a different location in the study array. After a response was made, a screen with two digits appeared and participants were required to state whether the two digits on the screen were the same as those they had been rehearsing throughout the trial.**

faces that extends beyond the time required to encode the faces, leading them to be represented and stored more efficiently in VSTM than inverted faces even at long encoding times. Experiment 1 tests this prediction.

## EXPERIMENT 1

### Method

**Participants**. Twenty-four (18 female) individuals (mean age = 24.3, *SE* = 3.86) participated for payment.

**Stimuli and Procedure**. Seventy-two grayscale faces (1.9º × 2.3º) from the Max-Planck Institute were used. Each face either appeared upright or inverted. Participants performed a delayed match-to-sample probe recognition task (see Figure 1) involving 900 trials across three sessions, each consisting of 10 alternating blocks (30 randomized trials/block) of either upright or inverted faces. There were 450 trials with upright faces and 450 with inverted faces. For each orientation, there were 15 conditions (5 set sizes × 3 encoding durations), each including 30 trials.

**Analysis**. One participant's data were removed due to poor performance in the auditory task (all others exceeded 95% accuracy). Incorrect articulatory suppression trials (<2%) were removed from further analyses. For each participant and condition, Cowan's K was calculated providing an estimate of the number of objects successfully encoded in VSTM (Cowan, 2001).[1] The maximum K (K-max) was identified for each encoding duration, regardless of set size. In addition to analyses on the standard K measure, analyses on the K-max measure were performed because some participants tend to show a drop in performance in conditions where the set size exceeds their capacity, especially for faces, thus it is possible that K-max may better capture VSTM capacity for larger set size conditions.

### Results

Increased encoding time led to a greater increase in VSTM capacity for upright than inverted faces (Figure 2, Table 1). With 500-msec encoding time, orientation had little effect on VSTM capacity, but longer encoding times (1,500 msec or 2,500 msec) led to greater VSTM capacity for upright than for inverted faces. Thus, inverted faces benefited less than upright faces from additional encoding time.

A 2 (upright vs. inverted faces) × 5 (set sizes 1 5) × 3 (500- vs. 1,500- vs. 2,500-msec encoding time) ANOVA revealed main effects of orientation and duration, with overall K greater for upright than inverted faces [$F(1,23) = 29.38$, $p \leq .0001$], and for longer encoding

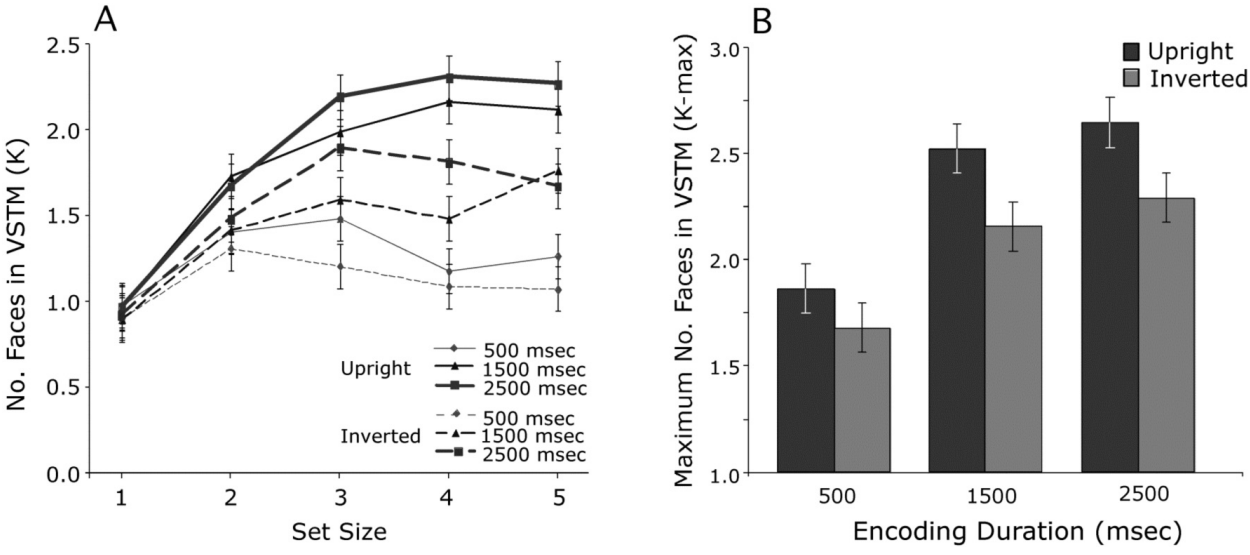


**Figure 2. (A) Estimated number of objects stored in visual short-term memory (VSTM) using Cowan's (2001) formula for 500-, 1,500-, and 2,500-msec encoding time for upright and inverted faces at each of the different set sizes, and (B) the maximum VSTM capacity (K-max) for each category and encoding duration. K and K-max values for upright and inverted faces were similar at the shortest encoding time, and increased with additional encoding time for both upright and inverted faces. However, K and K-max values for upright faces were greater than that for inverted faces when ample encoding time was allowed, suggesting that affects of processing mode on VSTM capacity emerge with sufficient encoding time.**

times [$F(2,46) = 83.43$, $p \leq .0001$]. Most importantly, orientation and encoding duration interacted, with the effect of orientation greater for longer durations [$F(2,46) = 4.35$, $p = .019$].[2]

A supplementary ANOVA on the maximum VSTM capacity (K-max) across the different set sizes produced similar results, with main effects of orientation [$F(1,23) = 15.26$, $p = .0007$], and duration [$F(2,46) = 30.76$, $p \leq .0001$]. Although the interaction between orientation and duration was not significant ($F < 1$, n.s.), paired $t$ tests revealed differences in K-max between upright and inverted faces with 1,500-msec [$t(23) = 3.095$, $p = .0051$] and 2,500-msec [$t(23) = 2.59$, $p = .016$] encoding time, but not with 500 msec [$t(23) = 1.66$, $p = .11$].

### Discussion

VSTM capacity was greater for upright than inverted faces, provided there was sufficient encoding time. It is possible that performance was constrained by a floor effect at 500 msec; however, previous studies have reported lower VSTM capacities for face and nonface objects, suggesting that lower performance is possible (Eng et al., 2006; Olsson & Poom, 2005). Importantly, a floor effect in this condition would not affect our main finding of an advantage for faces with extended encoding time.

This advantage for upright faces likely results from holistic processing mechanisms unavailable or less efficient for inverted faces, but could also reflect more general differences between upright and inverted stimuli (costs for unfamiliar orientations, Lawson & Jolicœur, 1998). If so, our effects should not extend to a comparison between upright faces and upright nonholistically processed objects (e.g., cars or watches). Experiment 2 addresses this possibility.

## EXPERIMENT 2

### Method

**Participants**. Twenty-one (11 female) individuals (mean age = 26.3, $SE = 5.2$) participated for payment.

**Stimuli and Procedure**. Seventy-two grayscale images each of upright faces (1.9º × 2.3º), watches (1.9º × 2.3º), and cars (2.3º × 1.5º) (Figure 3). The faces were the same as in Experiment 1. The watch images all depicted front-on upright views. The viewpoint of the cars varied from three-quarter to side views across images.

The procedure and analysis was similar to Experiment 1.[3] Participants performed 420 trials for each stimulus category in three separate sessions (sessions were counterbalanced across subjects). For each category, the 28 trials of each of the 15 conditions were randomized.

### Results

As in Experiment 1, additional encoding time increased participants' performance at identifying an object from the study array. Also, the relative VSTM capacity between categories again depended on encoding time. At short encoding times (i.e., 500 msec), capacity for faces was less than that for other categories; however, with 2,500-msec encoding time there was no difference between face and object VSTM capacity (Figures 4 and 5, Table 1).

An ANOVA revealed main effects of encoding duration and category, with K greater for longer encoding durations [$F(2,40) = 128.76$, $p \leq .0001$], and for watches and cars compared with faces [$F(2,40) = 12.06$, $p \leq .0001$]. Most importantly, category interacted with duration, with the effect of presentation duration larger for faces than watches or cars [$F(4,80) = 6.52$, $p \leq .0001$].[4]

A supplementary ANOVA on the K-max values also revealed main effects of category [$F(2,40) = 5.13$, $p = .01$], and encoding duration [$F(2,40) = 57.89$, $p \leq .0001$]. The interaction between category and duration approached

**Table 1**
**Mean Percent Correct for Each Condition in Each of the Three Experiments**

| Category | Duration (msec) | Set Size 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|
| | | Experiment 1 | | | | |
| Upright faces | 500 | 98.7 | 85.0 | 74.6 | 64.8 | 62.7 |
| | 1,500 | 97.7 | 93.1 | 82.9 | 76.9 | 71.1 |
| | 2,500 | 98.0 | 91.6 | 86.6 | 78.8 | 72.5 |
| Inverted faces | 500 | 94.9 | 82.7 | 69.8 | 63.6 | 60.6 |
| | 1,500 | 94.4 | 85.2 | 76.5 | 68.4 | 67.5 |
| | 2,500 | 95.7 | 86.9 | 81.4 | 72.7 | 66.8 |
| | | Experiment 2 | | | | |
| Faces | 500 | 96.9 | 84.5 | 70.8 | 66.8 | 60.1 |
| | 1,500 | 96.4 | 92.5 | 86.3 | 76.2 | 67.7 |
| | 2,500 | 97.2 | 92.0 | 85.5 | 81.8 | 75.4 |
| Watches | 500 | 96.9 | 87.0 | 80.1 | 74.4 | 66.0 |
| | 1,500 | 98.8 | 92.4 | 84.7 | 78.7 | 75.1 |
| | 2,500 | 98.6 | 91.9 | 86.9 | 80.7 | 73.7 |
| Cars | 500 | 98.3 | 92.2 | 81.1 | 71.7 | 70.1 |
| | 1,500 | 98.8 | 92.4 | 84.7 | 78.7 | 75.1 |
| | 2,500 | 99.1 | 94.6 | 87.4 | 82.5 | 76.4 |
| | | Experiment 3 | | | | |
| Upright faces | 500 | 96.0 | | 71.0 | | 62.6 |
| | 2,500 | 97.9 | | 86.4 | | 75.3 |
| | 4,000 | 97.8 | | 89.7 | | 77.0 |
| Inverted faces | 500 | 92.0 | | 68.9 | | 62.6 |
| | 2,500 | 94.7 | | 75.3 | | 68.7 |
| | 4,000 | 94.9 | | 81.2 | | 69.5 |
| Upright cars | 500 | 94.0 | | 75.8 | | 64.2 |
| | 2,500 | 95.3 | | 80.8 | | 73.0 |
| | 4,000 | 96.4 | | 82.6 | | 71.8 |
| Inverted cars | 500 | 94.3 | | 71.5 | | 63.4 |
| | 2,500 | 94.2 | | 77.7 | | 69.5 |
| | 4,000 | 94.2 | | 79.0 | | 71.8 |

significance [$F(4,80) = 2.14$, $p = .084$]. There appeared to be an overall advantage for cars, perhaps reflecting participants' use of the variability in viewpoint to facilitate VSTM performance. When cars were removed from the analysis the interaction between duration and category reached significance [$F(2,40) = 3.76$, $p = .032$].[5] Paired $t$ tests revealed differences in K-max between faces and cars [$t(20) = 4.053$, $p = .0006$], and faces and watches [$t(20) = 3.27$, $p = .0038$], with 500-msec but not with 1,500-msec or 2,500-msec encoding time (all $p$s $> .09$).

### Discussion

Participants experienced a greater increase in VSTM capacity with additional encoding time for faces than objects. This suggests that the results from Experiment 1 did not depend on differences between canonical and noncanonical stimulus views. However, inconsistent with our predictions, VSTM for faces did not exceed that for objects at the longest encoding time. Possible explanations are discussed later and explored in Experiment 3.

The smaller VSTM capacity for faces compared to objects at the shortest encoding duration (500 msec) likely reflects a difference in perceptual complexity or within-category homogeneity. Indeed, the absence of this difference between upright and inverted faces at short encoding times suggests this effect is not related to processing strategy. Faces may place a greater burden on encoding mechanisms compared to watches or cars, consistent with the slower search rates for faces compared to other complex object categories (Eng et al., 2006).

Specific measures of object complexity, such as visual search rate, were not obtained for the face and nonface objects. Not only is object complexity a somewhat elusive construct to define (see Donderi, 2006), but it is also unclear how useful such indexes would be, given that they appear to be predictive only at limited encoding durations (Eng et al., 2006). Importantly, the interaction between encoding duration and stimulus category does not depend on specific differences in complexity or information load between categories.

Our results also suggest that additional encoding time can partly compensate for differences in complexity (Eng et al., 2006). Despite the smaller VSTM capacity for faces compared to cars or watches at short encoding durations, with sufficient encoding time capacity for faces approximates that of the other categories. The influence of encoding time may depend on processing strategy, with greater benefits for more holistically processed faces.

Based on the results of Experiment 1, we expected VSTM capacity for upright faces to exceed that of objects at the longest encoding time (2,500 msec). However, the lower VSTM capacity for faces than objects at short encoding times suggests that VSTM for faces is handicapped by factors influencing information load (e.g., complexity or homogeneity) and this effect may be too strong to be fully offset by a holistic strategy. Alternatively, 2,500 msec may be insufficient to eliminate perceptual encoding limitations for faces (Eng et al., 2006). Experiment 3 addresses these possibilities.

## EXPERIMENT 3

Experiment 3 examines VSTM capacity under conditions known to interact with processing strategy. The inversion of objects does not produce the same qualitative effects documented for faces (Yin, 1969), therefore if the larger capacity for upright compared to inverted faces is due to a difference in holistic processing, inversion of cars should not have the same influence on VSTM. We also explore whether capacity for inverted faces reaches that for upright faces with additional encoding time: this would be expected if VSTM capacity is entirely determined by complexity, whereas it is not predicted if holistic processing is important. Finally, we predict greater VSTM capacity for upright faces than upright cars when encoding time exceeds 2,500 msec.

### Method

**Participants**. Twenty-seven (17 female) individuals (mean age = 20.2, $SE = 2.2$) participated for payment.

**Stimuli and Procedure**. The faces were the same as in Experiments 1 and 2. A different set of 72 images of car profiles was used. The procedure was similar to Experiment 1 except that the study array was presented for 500, 2,500, or 4,000 msec and only in set

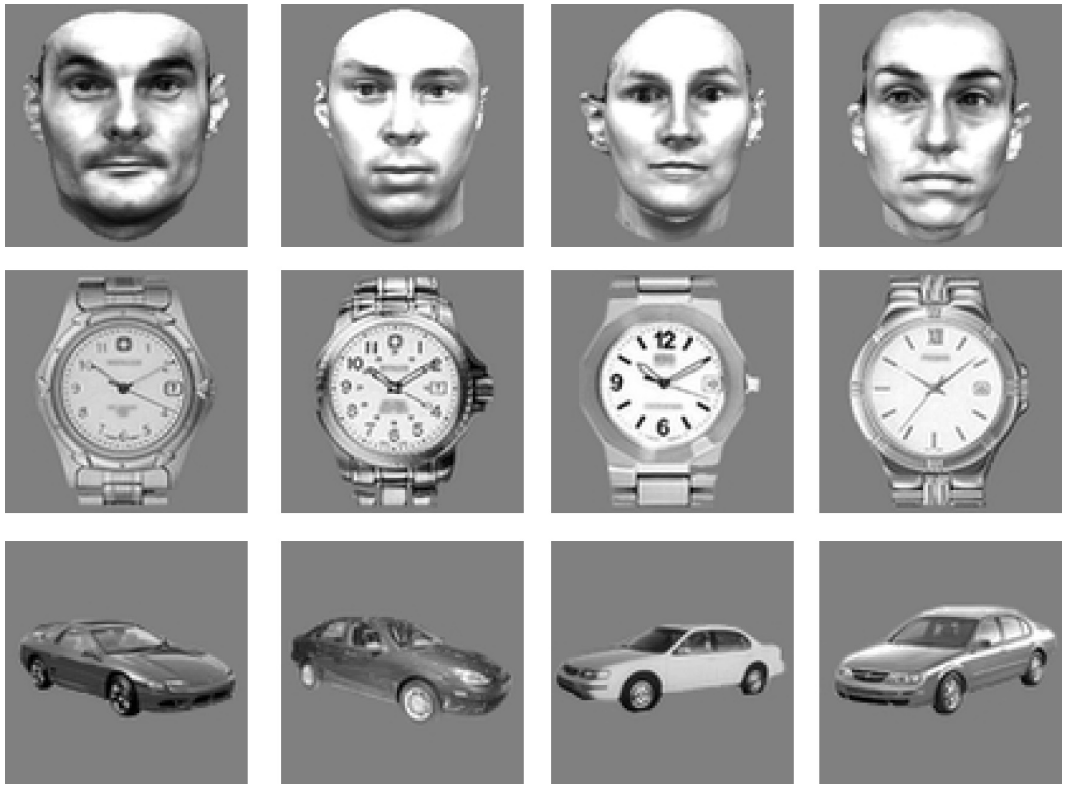

**Figure 3. Examples of the stimuli presented in Experiment 3.**

sizes 1, 3, and 5. Participants performed 1,152 trials across four sessions (counterbalanced across participants), each consisting of 8 alternating blocks (36 randomized trials/block) of either upright or inverted images. Two sessions consisted of either upright or inverted faces; the remaining sessions showed upright or inverted cars. There were 288 trials for each of the four stimulus categories, within which there were 9 conditions (3 set sizes × 3 encoding durations) presented 32 times.[6]

## Results

Increased encoding time led to a greater increase in VSTM capacity for upright faces compared to upright cars and inverted faces and cars (Figures 6 and 7, Table 1). With 500-msec encoding time, orientation had little effect on VSTM capacity, but longer encoding time (4,000 msec) led to greater capacity for upright faces compared to in-

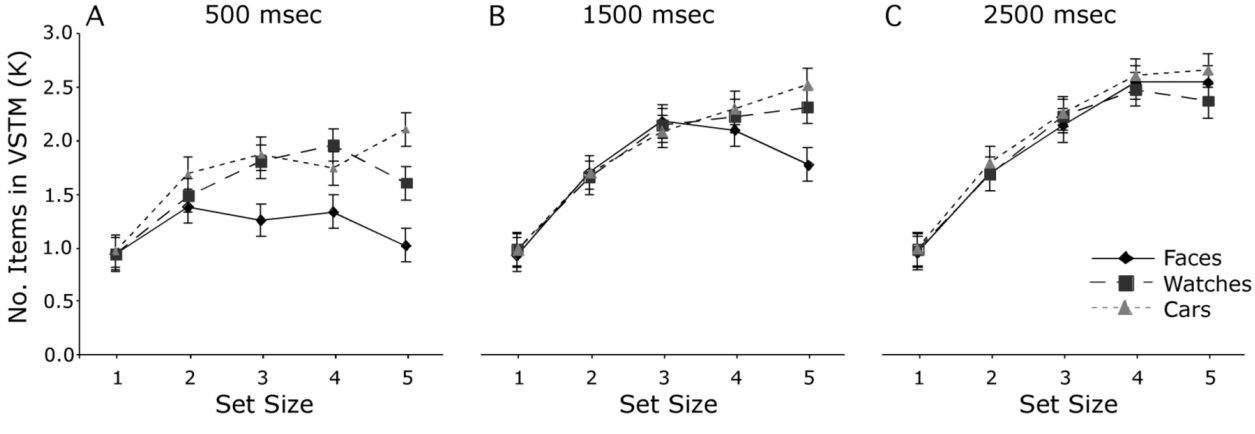


**Figure 4. Estimated number of objects in visual short-term memory (VSTM) using Cowan's (2001) formula for (A) 500-msec, (B) 1,500-msec, and (C) 2,500-msec encoding time for each of the different set sizes and stimulus categories in Experiment 2. K increases with additional encoding time for all three categories of stimuli, and is similar across the three categories when ample encoding time is allowed. However, faces have a lower K in the shorter encoding time conditions compared to the object categories, suggesting that they may place a greater burden on encoding mechanisms than watches and cars.**

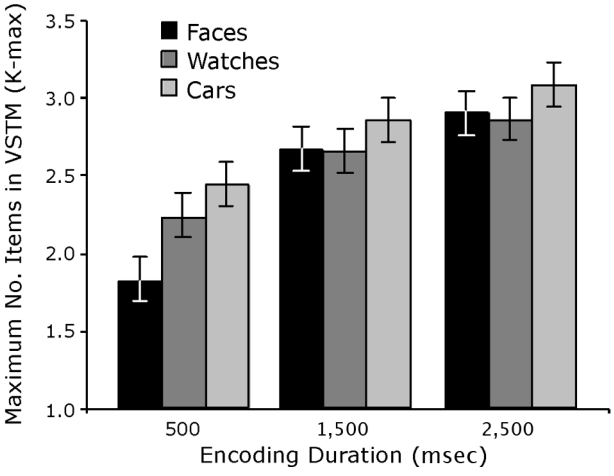


**Figure 5. The maximum number of objects (K-max) in visual short-term memory (VSTM) for 500-msec, 1,500-msec, and 2,500-msec encoding time for faces, watches, and cars. K-max increased with encoding time for each of the three categories, however the benefit of additional encoding time was greatest for faces.**

verted faces, upright cars, or inverted cars. Therefore, capacity for upright faces benefited more from additional encoding time than did that for inverted faces, or cars irrespective of their orientation.

A 2 (upright vs. inverted) × 2 (cars, faces) × 3 (set sizes 1, 3, 5) × 3 (500- vs. 2,500- vs. 4,000-msec encoding time) ANOVA revealed main effects of encoding duration and orientation but not category ($t < 1$), with K greater for longer encoding times [$F(2,52) = 81.78$, $p \leq .0001$], and for upright compared to inverted items [$F(1,26) = 42.10$, $p \leq .0001$]. Category interacted with both orientation [$F(1,26) = 11.58$, $p = .0022$], and encoding duration [$F(2,52) = 7.77$, $p = .0011$], with the effect of orientation and duration on VSTM greater for faces than cars. In addition, an interaction between orientation and encoding duration [$F(2,52) = 5.75$, $p = .0056$], revealed that the effect of orientation was greater for longer encoding durations. Most importantly, there was a three-way interaction between orientation, category and encoding duration, confirming a greater influence of orientation for faces than cars, especially at longer encoding durations [$F(2,52) = 4.56$, $p = .015$].[7]

A supplementary ANOVA on the K-max values mirrored the results reported above, with main effects of orientation [$F(1,26) = 47.04$, $p \leq .0001$], and duration [$F(2,52) = 56.65$, $p \leq .0001$], but not category ($F < 1$). There was also an interaction between orientation and category [$F(1,26) = 7.78$, $p = .0098$], orientation and encoding duration [$F(2,52) = 3.29$, $p = .045$], and category and encoding duration [$F(2,52) = 3.48$, $p = .038$]. Most importantly, there was again a three-way interaction between orientation, category and encoding duration [$F(2,52) = 3.29$, $p = .045$].

Consistent with the ANOVA results, paired $t$ tests revealed that K-max was reduced by inversion for faces [$t(26) = 8.52$, $p \leq .0001$], but not cars [$t(26) = 1.67$, $p = .11$]. In addition, K-max for upright faces was greater than for upright [$t(26) = 2.44$, $p = .022$], and inverted cars [$t(26) = 3.62$, $p = .0012$]. Paired $t$ tests on the different encoding durations revealed strong effects of inversion on K-max for faces with 2,500-msec [$t(26) = 4.473$, $p \leq .0001$] or 4,000-msec [$t(26) = 5.8$, $p \leq .0001$] but not with 500-msec [$t(26) = 1.66$, $p = .11$] encoding time. In contrast, the influence of inversion on K-max for cars was less consistent; there was no influence of inversion in the 4,000-msec ($t < 1$) and 500-msec encoding conditions [$t(26) = 1.10$, $p = .28$], although it just reached significance in the 2,500-msec encoding condition [$t(26) = 2.107$, $p = .045$]. Some degree of an inversion cost for

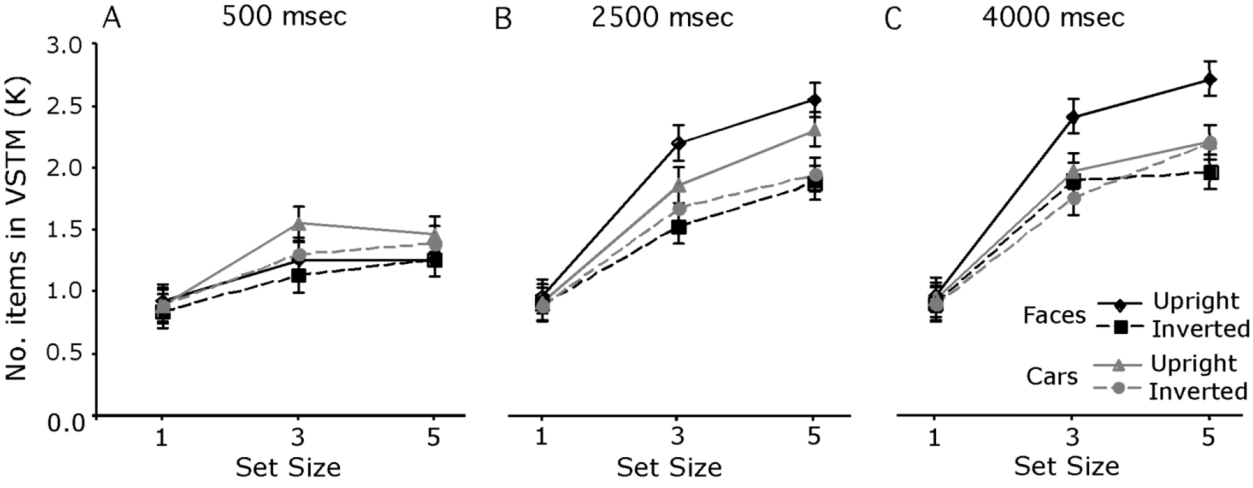


**Figure 6. Estimated number of objects in visual short-term memory (VSTM) using Cowan's (2001) formula for (A) 500-msec, (B) 2,500-msec, and (C) 4,000-msec encoding time for each of the different set sizes and stimulus categories. K increases with additional encoding time for all four categories of stimuli, however the benefit of additional encoding time was greatest for faces. The effect of inversion on VSTM capacity was larger for faces than cars.**

cars is not surprising as the inversion effect for faces is specifically defined as a larger cost due to inversion relative to that for nonface objects (Yin, 1969). Notably, the inversion cost was significantly greater for faces than cars in the 2,500-msec condition ($p = .0101$).

### Discussion

Consistent with the proposed influence of holistic processing on VSTM capacity, with sufficient encoding time, capacity for upright faces exceeded that for inverted faces, upright cars, and inverted cars. Cars showed a smaller effect of orientation with 2,500-msec encoding, which disappeared with additional encoding time. This is likely due to a familiarity with these mono-oriented objects (Yin, 1969), but it does not account for the larger inversion effect obtained for faces. In addition, capacity for inverted faces appears fundamentally limited compared to that for upright faces, independent of encoding duration.

## GENERAL DISCUSSION

Different categories of objects can recruit different processing styles through either experience or innate biases. A holistic strategy for face processing confers an advantage, making face recognition less susceptible to irrelevant feature changes (e.g., hairstyle) and increasing sensitivity to subtle configural differences (Tanaka & Sengco, 1997). Our results suggest that holistic processing also influences VSTM. This cannot be explained by differences in stimulus-based complexity or by a general advantage for objects in familiar orientations. The use of unfamiliar faces and of articulatory suppression also renders alternative explanations involving a contribution from verbal short-term memory unlikely. Long-term memory is also unlikely to explain these results: not only were the faces unfamiliar, but long-term memory traces created in the context of a VSTM task have limited impact on capacity (Chen, Eng, & Jiang, 2006).

One possibility is that the VSTM advantage for faces arises from a difference in eye movements. For instance, a VSTM advantage for faces could result from a greater tendency to fixate upright faces, particularly at longer encoding times. However, this is unlikely because the VSTM advantage for faces over nonface objects persisted even in a further control experiment where items were presented sequentially and each item was fixated for the same duration.[8]

It may seem that more direct manipulations of holistic processing, such as dividing a face into small spatially separated pieces to disrupt this processing strategy (Farah, Tanaka, & Drain, 1995), could better address the influence of holistic processing on VSTM capacity. However, such manipulations have consequences not only for holistic processing, but also for VSTM more generally; VSTM capacity is facilitated by the organization of features into objects, e.g., through proximity and connectedness (Xu, 2006). Inversion robustly influences the degree of holistic processing of faces without disrupting VSTM more generally and is therefore a more controlled manipulation of holistic processing.

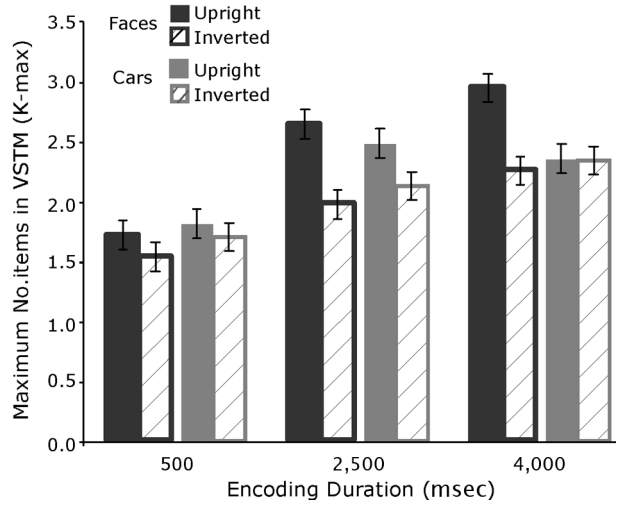


**Figure 7. The maximum number of objects (K-max) in visual short-term memory with 500-, 2,500-, and 4,000-msec encoding time for each of the stimulus categories. When given 4000 msec encoding time capacity for upright faces exceeded that for upright or inverted cars. The influence of inversion on VSTM was greater for faces than cars. In addition, controlling viewpoint eliminated the advantage for cars found in Experiment 2.**

A recent study proposed that VSTM capacity for intracategory objects that cannot be easily labeled is only one object (Olsson & Poom, 2005). The authors argued that estimates of VSTM capacity can be inflated when observers assign verbal labels to objects that cross category boundaries. However, we find VSTM capacity greater than two items for cars and watches, even reaching three items for faces, despite the fact that our stimuli were intracategorical and did not have obvious labels, at least for the watches and faces. The car labels may have been more familiar to experts, although it is unlikely that our novice subjects could label most of them. Our stimuli were also more complex, which should have reduced VSTM capacity (Alvarez & Cavanagh, 2004). Indeed, one factor not considered by Olsson & Poom (2005) is that intracategory objects may require more encoding time than objects that cross category boundaries. Our results and those of Eng et al. (2006) suggest that VSTM capacity for complex objects is underestimated at 500 msec of encoding time because of perceptual encoding limitations. Perhaps with enough encoding time, capacity for the geometrical objects used by Olsson & Poom (2005) would have met that which we found for watches, for instance.

Holistic processing mechanisms, generally linked with visual expertise (Gauthier & Tarr, 2002), may increase VSTM capacity by creating more efficient representations for storage. Feature-based theories propose separate capacity limits for different types of featural information, such as color and shape. Holistic processing may increase VSTM capacity by integrating otherwise independent and even spatially separate features, such as the shape of the eyes and nose. Integrating this information into a single "shape unit" may effectively circumvent feature-based capacity limits. In addition, VSTM capacity for multifeatured

objects may also be constrained by capacity-limited attention required to bind different features together (Delvenne & Bryuer, 2004; Wheeler & Treisman, 2002). In this case, holistic processing could reduce the burden to bind the features within a face together.

Holistic processing could also increase VSTM capacity in ways more consistent with object-based accounts of VSTM (Vogel et al., 2001): It may allow participants to incorporate more object features into the unified representations believed to serve as the units of VSTM. A more detailed and complete representation of the study array items would provide a considerable advantage when participants are required to differentiate between highly similar exemplars (Awh, Barton, & Vogel, 2007).

Alternatively, holistic representations for upright faces may be more robust and less susceptible to decay or interference than are the more feature-based representations created for inverted faces or objects. However, a recent study provides evidence against the possibility: Freire, Lee, and Symons (2000) found no effect of memory delay (1–10 sec) on the size of the inversion effect. This suggests the VSTM advantage for faces does not result from more robust representations, but rather from more qualitative differences between holistic and more featural representations.

An interesting question is why the advantage for faces in VSTM capacity only appears at encoding times longer than 500 msec, given that differences between the neural response to upright faces and inverted faces or objects occur as early as 170 msec after presentation (Rossion & Gauthier, 2002). One possibility is that holistic processing already has an effect at 500 msec, but that its contribution cannot overcome perceptual limitations under these conditions (Eng et al., 2006). Alternatively, because consolidation into VSTM is capacity limited, it may have insufficient time to consolidate more faces than objects even if more could be stored (Jolicœur & Dell'Acqua, 1998). Thus, the advantage for upright faces may only appear with sufficient time to complete consolidation. Consolidation time has been estimated to be as long as 500 msec per item (Chun & Potter, 1995; Jolicœur & Dell'Acqua, 1998), although it could be as short as 50 msec for simple objects (Vogel, Woodman, & Luck, 2006). Further studies should explore the interaction between consolidation mechanisms and perceptual processing efficiency.

Not only do we perceive faces differently, but this difference extends to our capacity to store them in VSTM. Indeed, more efficient perception of important categories in our environment would not be maximally adaptive if this advantage disappeared with any visual disruption. Interestingly, because holistic processing is also observed for objects of expertise (Gauthier & Tarr, 2002), perceptual experts may also demonstrate a greater VSTM capacity. If so, this would lead to an intriguing prediction: although practice on VSTM tasks and on visual search has not been found to improve VSTM capacity substantially (Chen, et al., 2006; Wolfe, Klempen, & Dahlen, 2000), expertise training procedures shown to increase holistic processing for objects may be more likely to impact VSTM. This would be worthy of further investigation.

## AUTHOR NOTE

This study was supported by NSF (0091752), NIH (EY13441), and James S. McDonnell Foundation awards. This research was also supported in part by the Temporal Dynamics of Learning Center (NSF Science of Learning Center SBE-0542013). We thank René Marois, Steven Most, Geoff Woodman, and PEN members for helpful discussions; Jay Todd for programming advice; and Ludvik Bukach for help with data collection. Correspondence regarding this article should be addressed to K. M. Curby, Temple University, Department of Psychology, Weiss Hall, 1701 N. 13th St., Philadelphia, PA 19122-6085 (e-mail: kim.curby@temple.edu).

*Note—Accepted by David A. Balota's editorial team.*

## REFERENCES

Alvarez, G. A., & Cavanagh, P. (2004). The capacity of visual short term memory is set both by visual information load and by number of objects. *Psychological Science*, **15**, 106-111.

Awh, E., Barton, B., & Vogel, E. K. (2007). Visual working memory represents a fixed number of items regardless of complexity. *Psychological Science*, **18**, 622-628.

Chen, D., Eng, H. Y., & Jiang, Y. (2006). Visual working memory for trained and novel polygons. *Visual Cognition*, **12**, 37-54.

Chun, M. M., & Potter, M. C. (1995). A two-stage model for multiple target detection in rapid serial visual presentation. *Journal of Experimental Psychology: Human Perception & Performance*, **21**, 109-127.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral & Brain Sciences*, **24**, 87-185.

Delvenne, J.-F., & Bruyer, R. (2004). Does visual short-term memory store bound features? *Visual Cognition*, **11**, 1-27.

Donderi, D.C. (2006). Visual complexity: A review. *Psychological Bulletin*, **132**, 73-97.

Eng, H. Y., Chen, D., & Jiang, Y. (2006). Visual working memory for simple and complex visual stimuli. *Psychonomic Bulletin & Review*, **12**, 1127-1133.

Farah, M. J., Tanaka, J. W., & Drain, H. M. (1995). What causes the face inversion effect? *Journal of Experimental Psychology: Human Perception & Performance*, **21**, 628-634.

Freire, A., Lee, K., & Symons, L. A. (2000). The face-inversion effect as a deficit in the encoding of configural information: Direct evidence. *Perception*, **29**, 159-170.

Gauthier, I., & Tarr, M. J. (2002). Unraveling mechanisms for expert object recognition: Bridging brain activity and behavior. *Journal of Experimental Psychology: Human Perception & Performance*, **28**, 431-446.

Jolicœur, P., & Dell'Acqua, R. (1998). The demonstration of short-term consolidation. *Cognitive Psychology*, **36**, 138-202.

Lawson, R., & Jolicœur, P. (1998). The effects of plane rotation on the recognition of brief masked pictures of familiar objects. *Memory & Cognition*, **26**, 791-803.

Olsson, H., & Poom, L. (2005) Visual memory needs categories. *Proceedings of the National Academy of Sciences*, **102**, 8776-8780.

Rossion, B., & Gauthier, I. (2002). How does the brain process upright and inverted faces? *Behavioral & Cognitive Neuroscience Reviews*, **1**, 63-75.

Tanaka, J. W., & Sengco, J. A. (1997). Features and their configuration in face recognition. *Memory & Cognition*, **25**, 583-592.

Tong, F., & Nakayama, K. (1999). Robust representations for faces: Evidence from visual search. *Journal of Experimental Psychology: Human Perception & Performance*, **25**, 1016-1035.

Vogel, E. K., Woodman, G. F., & Luck, S. J. (2001). Storage of features, conjunctions, and objects in visual working memory. *Journal of Experimental Psychology: Human Perception & Performance*, **27**, 92-114.

Vogel, E. K., Woodman, G. F., & Luck, S. J. (2006). The time course of consolidation in visual working memory. *Journal of Experimental Psychology: Human Perception & Performance*, **32**, 1436-1451.

Wheeler, M. E., & Treisman, A. M. (2002). Binding in short-term visual memory. *Journal of Experimental Psychology: General*, **131**, 48-64.

Wolfe, J. M., Klempen, N., & Dahlen, K. (2000). Post-attentive vi-

sion. *Journal of Experimental Psychology: Human Perception & Performance*, **26**, 693-716.

XU, Y. (2006). Understanding the object benefit in visual short-term memory: The roles of proximity and connectedness. *Perception & Psychophysics*, **68**, 815-828.

YIN, R. K. (1969). Looking at upside-down faces. *Journal of Experimental Psychology*, **81**, 141-145.

## NOTES

1. K = (hit rate + correct rejection rate − 1) ∗ set size.

2. The main effect of set size and its interactions with encoding duration and orientation were also significant ($p < .05$) but are not central to our research question.

3. Removed incorrect articulatory suppression trials accounted for <2% of trials.

4. The main effect of set size and its interactions with encoding duration and with stimulus category were also significant ($p < .05$) but were not central to our research question.

5. Further supporting this account is the significant interaction between stimulus category and orientation in Experiment 3, where the viewpoint of the car stimuli was controlled ($p = .0098$).

6. Removed incorrect articulatory suppression trials accounted for <2.3% of trials.

7. The main effect of set size and its interactions with encoding duration and orientation were also significant ($p < .05$) but were not central to our research question.

8. In this control experiment, there were both main effects of presentation format (sequential, simultaneous) [$F(1,28) = 7.12$, $p = .0125$], category (face, car) [$F(1,28) = 10.97$, $p = .0026$], and duration (500, 4,000 msec) [$F(2,56) = 57.94$, $p \leq .0001$]. Notably, there was no interaction between presentation format and category [$F(1,28) = 1.98$, $p = .170$] or duration ($F < 1$). However, there was an interaction between category and duration [$F(2,56) = 6.86$, $p = .0022$], but this did not interact with presentation format ($F < 1$). These results suggest that the advantage for faces over cars was not influenced by the presentation format of the stimuli and thus not by the pattern of eye movements used to encode the stimuli.

(Manuscript received September 24, 2005;
revision accepted for publication August 23, 2006.)