Acta Psychologica 143 (2013) 40–51

Contents lists available at SciVerse ScienceDirect

Acta Psychologica

journal homepage: www.elsevier.com/ locate/actpsy

The fidelity of visual memory for faces and non-face objects

Corrado Caudek *

Dipartimento di Psicologia, Università degli Studi di Firenze, Via di San Salvi 12, Complesso di S. Salvi, Padiglione 26, 50135 Firenze (FI), Italy
Center for Neuroscience and Cognitive Systems@UniTn, Istituto Italiano di Tecnologia, Corso Bettini 31, 38068 Rovereto (TN), Italy

A R T I C L E I N F O

Article history:
Received 19 May 2011
Received in revised form 3 December 2012
Accepted 6 December 2012
Available online 20 March 2013

PsycINFO classification:
2343 Learning & Memory

Keywords:
Visual working memory
Face recognition
Old–new recognition
Objects of expertise
Summed-similarity model

A B S T R A C T

The fidelity of visual working memory was assessed for faces and non-face objects. In two experiments, four levels of memory load (1, 2, 3, or 4 items) were combined with four perceptual distances between probe and study items, with maximum item confusability occurring for the minimum memory load. Under these conditions, recognition memory for multiple faces exceeded that of a single face. This result was primarily due to the higher false alarm rates for faces than non-face objects, even though the two classes of stimuli had been matched for perceptual discriminability. Control experiments revealed that this counterintuitive result emerged only for old–new recognition choices based on near-threshold image differences. For non-face objects, instead, recognition performance decreased with increasing memory load. It is speculated that the low memorial discriminability of the transient properties of a face may serve the purpose of enhancing recognition at the individual-exemplar level.

1. Introduction

Visual working memory (WM) provides temporary storage and manipulation of task-relevant information in cognitive processes such as perception (Simons & Rensink, 2005), attention (Awh & Jonides, 2001), and visual search (Emrich, Al-Aidroos, Pratt, & Ferber, 2010). WM maintains representations in an active and accessible state, but it has a limited capacity (Cowan, 2006).

The fidelity with which visual information can be maintained in WM depends on several factors. Large-scale or holistic information is extracted over a very short time, whereas the consolidation of this information, with the extraction of further details, requires longer presentation times (Hollingworth & Henderson, 2002; Melcher, 2001, 2006). The precision with which items are stored in WM is affected not only by *encoding time* but also by *set size*. With the increase of set size, less memory resources are allocated to each item and the precision with which items are stored in WM decreases (Alvarez & Cavanagh, 2004; Bays, Catalao, & Husain, 2009; Bays & Husain, 2008; Bays, Wu, & Husain, 2011; Brady, Konkle, & Alvarez, 2011; Wilken & Ma, 2004). The fidelity of WM also depends on *task demands*. Within a change-blindness paradigm, for example, the probability of a correct change detection is higher for the objects of central interest in the visual scene (Rensink, O'Regan, & Clark, 1997). Interestingly, the fidelity of WM is also influenced by *domain-specific expertise*. Wagar and Dixon (2005) showed that the properties of the information stored in WM depend on previous experience requiring the repeated categorization of the target objects into different families. In their study, a categorization learning phase improved the fidelity of WM for features diagnostic of category membership and *impaired* WM performance for non-diagnostic features.

The findings of Wagar and Dixon (2005) are consistent with recent studies suggesting that learning to categorize objects causes (1) an increase of perceptual discriminability along the dimensions relevant to the learned categories ("acquired distinctiveness"), and (2) a decrease in discriminability along the irrelevant dimensions ("acquired equivalence"). Goldstone and colleagues have proposed that "acquired distinctiveness" and "acquired equivalence" occur under both explicitly reinforced (i.e., supervised) and incidental (i.e., unsupervised) category acquisition (Gureckis & Goldstone, 2008). Many studies have provided empirical support for acquired distinctiveness (Goldstone & Steyvers, 2001; Notman, Sowden, & Özgen, 2005; Op de Beeck, Wagemans, & Vogels, 2003; Özgen & Davies, 2002), but empirical evidence in support to acquire equivalence is more elusive (e.g., Folstein, Palmeri, & Gauthier, 2012).

Here I propose that "acquired distinctiveness" and "acquired equivalence" modulate not only perceptual expertise, but also WM recognition. For a WM task, "acquired equivalence" translates into low-fidelity maintenance of transient and non-diagnostic features. The present experiments study this phenomenon for objects of expertise and objects of non-expertise.

1.1. "Acquired equivalence" in working memory

Based on the empirical findings described in the previous section, it is here proposed that "acquired distinctiveness" and "acquired

* Corresponding author. Tel.: +39 055 2055832; fax: +39 055 4271280.
E-mail address: corrado.caudek@unifi.it.

0001-6918/$ – see front matter
http://dx.doi.org/10.1016/j.actpsy.2012.12.002

equivalence" modulate the fidelity of the of the representations held in WM whenever experience induces a subordinate-shift in which objects are identified at a subordinate level rather than at the basic-level of categorization (Gauthier & Tarr, 1997; Johnson & Mervis, 1997; McGugin, Tanaka, Lebrecht, Tarr, & Gauthier, 2011; Nishimura & Maurer, 2008; Scott, Tanaka, Sheinberg, & Curran, 2006; Tanaka, Curran, & Sheinberg, 2005; Tanaka & Taylor, 1991). I propose that (1) perceptual dimensions that are relevant for identification at the individual level may receive a stronger memorial representation for objects of expertise than non-expertise ("acquired distinctiveness"), and (2) within-category image transformations that are irrelevant for identification at the individual level may manifest a lower memorial discriminability (i.e., may be represented with lower fidelity in WM) for objects of expertise than non-expertise ("acquired equivalence").

Face identity recognition always requires the selection of the invariant aspects that underlie face identity from the transient features generated by speech production, facial expression, and variations of the viewing conditions. Therefore, "acquired equivalence" may be especially important for faces, also considering that recognition at the individual exemplar is more important for faces than objects (Kanwisher, 2000).[1]

Previous work has shown a WM recognition advantage for faces over objects. For example, Curby and Gauthier (2007) showed that more faces can be stored in WM than other complex objects. Instead, the present data will show that faces can be at a disadvantage with respect to non-face objects, if the WM task concerns the recognition of transient changes in appearance (i.e., subtle image variations) that preserve the identity of the study items.

1.2. Plan of the experiments

In a pretest participants completed a same-different simultaneous matching task to measure perceptual discriminability for pairs of faces or cars lying on six morphing continua. In Experiments 1 and 2, participants performed a delayed matching task (Fig. 1) with the stimuli generated from the morph continua analyzed in the pretest. In Experiments 3a and 3b the difficulty of stimulus discriminability was decreased, in order to facilitate recognition performance.

Four levels of memory load (1, 2, 3, or 4 items to be retained in memory) were combined with four distances between the probe and the study items. In "different" trials, the physical differences between the probe and the to-be-remembered items comprised (1) small or large within-category distances (20 and 40 morphing steps, respectively), and (2) small or large distances crossing the category boundary (60 and 80 morphing steps, respectively) — see Tables 1 and 2. For the present stimuli, a physical difference of 20 morph steps is near the perceptual threshold for discrimination and it corresponds to subtle changes in appearance that preserve the identity of the item.

The items used in each single trial of Experiments 1 and 2 were selected from one morphing continuum generated between two faces or two cars. Different morphing continua were used to generate the items employed in different trials (Fig. 2).

1.3. Hypotheses of the present study

Memory loads of 1 and 2 included small and large within-category distances between the memory probe and the study items; memory loads of 3 and 4 included small and large across-category distances between the memory probe and the study items (see Table 2). The focus of the present study is on the small within-category differences, which were matched for perceptual discrimination across faces and cars, for stimulus presentation durations of 1000 ms (see Pretest).

[1] The perceptual expertise at recognizing objects at a more specific categorical level than the "basic level" of categorization (Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976) has been referred to as "individuation training" by McGugin et al. (2011).

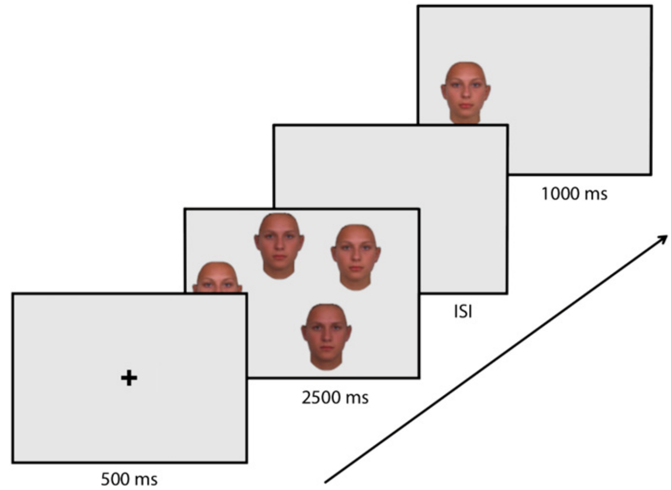

Fig. 1. An example trial from Experiments 1 and 2. After a fixation mark, a memory array of 1, 2, 3, or 4 items (faces, cat faces, or cars), a blank screen (ISI = 250, 750 ms, or 2500 ms), a memory probe, and another blank screen were presented sequentially. The participants reported whether the memory probe was the same as one of the items in the memory array.

"Acquired equivalence" predicts that subtle image differences, which are irrelevant for identification at the individual level, are represented in WM with lower fidelity for objects of expertise than non-expertise. In the present design, when the memory load was 1, the "new" probe differed from the study item only in terms of subtle image characteristics. Under these conditions, "acquired equivalence" predicts a higher false alarm rate for faces than cars.

The present design does not allow to test whether the hit rates are higher for faces than cars when the memory probe and the study items are separated by the category boundary. In fact, the memory loads 3 and 4 included both within-category and across-categories differences between the probe and the study items. Note, moreover, that the memory load of 2 included both small and large within-category differences between the probe and the study items. Under those conditions, "acquired equivalence" is not expected to occur.

In summary, when the memory load is larger than 1, there is no reason to expect that "acquired equivalence" and "acquired distinctiveness" may modulate in a different manner the hit rates and the false alarm rates of faces and cars. When the memory load is 1, instead, "acquired equivalence" predicts larger false alarm rates for faces than cars. As a consequence, in the present design the relationship between memory load and recognition accuracy, as assessed by d', is expected to be qualitatively different for the two classes of stimuli.

2. Experiment 1

Recognition accuracy for morphed Caucasian faces and cars was measured as the perceptual distance between the probe and the study items, the memory load, and the ISI were manipulated. The experiment was preceded by a pretest of the materials used in the old–new recognition task.

Table 1
Morph steps in the continuum between the memory probe and each item of the memory array in Experiments 1 and 2.

Memory load	Old trials				New trials			
1	0				20			
2	0	20			20	40		
3	0	20	40		20	40	60	
4	0	20	40	60	20	40	60	80

Table 2
Experiments 1 and 2. For each memory load, the first column indicates the positions on the morph continuum from which the items of the memory array were selected; the second column indicates the positions on the morph continuum from which the probe was selected in "new" trials; the third column indicates the positions on the morph continuum from which the probe was selected in "old" trials.

Memory array	"New" probe	"Old" probe	Memory array	"New" probe	"Old" probe
Memory load 1			Memory load 2		
5	25	5	5, 25	45	25
15	35	15	15, 35	55	35
65	85	65	45, 65	85	65
75	95	75	55, 75	95	75
30	10	30	30, 50	10	30
40	20	40	40, 60	20	40
90	70	90	70, 90	50	70
100	80	100	80, 100	60	80
Memory Load 3			Memory Load 4		
5, 25, 45	65	45	5, 25, 45, 65	85	65
15, 35, 55	75	55	10, 30, 50, 70	90	70
25, 45, 65	85	65	15, 35, 55, 75	95	75
35, 55, 75	95	75	20, 40, 60, 80	100	80
30, 50, 70	10	30	25, 45, 65, 85	5	25
40, 60, 80	20	40	30, 50, 70, 90	10	30
50, 70, 90	30	50	35, 55, 75, 95	15	35
60, 80, 100	40	60	40, 60, 80, 100	20	40

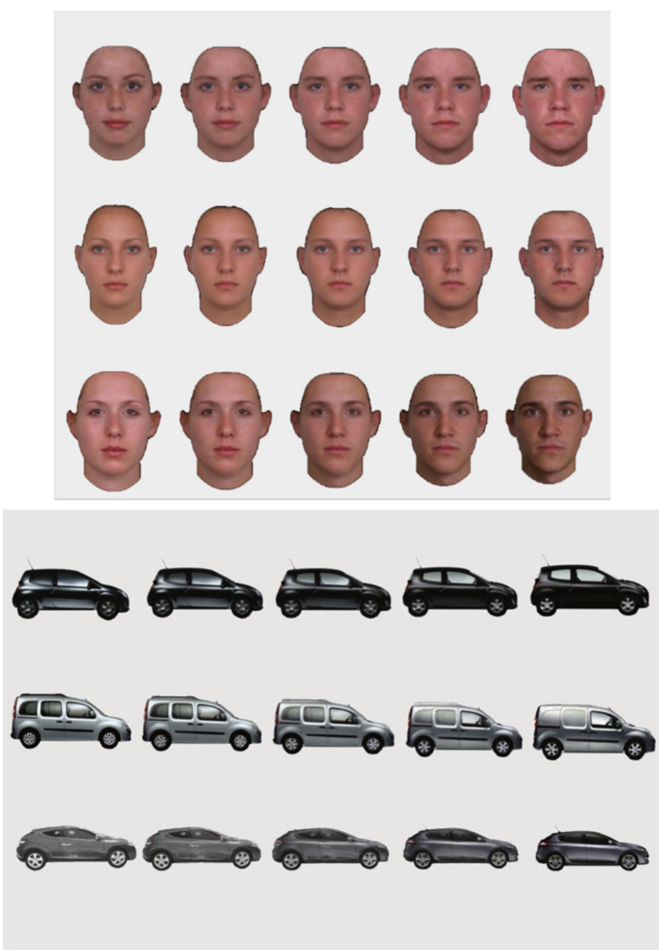

Fig. 2. The spectrum of face and car morphs in the pretest and in Experiment 1. Three face and three car morph continua were generated. The five images in each row represent the positions 1, 25, 50, 75, and 100 along the morph continuum. In a given trial, the items in the memory array and the probe were selected from one of these morph continua according to the scheme indicated in Table 2.

2.1. Pretests of materials for Experiment 1

Psychometric functions in discrimination of test stimuli with different levels of morphing between two faces or two cars were generated. To enhance holistic processing, exposure duration at test was limited (1000 ms). Unlimited exposure duration at test, in fact, can encourage a local analysis which disrupts holistic processing (Pallett & MacLeod, 2011; see also Hegde, 2008; Richler, Mack, Gauthier, & Palmeri, 2009).

2.1.1. Method

2.1.1.1. Participants. Four undergraduate students of the University of Florence with normal or corrected to normal vision participated in the pretest. 5070 trials were collected from each participant.

2.1.1.2. Materials and procedure. Face stimuli were derived from full-front digital renderings (300×400 pixels) of three Caucasian men and three Caucasian women (e.g., Webster, Kaping, Mizokami, & Duhamel, 2004). The study faces were generated by using a three-dimensional face modeling software (FaceGen Modeller, Version 3.1, Singular Inversions, Vancouver, BC, Canada). This software generates facial structures of male and female faces semi-randomly, with a high level of realism. Faces had no visible gender-specific features (e.g., facial hair or make up). Morphs were generated by entering male–female pairs into a morphing algorithm (Morph Man 4). Artificial continua were generated by morphing between pairs of faces differing in gender. Faces were morphed between genders in order to maximally differentiate the two extremes of each morph continuum. The stimuli were generated by using the same procedure as described by Afraz, Vaziri-Pashkam, and Cavanagh (2010). For each pairing, the morphing procedure resulted in 100 images. Car stimuli were generated in a similar way. Car pictures were recovered from catalogs of auto manufacturers of different years, in order to create morph continua between images of cars having similar shapes. Three morph continua were generated between pairs of faces and three continua were generated between pairs of cars (see Fig. 2). Stimuli were presented and responses collected using a custom script written with the PsychToolbox extension (Brainard, 1997; Pelli, 1997) of MATLAB (Mathworks, Massachusetts) on a 486-based PC-compatible computer connected to a 17-in. video monitor operating at 72 Hz.

Each trial started with the appearance of a small fixation point in the middle of the screen. After 500 ms, a pair of faces or cars was shown side-by-side for 1000 ms. In half of the trials, two identical faces or cars were shown; in the other half, the two faces or cars were slightly different. The method of constant stimuli was used to measure two psychometric functions for each morph continuum. In one psychometric function, the standard stimulus was 5% morphed toward one extreme of the continuum; in the other, the standard was 95% morphed toward the opposite extreme. The images used as comparative stimuli were separated by 5, 10, 15, 20, 25, 30, or 35 morphing steps from the standard toward the opposite extreme of the continuum. Subjects were not given feedback for their correct and incorrect key presses.

2.1.1.3. Results. Psychometric functions were fit with a nonparametric approach based on local linear fitting (Zchaluk & Foster, 2009). The PSE and the slope were computed for each of the twelve psychometric functions for each participant. There was not a statistically significant difference between the average PSEs of the two types of stimuli, $t_{46} = 0.55$, n.s., nor between the average slopes of the psychometric functions, $t_{46} = 0.30$, n.s. (see Table 3). The average PSEs were equal to 17.7 morph steps ($S.D. = 6.4$) and 16.2 morph steps ($S.D. = 7.8$) for the face and car morph continua, respectively. The average slopes of the psychometric functions were equal to 0.06 ($S.D. = 0.04$) and 0.06 ($S.D. = 0.03$) for the face and car morph continua, respectively.

Table 3
Average Points of Subjective Equality (PSE) and slopes of the psychometric functions for the perceptual discrimination task in the pretest. Standard errors are indicated in parenthesis.

Stimuli		PSE	Slope
Faces	Continuum 1	14.66 (3.31)	0.084 (0.03)
	Continuum 2	21.78 (3.23)	0.042 (0.01)
	Continuum 3	16.56 (2.08)	0.058 (0.01)
Cars	Continuum 1	20.63 (4.93)	0.046 (0.01)
	Continuum 2	12.36 (3.43)	0.051 (0.02)
	Continuum 3	15.49 (2.03)	0.073 (0.02)

2.1.2. Discussion

The pretest indicates that, for exposure durations of 1000 ms (the same as in Experiments 1–3), the face and car stimuli elicited comparable levels of discrimination performance in a simultaneous matching task. The results of the pretest provide the basis for using the physical distance on the morph continuum as an estimate of the perceptual distance across the two stimulus types. If the perceptual distance between two items is directly related to their discriminability (D'Lauro, Tanaka, & Curran, 2008), then the physical distance on the morphing continuum can be used as an estimate of perceptual distance. It is important to note, however, that the results of the pretest depend on stimulus duration. Two further pretest experiments provided evidence that varying exposure duration can lead to qualitatively different results (see Section 4).

2.2. Experiment

Recognition accuracy for morphed Caucasian faces and cars was measured for retention intervals of 250 ms, 750 ms, and 2500 sm.

2.2.1. Method

2.2.1.1. Participants. 28 undergraduate students from the University of Firenze, Italy, participated in the experiment. They all had normal or corrected-to-normal vision. All participants were naïve to the purpose of the study and had not participated in the pretest experiments.

2.2.1.2. Materials and procedure. From each of the six morph continua used in the pretest, one, two, three, or four positions were selected for generating the memory array [memory load 1 (1 item in the memory array), …, memory load 4 (4 items in the memory array)] and one position was selected for the memory probe. A scheme of stimuli generation is shown in Table 2.

Each of the 192 combinations of the morphed images used for the memory array and for the probe (three continua × the 64 combinations) was presented six times. In each trial, both the probe and the items of the memory array were randomly positioned within a 10° × 8° region. Each participant completed the 1152 trials over the course of four blocks in two 60-minute sessions. Trial order was randomized for each participant. Experimental trials were preceded by a short practice session allowing the participants to familiarize with the task and the stimuli. Stimulus type (faces versus cars) was manipulated between subjects.

After a fixation mark appeared for 500 ms in the center of the screen, between one and four faces or cars were shown for encoding for 2500 ms, followed by a 250, 750, or 2500 ms blank retention/maintenance phase, a 1000 ms retrieval phase with a single probe image at a random location, and a black screen until participants stated whether the probe replicated or not one of the study items (see Fig. 1). The response was followed by a 2000 ms inter-trial interval (blank screen). Half of all trials were "new" trials (the probe did not replicate any of the studies items). All stimuli were presented at a viewing distance of 70 cm, which was kept constant by using a chin rest.

2.2.1.3. Data analysis. A measure of the ability of participants to determine whether the probe replicated or not one of the items in the encoding display is provided by the sensitivity index d' used in signal detection theory. The statistic d' can be computed as $d' = \text{probit}(\text{hit rate}) - \text{probit}(\text{false alarm rate})$. Equivalently, d' can be estimated by a probit regression (DeCarlo, 1998), with the advantage that, when the analysis is framed as a regression model, other predictor variables can be considered (Wright, Horry, & Skagerberg, 2009). By following this second approach, a mixed-effect linear model with binomial error structure and a probit link function was used to analyze the participants' binary responses, with participants and items as crossed random effects, and Faceold (whether the probe was old or new), Setsize (1, 2, 3, or 4), and ISI as fixed effects. These analyses were performed using the lme4 package (Bates & Sarkar, 2007) for the R statistical environment (version 2.15.1, RDevelopment Core Team, 2012). The participants' accuracy was allowed to vary by adding to the random part of the model both a term for the variance of accuracy and the covariance between accuracy and responding "old".[2]

2.2.2. Results

2.2.2.1. Recognition accuracy. Fig. 3 (left) shows recognition accuracy (d') for face stimuli as a function of memory load and ISI. Overall d' was equal to 0.84, $z = 8.67$, $p = .001$. The memory load × ISI interaction was not statistically significant, $\chi^2_6 = 3.43$, $p = .754$. The effect of memory load was statistically significant, $\chi^2_{12} = 87.75$, $p = .001$. Accuracy was significantly lower for memory load 1 than for memory loads 2 and 3 ($\Delta d'_{1-2} = -0.42$, $z = -3.44$, $p = .001$; $\Delta d'_{1-3} = -0.26$, $z = -2.17$, $p = .030$). The effect of ISI was statistically significant, $\chi^2_{10} = 104.35$, $p = .001$. For retention intervals of 750 ms and 2500 ms, accuracy was significantly lower than for a retention interval of 250 ms ($\Delta d'_{250-750} = -0.22$, $z = -3.85$, $p = .001$; $\Delta d'_{250-2500} = -0.30$, $z = -5.43$, $p = .001$).

Fig. 3 (right) shows recognition accuracy (d') for car stimuli as a function of memory load and ISI. For cars, the memory load × ISI interaction was not statistically significant, $\chi^2_6 = 7.37$, $p = .288$. The effect of memory load was significant, $\chi^2_{12} = 65.00$, $p = .001$. Accuracy was not significantly different for memory loads 2 and 3 than for memory load 1 ($\Delta d'_{1-2} = -0.16$, $z = -1.06$, $p = .289$; $\Delta d'_{1-3} = -0.27$, $z = -1.79$, $p = .074$). Accuracy was significantly lower for memory load 4 than for memory load 1 ($\Delta d'_{1-4} = -0.32$, $z = -2.10$, $p = .036$). The effect of ISI was statistically significant, $\chi^2_{10} = 56.18$, $p = .001$. For retention intervals of 750 ms and 2500 ms, accuracy was significantly lower than for a retention intervals of 250 ms ($\Delta d'_{250-750} = -0.18$, $z = -3.37$, $p = .001$; $\Delta d'_{250-2500} = -0.28$, $z = -5.48$, $p = .001$). Overall, d' was 0.38 higher for cars than for faces, $z = 4.63$, $p = .01$.

For memory load 1, d' was 0.78 lower for faces than cars, $z = 4.91$, $p = .001$. There were no differences in recognition accuracy between stimulus categories for memory load 2, ($\Delta d' = 0.14$, $z = 0.96$, $p = .34$) and for memory load 3 ($\Delta d' = 0.21$, $z = 1.57$, $p = .12$). For memory load 4, d' was 0.43 lower for faces than cars, $z = 3.82$, $p = .001$.

2.2.2.2. Hit rates and false-alarm rates. Hit rates and false alarm rates are shown in Fig. 4. For memory load 1 (small intra-class differences), the false alarm rates were expected to be higher for faces than cars. This hypothesis was tested by means of a by-subject random-intercepts

[2] An advantage of the mixed-effect analyses over the traditional approach (where an ANOVA is performed on the d' scores computed for each participant in each cell of the factorial design), is the possibility to specify crossed (or partially crossed) random effects for participants and items (see also Caudek & Domini, 2013; Wright & London, 2009). In this respect, mixed models can replace the by-subjects (F1) and by-items (F2) ANOVAs (Judd, Westfall, & Kenny, 2012). The significance of the fixed-effects was assessed by computing the deviance statistics (minus 2 times the log-likelihood) of nested models; change in deviance is distributed as chi-square, with degrees of freedom equal to the number of parameters deleted from the model (e.g., Baayen, Davidson, & Bates, 2008).

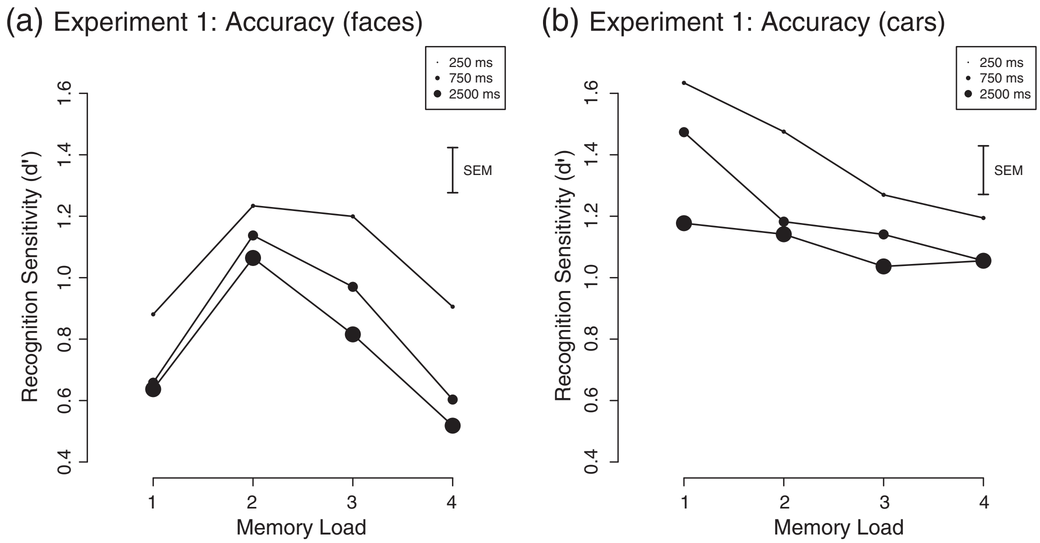

Fig. 3. Experiment 1. Recognition sensitivity (d') a function of memory load and ISI: (a) faces, (b) cars.

random-slopes mixed-effect model with logit(false alarm rate) as the DV and memory load, ISI, and stimulus type (faces, cars) as the fixed-effects, with participants and items as random effects. The 3-way interaction was statistically significant, $\chi^2_{28} = 229.35$, $p = .001$. For memory load 1, logit(false alarm rate) was higher for faces than cars at all ISIs (250 ms: $t_{312} = 3.56$, $p = .001$; 750 ms: $t_{312} = 3.10$, $p = .002$; and 2500 ms: $t_{312} = 2.42$, $p = .015$). For memory loads larger than 1, the effects of stimulus type and ISI were not significant, nor was significant the stimulus type × ISI interaction ($\chi^2_{15} = 21.66$, $p = .117$); logit(false alarm rate) significantly decreased as a function of memory load ($\chi^2_2 = 8.46$, $p = .014$).

A similar analysis used logit(hit rate) as the DV. The effect of stimulus type was not statistically significant, nor were any interactions between stimulus type and the other variables, $\chi^2_{25} = 36.59$, $p = .063$. The effects of ISI ($\chi^2_2 = 28.34$, $p = .001$) and memory load ($\chi^2_3 = 44.795$, $p = .001$) were statistically significant (see Fig. 4).

2.2.2.3. Summed-similarity model of short-term item recognition. The idea that WM fidelity is modulated by "acquired distinctiveness" and "acquired equivalence" to maximize identity recognition is consistent with the "summed-similarity models" of recognition, categorization, and identification (e.g., Huang & Sekuler, 2010; Kahana & Sekuler, 2002; Nosofsky, 1988; Sekuler & Kahana, 2007). According to these models, in an old–new recognition task the probe acts as a cue that activates matching memory representations. The probability of a "yes/match" response increases with the summed similarity between the probe and the memory samples (Kahana & Sekuler, 2002; Zhou, Kahana, & Sekuler, 2004). Importantly, these models propose that different stimulus dimensions are represented according to their diagnosticity for category membership: A greater "weight" is given to psychological dimensions that are relevant for the categorization and a smaller "weight" is given to irrelevant dimensions.

The summed-similarity model was implemented as described in Eq. (A.5) of the Appendix A. Tables 4 and 5 report the estimated β coefficients for the distances S1 (20 morph steps), S2 (40 morph steps), S3 (60 morph steps), and S4 (80 morph steps) between the memory probe and each item of the memory array (see Table 1). The coefficients β_{S1} and β_{S2} describe the contributions of small and large

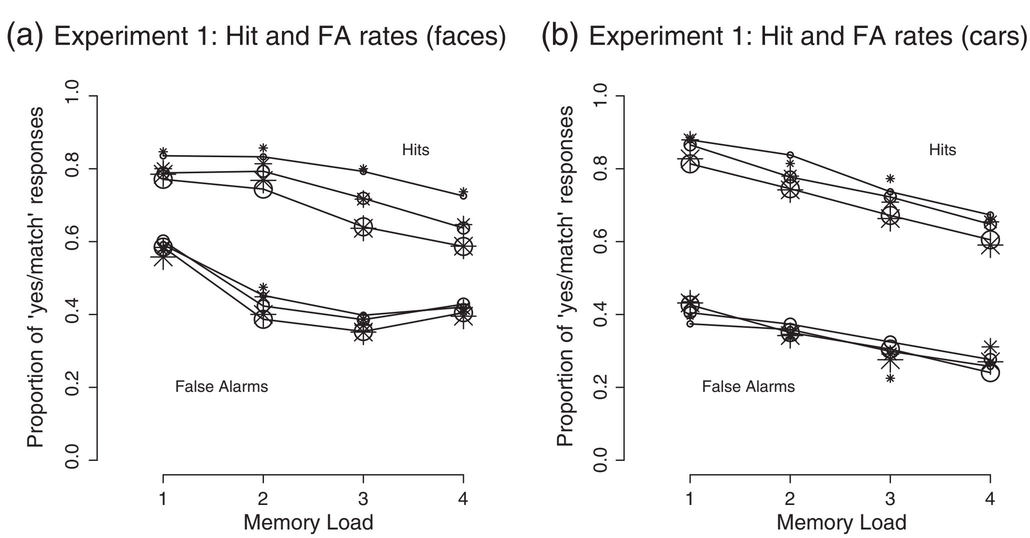

Fig. 4. Experiment 1. Hit rates and false alarm rates as a function of memory load and ISI (small open circles: 250 ms; medium open circles: 750 ms; large open circles: and 2500 ms) for faces (left panel) and cars (right panel). The asterisks denote the predicted probabilities computed according to Eq. (A.5) with the parameters shown in Tables 4 and 5 (small asterisks: 250 ms ISI; medium asterisks: 750 ms ISI; large asterisks: and 2500 ms ISI).

Table 4
Estimated parameters of Eq. (A.5) for the face stimuli in Experiment 1. Significance levels were computed by logistic mixed effects models with the participants' binary responses as the dependent variable with binomial error structure, participants and items as random effects, and *M*, *S1*, *S2*, *S3*, and *S4* as fixed effects.

	Coef β	SE(β)	z	p
ISI = 250 ms				
Intercept	0.22	0.20	1.1	>0.3
M	1.49	0.09	17.1	**<.0001**
S1	0.10	0.19	0.5	>0.6
S2	−0.60	0.16	−3.7	**<.001**
S3	−0.65	0.19	−3.5	**<.001**
S4	0.19	0.27	0.7	>0.5
ISI = 750 ms				
Intercept	0.25	0.19	1.3	>0.2
M	1.13	0.08	13.8	**<.0001**
S1	0.11	0.17	0.6	>0.5
S2	−0.81	0.15	−5.3	**<.0001**
S3	−0.59	0.18	−3.3	**<.001**
S4	0.43	0.26	1.6	>0.1
ISI = 2500 ms				
Intercept	0.33	0.19	1.8	>0.1
M	0.97	0.08	11.9	**<.0001**
S1	−0.12	0.17	−0.7	>0.5
S2	−0.95	0.15	−6.2	**<.0001**
S3	−0.38	0.18	−2.1	**<.05**
S4	0.41	0.27	1.6	>0.1

intra-class differences to the probability of a "yes/match" response, respectively; the coefficients β_{S3} and β_{S4} describe the contributions of small and large cross-class differences to the probability of a "yes/match" response, respectively. The size of these coefficients indicates the relative contribution of each dissimilarity level to the probability of a "yes/match" response.

Note that, although the estimated coefficient β_{S1} for the car stimuli is statistically significant at all ISIs (Table 5), the coefficient β_{S1} for the face stimuli is not (Table 4). It is also worth noting that the coefficient β_{S4} (which is associated to the largest dissimilarity S4 = 80 morph steps) is not statistically significant at all ISIs, neither for the face nor for the car stimuli. This result is consistent with the summed-similarity model, which assumes that the probe does not activate highly dissimilar stored exemplars (Nosofsky, Little, Donkin, & Fific, 2011).

Table 5
Estimated parameters of Eq. (A.5) for the car stimuli in Experiment 1. Significance levels were computed by logistic mixed effects models with the participants' binary responses as the dependent variable with binomial error structure, participants and items as random effects, and *M*, *S1*, *S2*, *S3*, and *S4* as fixed effects.

	Coef β	SE(β)	z	p
ISI = 250 ms				
Intercept	0.14	0.20	0.7	>0.5
M	1.90	0.10	18.8	**<.0001**
S1	−0.70	0.19	−3.6	**<.001**
S2	−0.38	0.17	−2.3	**<.05**
S3	−1.02	0.25	−4.1	**<.0001**
S4	0.51	0.32	1.6	>0.1
ISI = 750 ms				
Intercept	0.46	0.19	2.4	**<.05**
M	1.53	0.10	16.0	**<.0001**
S1	−0.88	0.18	−4.8	**<.0001**
S2	−0.56	0.16	−3.6	**<.001**
S3	−0.46	0.23	−2.0	**<.05**
S4	0.21	0.31	0.7	>0.5
ISI = 2500 ms				
Intercept	0.24	0.20	1.2	>0.2
M	1.33	0.09	14.1	**<.0001**
S1	−0.62	0.17	−3.7	**<.001**
S2	−0.56	0.15	−3.7	**<.001**
S3	−0.57	0.23	−2.5	**<.05**
S4	−0.06	0.31	−0.2	>0.8

2.2.3. Discussion

The hypotheses described in Section 3 predict (1) higher false alarm rates for faces than cars in the case of memory load 1 (i.e., when the "new" memory probe and the study item differ only in terms of subtle image characteristics), and (2) no difference in false alarm rates and in hit rates between the two classes of stimuli for memory loads 2–4 (when the differences between the "new" memory probe and the study items are modulated as described in Table 2). The results reported in Section 2 are consistent with these predictions.

For faces, the hit rates and false alarm rates resulted in a lower recognition accuracy, as assessed by d', for memory load 1 than for memory loads 2 and 3 (Fig. 3 left). For cars, conversely, the results replicate the usual finding indicating that performance tends to decrease as memory load increases (e.g., Jackson & Raymond, 2008; Luck & Vogel, 1997; Vogel & Machizawa, 2004) — see Fig. 3 right.

The increase of false alarm rates with increased typicality has already been reported in old–new recognition experiments (e.g., Zaki & Nosofsky, 2001). The present results suggest that the false alarm rates in WM recognition depend, not only on perceptual discriminability, but also on whether the study items are (or are not) objects of expertise (i.e., faces or cars).

Consistent with the hypotheses, the coefficient β_{S1} of the summed-similarity model was statistically significant for the car stimuli, but not for the face stimuli. This indicates that the smallest intra-class differences provided a statistically significant contribution to the probability of participants saying the probe was seen before for the car stimuli, but not for the face stimuli.

2.2.4. Caveats

It is important to point out that the results described in the previous section depend critically on how similar the items are and how perceptual similarity is measured. (1) In the present study, memory fidelity was lower for faces than for cars, but only when the image differences between the probe and the study items were just over the threshold of perceptual discrimination (see also Section 4). (2) Perceptual similarity was indexed by perceptual discrimination performance. The stimuli were selected so that, for a stimulus presentation time of 1000 ms (i.e., the same presentation time used for the memory probe in the delayed matching task), performance in the simultaneous matching task was similar across face and car stimuli. However, as indicated in the two experiments briefly described below, perceptual similarity depends on encoding duration.

In one experiment (13 participants), the same stimuli of the Pretest were used to measure perceptual similarity with the procedure described in Corneille, Hugenberg, and Potter (2007). Presentation duration was 250 ms, after which the screen was blanked. The two images used on each trial were either identical or differed by 20 steps along the morph continuum. Fifteen unique different-face pairs were selected from each of the six morph continua. For each morph continuum, 15 same-face pairs and 15 different-face pairs were presented 22 times, totaling 660 trials per participant. Overall d' was equal to 1.24 for faces and to 0.81 for cars ($z = 7.17$, $p < .001$). Therefore, the morphed cars were closer to each other in perceptual space than the morphed faces, when processed within a temporal window of 250 ms.

In another experiment, the procedure was the same as above, except that participants were instructed to locally compare the two images shown side-by-side, with an unlimited presentation time. Under these conditions, discrimination accuracy was higher for cars than for faces. This is not surprising because the face images at the two endpoints of each morph continuum always comprised the same local features. But this was not true for the car images. As a consequence, when the image pairs were the same distance apart on the morph continuum, the *local comparison* of two car morphs allowed for an easier discrimination than the local comparison of two face morphs.

In summary, an equal number of steps on the morph continuum did not produce equivalent levels of perceptual discrimination performance across faces and car stimuli, if encoding duration was either very short or very long. It is thus possible that the results of Experiment 1 depend on the presentation duration of 1000 ms for the memory probe, which was adopted to maximize participants' reliance on holistic cues and to prevent a local encoding of individual face and car parts (Macchi Cassia, Picozzi, Kuefner, Bricolo, & Turati, 2008; Richler et al., 2009).

3. Experiment 2

The purpose of Experiment 2 was to determine whether the results of Experiment 1 could be replicated with other race/species faces.

3.1. Method

3.1.1. Participants

24 undergraduate students from the University of Florence participated in the experiment. All of them were Caucasian and had normal or corrected-to-normal vision. All participants were naïve to the purpose of the study and had not participated in Experiment 1. None of them had lived abroad (in an ethnically different environment) for a significant amount of time.[3]

3.1.2. Materials and procedure

The stimuli were generated by using 12 images (300 × 400 pixels) of human faces (six Caucasian facial images and six afro-American facial images) and 12 images of cat faces as indicated in Section 2. For human faces, artificial continua were generated by morphing between pairs of faces differing in gender but not race. The Caucasian faces were the same as in Experiment 1. For cats, different cat faces were morphed, as indicated in Fig. 5. Images of cats with a lighter or a darker fur were used. Because this variable had no effect on recognition performance, it was not further analyzed. Cat faces were expected to produce intermediate results between human faces and cars. Discrimination measures for cat faces were not collected because the present study focused on the two extreme categories. Task and procedure were the same as in Experiment 1, except that ISI was kept fixed at 750 ms.

Fig. 5. The spectrum of afro-American face and cat face morphs in Experiment 2. In the 25% of the trials of Experiment 2, the items of the memory array and the probe were selected from the three morph continua represented in the top panel of Fig. 2; in the 25% of the trials the stimuli were selected from the three morph continua of the afro-American faces (top panel); in the remaining 50% of the trials, the stimuli were selected from six morph continua of cat faces (bottom panel) — only three cat morph continua are shown here.

3.2. Results

3.2.1. Recognition accuracy

Fig. 6 (left) shows recognition accuracy (d') as a function of memory load and stimulus type (Caucasian faces, afro-American faces, and cat faces). The d' values were 1.10 ($z = 14.09$), 1.21 ($z = 15.59$), and 1.025 ($z = 15.14$), for Caucasian, afro-American, and cat faces, respectively. In a by-subject random-intercepts random-slopes mixed-effect model, the memory load × stimulus type interaction was statistically significant, $\chi^2_6 = 13.38$, $p = .037$.

For Caucasian faces, the effect of memory load was statistically significant, $\chi^2_3 = 18.77$, $p = .001$. Accuracy was significantly lower for memory load 1 than for memory load 2 ($\Delta d'_{1\text{–}2} = -0.59$, $z = 3.71$, $p = .001$). Accuracy for memory load 1 did not differ from accuracy for memory loads 3 and 4 ($\Delta d'_{1\text{–}3} = 0.06$, $z = 0.41$, and $\Delta d'_{1\text{–}4} = -0.13$, $z = -0.62$).

For afro-American faces, the effect of memory load was statistically significant, $\chi^2_3 = 16.53$, $p = .001$. Accuracy was significantly lower for memory load 1 than for memory load 2 ($\Delta d'_{1\text{–}2} = -0.37$, $z = 2.34$, $p = .020$). Accuracy for memory load 1 did not differ from accuracy for memory load 3 ($\Delta d'_{1\text{–}3} = 0.01$, $z = 0.052$). Accuracy was significantly lower for memory load 4 than for memory load 1 ($\Delta d'_{1\text{–}4} = 0.481$, $z = 2.33$, $p = .020$). Accuracy for afro-American faces did not differ from accuracy for Caucasian faces, $\chi^2_8 = 14.84$, $p = .062$.

For cat faces, the effect of the memory load on recognition accuracy was not statistically significant, $\chi^2_3 = 2.67$, $p = .446$. Accuracy was significantly lower for cat faces than for human faces ($\Delta d' = -0.143$, $z = -2.453$, $p = .014$).

3.2.2. Hit rates and false-alarm rates

Hit rates and false alarm rates are shown in Fig. 6 (right). In a by-subject random-intercepts random-slopes mixed-effect model having logit(false alarm rate) as the DV and memory load and stimulus type (Caucasian faces, afro-American faces, cat faces) as the as fixed-effect predictors, with participants and items as random effects, the interaction memory load × stimulus type was not statistically significant, $\chi^2_6 = 10.48$, $p = .106$. The effect of memory load was statistically significant, $\chi^2_3 = 41.53$, $p = .001$; the logit(false alarm rate) was significantly higher for the memory load 1 than for memory load 2, 3, or 4 (1 vs. 2: $t_{260} = 7.53$; 1 vs. 3: $t_{260} = 9.94$; 1 vs. 4: $t_{260} = 5.05$). The logit(false alarm rate) did not differ across Caucasian and cat faces, $t_{261} = 0.41$, but was higher for Caucasian than for afro-American faces, $t_{261} = 3.66$, $p = .001$.

When logit(hit rate) was the DV, the memory load × stimulus type interaction was not statistically significant, $\chi^2_6 = 9.98$, $p = .125$. The

[3] No information was collected about whether participants had extensive experience with cats.

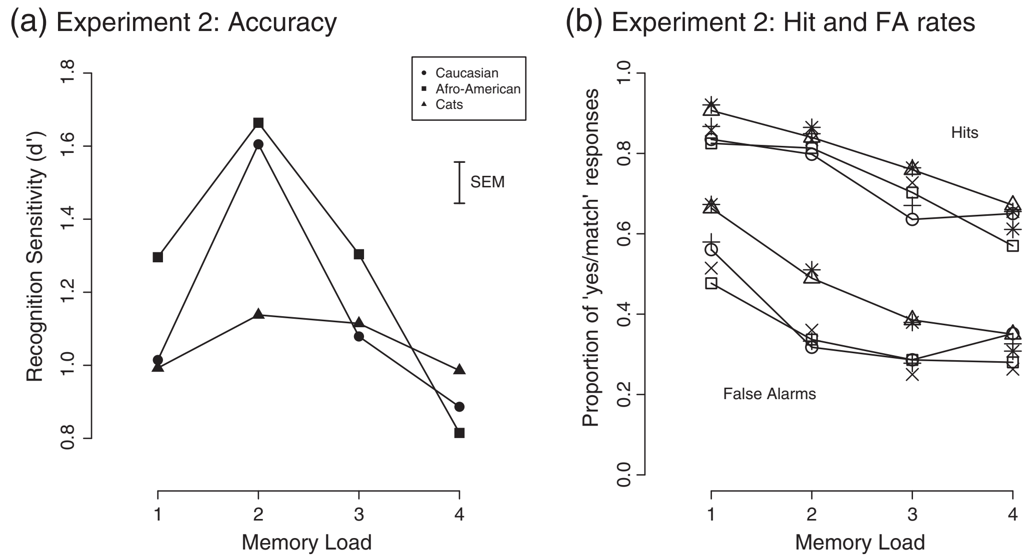

Fig. 6. Experiment 2. Left panel: Recognition sensitivity (d') as a function of memory load for Caucasian, afro-American, and cat faces. Right panel: Hit rates and false alarm rates as a function of memory load for Caucasian, afro-American, and cat faces. The $+$, $\times$, and * symbols denote the predicted probabilities computed according to Eq. (A.5) with the parameters shown in Table 6 for Caucasian, afro-American, and cat faces, respectively.

effect of stimulus type was not statistically significant, $\chi^2_2 = 4.46$, $p = .108$. The effect of memory load was statistically significant, $\chi^2_3 = 36.38$, $p = .001$. The logit(hit rate) was higher for memory load 1 than for memory load of 2, 3, or 4 (1 vs. 2: $t_{260} = 2.41$; 1 vs. 3: $t_{260} = 7.08$; 1 vs. 4: $t_{260} = 6.19$).

3.2.3. Summed-similarity model of short-term item recognition

Table 6 shows the estimated β coefficients for the distances *S1*, *S2*, *S3*, and *S4* between the memory probe and each item of the memory array. For Caucasian and afro-American faces, the estimated coefficients β_{S1} and β_{S4} are not statistically significant. These results replicate those of Experiment 1. For cat faces, instead, β_{S1} and β_{S4} are both statistically significant.

Table 6
Estimated parameters of Eq. (A.5) for the stimuli of Experiment 2. Significance levels were computed by logistic mixed effects models with the participants' binary responses as the dependent variable with binomial error structure, participants and items as random effects, and *M*, *S1*, *S2*, *S3*, and *S4* as fixed effects.

	Coef β	SE(β)	z	p
Caucasian faces				
Intercept	0.47	0.19	2.4	<**.05**
M	1.41	0.09	16.3	<**.0001**
S1	−0.18	0.17	−1.1	>0.3
S2	−1.52	0.15	−10.2	<**.0001**
S3	−0.47	0.19	−2.4	<**.05**
S4	0.51	0.38	1.3	>0.2
Afro-American faces				
Intercept	0.24	0.18	1.3	>0.2
M	1.56	0.08	18.9	<**.0001**
S1	−0.22	0.16	−1.4	>0.2
S2	−0.94	0.14	−6.8	<**.0001**
S3	−0.96	0.19	−5.1	<**.0001**
S4	0.15	0.38	0.4	>0.7
Cat faces				
Intercept	1.32	0.18	7.3	<**.0001**
M	1.13	0.07	16.3	<**.0001**
S1	−0.73	0.17	−4.3	<**.0001**
S2	−1.01	0.12	−8.2	<**.0001**
S3	−0.97	0.15	−6.5	<**.0001**
S4	−0.71	0.26	−2.7	<**.01**

3.3. Discussion

Accuracy was significantly lower for memory load 1 (i.e., for the smallest within-category distances between the memory probe and the study item) than for memory load 2, for both Caucasian and afro-American faces. As in Experiment 1, this result is primarily due to high false alarm rates when the memory array comprised one human face. For cat faces, accuracy was not affected by memory load.

The results of Experiment 2 are consistent with the hypothesis that visual expertise with human faces affects WM recognition through tuning of stored representations with "acquired equivalence" (i.e., low fidelity representation of irrelevant dimensions). For Caucasian participants, "acquired equivalence" affected WM recognition also for images of afro-American faces. The results obtained with cat faces, instead, differed from those obtained with human faces and with cars (Experiment 1). The importance of this difference, however, must be interpreted with caution, due to the possibility that it might depend on low-level differences between the two classes of images rather than on the cross-specie manipulation. It is possible to speculate that cat faces, because of their resemblance to human faces, combine WM processing of expert-domain objects and objects of non-expertise.

4. Experiments 3a and 3b

The two experiments briefly described here demonstrate that the results of Experiments 1 and 2 depend critically on item confusability.

4.1. Experiment 3a

Experiment 3a replicated the design of Experiment 2, except that the items were front-view photographs of easily discriminable faces of each sex (e.g., Morgan, Klein, Boehm, Shapiro, & Linden, 2008). The experiment consisted in 320 "old" trials and 320 "new" trials. Ten naive observers participated in the experiment. For memory loads 1, 2, 3, and 4, d' was equal to 4.17 (*S.E.* = 0.15), 2.35 (*S.E.* = 0.09), 1.39 (*S.E.* = 0.07), and 1.07 (*S.E.* = 0.07), respectively. Accuracy was significantly higher for memory load 1 than for memory load 2 (z = -10.75, p < .001), for memory load 2 than for memory load 3 (z = -8.35, p < .001), and for memory load 3 than for memory load 4 ($z = -3.19$, p < .002).

4.2. Experiment 3b

Experiment 3b replicated the design of Experiment 3a, except that (1) the memory array comprised only one or two faces, and (2) the stimuli were generated by morphing between images of real human faces. Only the inner portion of a face was made visible through an oval aperture. In "new" trials, the physical distances between the probe and the study items were comparable to the large within category differences used in Experiments 1 and 2. The experiment consisted in 354 "old" trials and 354 "new" trials. Seven naive observers participated in the experiment. For memory loads 1 and 2, d' was equal to 1.30 (*S.E.* = 0.05) and 1.04 (*S.E.* = 0.05), respectively. Accuracy was significantly higher for memory load 1 than for memory load 2 ($z = 5.85$, $p < .001$).

4.3. Discussion

Experiments 3a and 3b indicate that "acquired equivalence" affects the memorial discriminability of faces only when the memory probe and the study item are separated by within-category distances that are just over the threshold of perceptual discrimination. The results of Experiment 3b confirm that there is no effect of "acquired equivalence" on recognition performance for "large" within-category differences between the memory probe and the study items.

5. General discussion

Recognition at the individual-exemplar level tends to occur by default for expert-domain objects, but not for objects of non-expertise (e.g., Anaki & Bentin, 2009). Previous studies suggest that "acquired equivalence" and "acquired distinctiveness" enhance recognition at the individual-exemplar level (Goldstone, 1994; Nosofsky, Little, & James, 2012). The present study shows that "acquired equivalence" has the effect of decreasing memorial discriminability, more for objects of expertise (faces) than for objects of non-expertise (cars), even if the changes between study items and the probe had been matched for perceptual discriminability. When the task requires the recognition of subtle changes in appearance, which preserve the identity of the study items, the effect of "acquired equivalence" may be so dramatic as to alter the usual association between increased memory load and decreased recognition of faces (Experiments 1 and 2). This counterintuitive result in our data is primarily due to high false alarm rates that were observed when only one face was retained in memory and the differences between the memory probe and the study item were just over the threshold of perceptual discrimination.

5.1. Alternative interpretations

5.1.1. Encoding differences

An alternative explanation is that, during the learning phase, participants stored in WM a larger number of details for cars than faces. It is also possible that observers are sensitive to subtle image changes, but they are unable to access such information when the task requires to focus their attention to the identity of the items stored in memory. Furthermore, while subtle differences may be used to discriminate cars, other features may be responsible for facial identity discrimination.

5.1.2. Ensemble statistics

It has been suggested that WM stores not only information about individual items, but also information about the "featural context" or the "ensemble statistics" (Alvarez, 2011). This "relational" encoding could contribute to highlight the features that are more relevant for performing the old–new task (e.g., Haberman & Whitney, 2012; Jiang, Kwon, Shim, & Won, 2010; Magnussen, 2000). In the WM task, therefore, participants might have been at a disadvantage when the memory array comprised a single item. An explanation in terms of "relational" encoding, however, leaves open the question of why face and car stimuli produced different results.

5.1.3. Crowding

The inability to recognize objects in clutter, or crowding, is usually described between objects, but it can also occur within an object. For example, Martelli, Majaj, and Pelli (2005) presented a face in an eccentric position for 200 ms and found that it is unrecognizable, unless it is huge. They argued that this occurs because also faces are recognized by parts: When the image is small, the parts of the face cannot be isolated in a single glance and this impairs recognition. In another study, Louie, Bressler, and Whitney (2007) presented their stimuli in an eccentric position for 400 ms and found that recognition of an upright face was significantly worse when the target face was surrounded by upright flanker faces than when it was surrounded by inverted flankers, or none at all. Even though selective crowding can take place between high-level representations of faces, there are several reasons why the present findings cannot be explained by this attention mechanism. (1) The present results concern performance in a memory task, whereas crowding is a perceptual phenomenon. (2) In the present study, recognition performance was significantly better when the memory array comprised two faces rather than a single one. Crowding would predict the opposite result. (3) The memory array was presented for 2500 ms and comprised no more than four items. This presentation time is sufficient to foveate each item over multiple fixations. Therefore, crowding did not impair the encoding of the memory array. (4) 1000 ms provide enough time to foveate and also to scan the different parts of the test probe. Therefore, crowding did not impair the processing of the probe item.

5.2. Global versus local processing

A limited presentation time of 1000 ms for the memory probe biases participants toward a global processing strategy, which favors face recognition and holistic processing over feature-based processing (e.g., Calder & Young, 2005).[4] The present study suggests that, also under these conditions, "acquired equivalence" can put faces at a disadvantage with respect to a memory task for subtle image changes compared to other categories of stimuli.

5.3. The face-space model

Valentine (1991) proposed that faces are represented as points in a multidimensional similarity space centered on a prototypical average face. The dimensions spanning this face-space are assumed to encode the critical information used to discriminate faces (e.g., Busey, 1998; Busey & Arici, 2009; Busey & Tunnicliff, 1999; Hancock, Burton, & Bruce, 1996; Knapp, Nosofsky, & Busey, 2006). The tolerance to identity-preserving transformations has been accommodated within the face-space model by hypothesizing the existence of "identity clusters" that locate close to one another all the face images of the same person under varying viewing conditions (Blank & Yovel, 2011). The present results can be understood within this framework by postulating that it is more likely to retrieve from WM the mean of each "identity cluster" than any individual exemplar. Objects could also be represented in a similarity space, but the principles underlying their organization in similarity space might be different than for faces.

[4] Several lines of evidence suggest that a short presentation time favors face processing. For example, the face-identity aftereffect is stronger when the test stimuli are presented for shorter durations than for longer durations (Leopold, Rhodes, Müller, & Jeffery, 2005). A short presentation time has been also shown to be important for the emergence of the caricature effect (Lee & Perrett, 2000).

5.4. Separate face processing mechanisms

Multiple evidence suggest that separate modules in the brain process the expression of emotion and the identity of faces in a relatively independent fashion (Bruce & Young, 1986; Calder & Young, 2005; D'Argembeau & Van der Linden, 2011; Soto & Wasserman, 2011; but see also Martinez, 2003). It has also been proposed that facial expression recognition might be more efficient than facial identity memory or perception (Bankó, Gál, & Vidnyánszky, 2009). The ability to detect very small changes in the facial expression of emotion has been reported, among others, by Neth and Martinez (2009). What is interesting for the present discussion is that, in their study, observers were more likely to correctly report a difference between two sequentially-presented images of faces if the first image was closer to the mean face for a specific facial identity and the second was distant from that mean face, rather than the opposite. This result is consistent with the idea that the mean of an "identity cluster" is more likely to be recovered from WM than any individual exemplar. In fact, the closer the first image was to the mean facial identity, the smaller were the memory distortions. Note also that Neth and Martinez's results were found with images of faces displaying positive and negative expressions — not with faces with a neutral expression, as in the present study.

5.4.1. Evidence from upside-down faces

Results consistent with Neth and Martinez (2009) have also been provided by Caudek and Lorenzino (2012). In a preliminary phase of their experiments, pairs of upright or upside-down morphed faces were selected to be equally perceptually discriminable in a simultaneous matching task. The faces pairs were drawn from morph continua between two expressions or two identities. The selected face pairs were then used in a delayed matching task and the strength of the Face Inversion Effect (FIE) was measured.[5] For subtle variations in facial affect, recognition performance was better for upright than for upside-down faces. Instead, for subtle variations of face images displaying a neutral expression, the FIE reversed its direction: Caudek and Lorenzino found better recognition performance for upside-down than for upright faces. In accord with the present results, Caudek and Lorenzino's findings suggest that WM can represent with high fidelity fine-grained information about facial expressions, but not fine-grained information about faces with a neutral expression.

5.5. Conclusions

The present study utilized signal detection methodology to examine the effects of perceptual discriminability and memory load on recognition accuracy. The results obtained with objects of non-expertise (cars) confirm that increased memory load is associated with a decrement in recognition accuracy. The novel result of the present study is that, under specific stimulus conditions, the memory for multiple faces can exceed that of a single face. This result may be due to the low memorial discriminability of the transient properties of faces with a neutral expression, which may serve the purpose of enhancing recognition at the individual-exemplar level.

Acknowledgments

I would like to thank Tom Busey and two anonymous reviewers for helpful comments on an earlier version of this manuscript. I also thank David Badre and Fulvio Domini for helpful discussions, and Elisa Cianfrini, Camilla Corezzi, Sara Franceschi, Ilaria Lupi, Martina Lorenzino, Antonella Riillo, and Stefano Sabbatucci for their help in developing the stimuli and collecting data.

Appendix A. Fitting the summed-similarity model

According to the summed-similarity model, each member of the memory array is stored as a distinct exemplar in memory. When the memory probe i is observed, all the exemplars are activated in proportion to their similarity to the probe (e.g., Knapp et al., 2006). The greater the summed similarity between i and the stored exemplars, the greater is the probability with which the observer responds "old." Recognition choice probabilities are a positive function of the global familiarity F_i:

$$P(\text{say old}|i) = f(F_i). \tag{A.1}$$

The familiarity of the probe i is given by the summed similarity of i to each item j of the memory array:

$$F_i = \sum_j s(i,j), \tag{A.2}$$

where $s(i, j)$ is the similarity of the probe i to the jth stored exemplar. The similarity $s(i, j)$ is assumed to be an exponential decay function of the perceptual distance between i and j (e.g., Busey & Arici, 2009),

$$s(i,j) = exp[-\kappa \cdot d(i,j)]. \tag{A.3}$$

Finally, the similarity of the probe to itself (as a stored exemplar) is defined as

$$s(i,i) = \gamma \tag{A.4}$$

where $\gamma > 0$ is a freely estimated parameter that measures the importance of common-feature matches and the distinctiveness of the probe i (Knapp et al., 2006).

For the purposes of the present investigation, the summed-similarity model was instantiated in the following form[6] to predict the logit of participants saying the probe was seen before (i.e., say old):

$$log\frac{P\left(\text{say old}\middle|M, S_j\right)}{1-P\left(\text{say old}\middle|M, S_j\right)} = \alpha + \gamma M + \sum_j \beta_j S_j, \tag{A.5}$$

with $j = 1,\ldots,4$, where M is an indicator variable whose role is to indicate whether there is ($M = 1$) or there is not ($M = 0$) a perfect match between the probe and a study item, and S_j is an estimate of the similarity of the probe i to the jth stored exemplar. For the present purposes, S_j was estimated as $\exp[-d(i, j)]$, with $d(i, j)$ being the distance on the morph continuum between the probe i and the jth studied item (see Table 1). The parameter γ of Eq. (A.5) estimates $s(i, i)$ of Eq. (A.4), the parameters β_j estimate the contribution of $s(i, j)$ of Eq. (A.3) to the probability of participants saying the probe was seen before, and α is a parameter which modulates the overall probability of an "yes/match" response.[7]

---

[5] The FIE, which is one of the landmarks of holistic face processing, indicates that turning an image upside down impairs discrimination to a larger extent for faces than for non-face objects (Yin, 1969).

[6] Different instantiations of the summed-similarity model have been proposed and the relative fit of these different implementations has been discussed (e.g., Busey & Arici, 2009). These issues, however, are beyond the scope of the present investigation.

[7] Several methodological differences distinguish the present implementation of the summed-similarity model from previous work. In most previous studies (1) the probe remained visible until the subject's response was recorded, (2) feedback was often provided, and (3) the perceptual differences between the probe and the study items were well above the threshold of perceptual discrimination. In the present case, a short presentation time was chosen to limit local processing and to favor face processing. No feedback on correct judgments was provided in order to reduce learning effects, given that the goal was to study the biases in WM. Finally, stimulus displays with subtle image changes were used (rather than easily discriminable items), because the goal was to investigate the fidelity of WM.

References

Afraz, A., Vaziri-Pashkam, M., & Cavanagh, P. (2010). Spatial heterogeneity in the perception of face and form attributes. *Current Biology*, *20*, 2112–2116.
Alvarez, G. A. (2011). Representing multiple objects as an ensemble enhances visual cognition. *Trends in Cognitive Sciences*, *15*, 122–131.
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.
Anaki, D., & Bentin, S. (2009). Familiarity effects in the categorization levels of faces and objects. *Cognition*, *111*, 144–149.
Awh, E., & Jonides, J. (2001). Overlapping mechanisms of attention and spatial working memory. *Trends in Cognitive Sciences*, *5*, 119–126.
Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*, 390–412.
Bankó, É. M., Gál, V., & Vidnyánszky, Z. (2009). Flawless visual short-term memory for facial emotional expressions. *Journal of Vision*, *9*(1), 1–13 (12).
Bates, D. M., & Sarkar, D. (2007). *lme4: Linear mixed-effects models using S4 classes.* R package version 0.999999-0.
Bays, P. M., Catalao, R. F. G., & Husain, M. (2009). The precision of visual working memory is set by allocation of a shared resource. *Journal of Vision*, *9*(10), 1–11 (7).
Bays, P. M., & Husain, M. (2008). Dynamic shifts of limited working memory resources in human vision. *Science*, *321*, 851–854.
Bays, P. M., Wu, E. Y., & Husain, M. (2011). Storage and binding of object features in visual working memory. *Neuropsychologia*, *49*, 1622–1631.
Blank, I., & Yovel, G. (2011). The structure of face-space is tolerant to lighting and viewpoint transformations. *Journal of Vision*, *11*(8), 1–13 (15).
Brady, T. F., Konkle, T., & Alvarez, G. A. (2011). A review of visual memory capacity: Beyond individual items and toward structured representations. *Journal of Vision*, *11*(5), 1–34 (4).
Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, *10*, 433–436.
Bruce, V., & Young, A. (1986). Understanding face recognition. *British Journal of Psychology*, *77*, 305–327.
Busey, T. A. (1998). Physical and psychological representations of faces: Evidence from morphing. *Psychological Science*, *9*, 476–483.
Busey, T. A., & Arici, A. (2009). On the role of individual items in recognition memory and metacognition: Challenges for signal detection theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*, 1123–1136.
Busey, T. A., & Tunnicliff, J. (1999). Accounts of blending, typicality and distinctiveness in face recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*, 1210–1235.
Calder, A. J., & Young, A. W. (2005). Understanding facial identity and facial expression recognition. *Nature Neuroscience Reviews*, *6*, 641–653.
Caudek, C., & Domini, F. (2013). Priming effects under correct change detection and change blindness. *Consciousness and Cognition.* http://dx.doi.org/10.1016/j.concog.2012.08.003.
Caudek, C., & Lorenzino, M. (2012). Recognition memory is more accurate when faces are inverted than when they are upright. *Journal of Vision*, *12*(9). http://dx.doi.org/10.1167/12.9.630 (article 630).
Corneille, O., Hugenberg, K., & Potter, Y. (2007). Applying the attractor field model to social cognition: Perceptual discrimination is facilitated but memory is impaired for faces displaying evaluatively-congruent expressions. *Journal of Personality and Social Psychology*, *93*, 335–352.
Cowan, N. (2006). *Working memory capacity.* New York, NY: Psychology Press.
Curby, K. M., & Gauthier, I. (2007). A visual short-term memory advantage for faces. *Psychonomic Bulletin and Review*, *14*, 620–628.
D'Argembeau, A., & Van der Linden, M. (2011). Influence of facial expression on memory for facial identity: Effects of visual features or emotional meaning? *Emotion*, *11*, 199–202.
D'Lauro, C., Tanaka, J., & Curran, T. (2008). The preferred level of face categorization depends on discriminability. *Psychonomic Bulletin & Review*, *15*, 623–629.
DeCarlo, L. T. (1998). Signal detection theory and generalized linear models. *Psychological Methods*, *3*, 186–205.
Emrich, S. M., Al-Aidroos, N., Pratt, J., & Ferber, S. (2010). Finding memory in search: The effect of visual working memory load on visual search. *Quarterly Journal of Experimental Psychology*, *63*, 1457–1466.
Folstein, J. R., Palmeri, T. J., & Gauthier, I. (2012). Category learning increases discriminability of relevant object dimensions in visual cortex. *Cerebral Cortex.* http://dx.doi.org/10.1093/cercor/bhs067.
Gauthier, I., & Tarr, M. J. (1997). Becoming a "Greeble" expert: Exploring mechanisms for face recognition. *Vision Research*, *37*, 1673–1682.
Goldstone, R. L. (1994). Influences of categorization on perceptual discrimination. *Journal of Experimental Psychology. General*, *123*, 178–200.
Goldstone, R. L., & Steyvers, M. (2001). The sensitization and differentiation of dimensions during category learning. *Journal of Experimental Psychology. General*, *130*, 116–139.
Gureckis, T. M., & Goldstone, R. L. (2008). The effect of the internal structure of categories on perception. In B. C. Love, K. McRae, & V. M. Sloutskey (Eds.), *Proceedings of the 30th Annual Conference of the Cognitive Science Society, 2008 Jul 23-26* (pp. 1876–1881). Austin (TX): Cognitive Science Society.
Haberman, J., & Whitney, D. (2012). Ensemble perception: Summarizing the scene and broadening the limits of visual processing. In J. Wolfe, & L. Robertson (Eds.), *From perception to consciousness: Searching with Anne Treisman*. Oxford University Press.
Hancock, P., Burton, A., & Bruce, V. (1996). Face processing: Human perception and principal components analysis. *Memory and Cognition*, *24*, 26–40.
Hegde, J. (2008). Time course of visual perception: Coarse-to-fine processing and beyond. *Progress in Neurobiology*, *84*, 405–439.
Hollingworth, A., & Henderson, J. M. (2002). Accurate visual memory for previously attended objects in natural scenes. *Journal of Experimental Psychology. Human Perception and Performance*, *28*, 113–136.
Huang, J., & Sekuler, R. (2010). Distortions in recall from visual memory: Two classes of attractors at work. *Journal of Vision*, *10*(2), 1–27 (24).
Jackson, M. C., & Raymond, J. E. (2008). Familiarity enhances visual working memory for faces. *Journal of Experimental Psychology. Human Perception and Performance*, *34*, 556–568.
Jiang, Y., Kwon, M., Shim, W. M., & Won, B. Y. (2010). Redundancy effects in the perception and memory of visual objects. *Visual Cognition*, *18*, 1233–1252.
Johnson, K., & Mervis, C. (1997). Effects of varying levels of expertise on the basic-level of categorization. *Journal of Experimental Psychology. General*, *126*, 248–277.
Judd, C. M., Westfall, J., & Kenny, D. A. (2012). Treating stimuli as a random factor in social psychology: A new and comprehensive solution to a pervasive but largely ignored problem. *Journal of Personality and Social Psychology*, *103*, 54–69.
Kahana, M. J., & Sekuler, R. (2002). Recognizing spatial patterns: A noisy exemplar approach. *Vision Research*, *42*, 2177–2192.
Kanwisher, N. (2000). Domain specificity in face perception. *Nature Neuroscience*, *8*, 759–763.
Knapp, B. R., Nosofsky, R. M., & Busey, T. (2006). Recognizing distinctive faces: A hybrid-similarity exemplar-model account. *Memory & Cognition*, *34*, 877–889.
Lee, K. J., & Perrett, D. I. (2000). Manipulation of color and shape information and its consequence upon recognition and best-likeness judgments. *Perception*, *29*, 1291–1312.
Leopold, D. A., Rhodes, G., Müller, K. M., & Jeffery, L. (2005). The dynamics of visual adaptation to faces. *Proceedings of the Royal Society of London. Series B*, *272*, 897–904.
Louie, E. G., Bressler, D. W., & Whitney, D. (2007). Holistic crowding: Selective interference between configural representations of faces in crowded scenes. *Journal of Vision*, *7*(2), 1–11 (24).
Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, *309*, 279–281.
Macchi Cassia, V., Picozzi, M., Kuefner, D., Bricolo, E., & Turati, C. (2008). Holistic processing for faces and cars in preschool-aged children and adults: Evidence from the composite effect. *Developmental Science*, *12*, 236–248.
Magnussen, S. (2000). Low-level memory processes in vision. *Trends in Neurosciences*, *23*, 247–251.
Martelli, M., Majaj, N. J., & Pelli, D. G. (2005). Are faces processed like words? A diagnostic test for recognition by parts. *Journal of Vision*, *5*, 58–70.
Martinez, A. M. (2003). Matching expression variant faces. *Vision Research*, *43*, 1047–1060.
McGugin, R. W., Tanaka, J. W., Lebrecht, S., Tarr, M. J., & Gauthier, I. (2011). Race-specific perceptual discrimination improvement following short individuation training with faces. *Cognitive Science*, *35*, 330–347.
Melcher, D. (2001). Persistence of visual memory for scenes — A medium-term memory may help us to keep track of objects during visual tasks. *Nature*, *412*, 401-401.
Melcher, D. (2006). Accumulation and persistence of memory for natural scenes. *Journal of Vision*, *6*, 8–17.
Morgan, H. M., Klein, C., Boehm, S. G., Shapiro, K. L., & Linden, D. E. J. (2008). Working memory load for faces modulates P300, N170, and N250r. *Journal of Cognitive Neuroscience*, *20*, 989–1002.
Neth, D., & Martinez, A. M. (2009). Emotion perception in emotionless face images suggests a norm-based representation. *Journal of Vision*, *9*(1), 1–11 (5).
Nishimura, M., & Maurer, D. (2008). The effect of categorisation on sensitivity to second order relations in novel objects. *Perception*, *37*, 584–601.
Nosofsky, R. M. (1988). Exemplar-based accounts of relations between classification, recognition, and typicality. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*, 700–708.
Nosofsky, R. M., Little, D. L., Donkin, C., & Fific, M. (2011). Short-term memory scanning viewed as exemplar-based categorization. *Psychological Review*, *118*, 280–315.
Nosofsky, R. M., Little, D. R., & James, T. W. (2012). Activation in the neural network responsible for categorization and recognition reflects parameter changes. *Proceedings of the National Academy of Sciences*, *109*, 333–338.
Notman, L. A., Sowden, P. T., & Özgen, E. (2005). The nature of learned categorical perception effects: a psychophysical approach. *Cognition*, *95*, B1–B14.
Op de Beeck, H., Wagemans, J., & Vogels, R. (2003). The effect of category learning on the representation of shape: Dimensions can be biased, but not differentiated. *Journal of Experimental Psychology. General*, *132*, 491–511.
Özgen, E., & Davies, I. R. L. (2002). Acquisition of categorical color perception: A perceptual learning approach to the linguistic relativity hypothesis. *Journal of Experimental Psychology. General*, *131*, 477–493.
Pallett, P. M., & MacLeod, D. I. (2011). Seeing faces as objects: No face inversion effect with geometrical discrimination. *Attention, Perception, & Psychophysics*, *73*, 504–520.
Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*, 437–442.
RDevelopment Core Team (2012). *R: A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing3-900051-07-0 (URL , http://www.R-project.org)
Rensink, R. A., O'Regan, J. K., & Clark, J. J. (1997). To see or not to see: The need for attention to perceive changes in scenes. *Psychological Science*, *8*, 368–373.
Richler, J. J., Mack, M. L., Gauthier, I., & Palmeri, T. J. (2009). Holistic processing of faces happens at a glance. *Vision Research*, *49*, 2856–2861.
Rosch, E. H., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, *8*, 382–439.

Scott, L., Tanaka, J. W., Sheinberg, D., & Curran, T. (2006). A reevaluation of the electrophysiological correlates of expert object processing. *Journal of Cognitive Neuroscience*, *18*, 1453–1465.

Sekuler, R., & Kahana, M. J. (2007). A stimulus-oriented approach to memory. *Current Directions in Psychological Science*, *16*, 305–310.

Simons, D. J., & Rensink, R. A. (2005). Change blindness: Past, present, and future. *Trends in Cognitive Sciences*, *9*, 16–20.

Soto, F. A., & Wasserman, E. A. (2011). Asymmetrical interactions in the perception of face identity and emotional expression are not unique to the primate visual system. *Journal of Vision*, *11*(3), 1–18 (24).

Tanaka, J. W., Curran, T., & Sheinberg, D. (2005). The training and transfer of real world perceptual expertise. *Psychological Science*, *16*, 145–151.

Tanaka, J., & Taylor, M. (1991). Object categories and expertise: Is the basic-level in the eye of the beholder? *Cognitive Psychology*, *23*, 457–482.

Valentine, T. (1991). Representation and process in face recognition. In: R. Watt (ed.) *Pattern recognition by Man and Machine*. (Vol. 14 in 'Vision and Visual Dysfunction' series edited by J. Cronly-Dillon). London: Macmillan Press.

Vogel, E. K., & Machizawa, M. G. (2004). Neural activity predicts individual differences in visual working memory capacity. *Nature*, *428*, 748–751.

Wagar, B. M., & Dixon, M. J. (2005). Past experience influences object representation in working memory. *Brain and Cognition*, *57*, 248–256.

Webster, M. A., Kaping, D., Mizokami, Y., & Duhamel, P. (2004). Adaptation to natural facial categories. *Nature*, *428*, 557–561.

Wilken, P., & Ma, W. J. (2004). A detection theory account of change detection. *Journal of Vision*, *4*(12), 1120–1135 (11).

Wright, D. B., Horry, R., & Skagerberg, E. M. (2009). Functions for traditional and multilevel approaches to signal detection theory. *Behavior Research Methods*, *41*, 257–267.

Wright, D. B., & London, K. (2009). Multilevel modelling: Beyond the basic applications. *British Journal of Mathematical and Statistical Psychology*, *62*, 439–456.

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

Zaki, S. R., & Nosofsky, R. M. (2001). Exemplar accounts of blending and distinctiveness effects in perceptual old–new recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*, 1022–1041.

Zchaluk, K., & Foster, D. H. (2009). Model-free estimation of the psychometric function. *Attention, Perception, & Psychophysics*, *71*, 1414–1425.

Zhou, F., Kahana, M. J., & Sekuler, R. (2004). Short-term episodic memory for visual textures: A roving probe gathers some memory. *Psychological Science*, *15*, 112–118.