Attention, Perception, & Psychophysics (2022) 84:638–646
https://doi.org/10.3758/s13414-022-02442-1

SHORT REPORT

Revealing a competitive dynamic in rapid categorization with object substitution masking

Jason K. Chow[1] · Thomas J. Palmeri[1] · Michael L. Mack[2]

Accepted: 5 January 2022 / Published online: 23 February 2022

Abstract

Categorization at different levels of abstraction have distinct time courses, but the different levels are often considered separately. Superordinate-level categorization is typically faster than basic-level categorization at ultra-rapid exposure durations (< 33 ms) while basic-level categorization is faster than superordinate-level categorization at longer exposure durations. This difference may be due to a competitive dynamic between levels of categorization. By leveraging object substitution masking, we found a distinct time course of masking effects for each level of categorization. Superordinate-level categorization showed a masking effect earlier than basic-level categorization. However, when basic-level categorization first showed a masking effects, superordinate-level categorization was spared despite its earlier masking effect. This unique pattern suggests a trade-off between the two levels of categorization over time. Such an effect supports an account of categorization that depends on the interaction of perceptual encoding, selective attention, and competition between levels of category representation.

Keywords Object categorization · Object substitution masking · Time course · Temporal dynamics

Introduction

Categorization, the ability to rapidly link perceptual input from the environment onto semantic knowledge, is a fundamental aspect of human experience. This ability is marked by a balance of detecting the regularities and co-occurrences of object features (e.g., dogs have four legs, fur, and a snout; birds have two legs, feathers, and a beak) and cognitive utility (e.g., labeling an object as a dog as opposed to an animal or Golden Retriever is most informative to most situations; Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976). In fact, most objects are most quickly and accurately categorized at an intermediate level of abstraction in the semantic hierarchy known as the basic level (e.g., dog). However, several factors can influence this entry level of categorization including development (Mandler, Bauer, & McDonough, 1991), expertise in visual domains (e.g., Tanaka & Taylor, 1991; Johnson & Mervis, 1997), and social relevance (e.g., Mack & Palmeri, 2010). Most strikingly, when exposure to visual information is severely limited, objects are more easily categorized at superordinate levels relative to basic levels (e.g., animal instead of dog; Macé, Joubert, Nespoulous, & Fabre-Thorpe, 2009; Mack & Palmeri, 2015; Thorpe, Fize, & Marlot, 1996; Vanmarcke, Calders, & Wagemans, 2016). Here, we extend this research to ask if interactions across the levels of the semantic hierarchy itself play an important role in the time course of rapid visual categorization with a novel application of object substitution masking (OSM; Enns & Di Lollo, 1997).

Research on the time course of visual categorization is dominated by carefully controlled studies that limit exposure to visual stimuli with brief presentations and backward masking (e.g., Bacon-Macé, Macé, Fabre-Thorpe, & Thorpe, 2005; Bacon-Macé, Kirchner, Fabre-Thorpe, & Thorpe, 2007; Fabre-Thorpe, 2011; Macé et al., 2009; Mack & Palmeri, 2011; Mack & Palmeri, 2015; Thorpe et al., 1996). By limiting the duration of visual input and effectively halting further perceptual processing with a backward visual pattern mask (a spatially overlapping meaningless image appearing after the inputs), these paradigms have revealed important mechanistic trade-offs for categorization at different levels of abstraction. However, there is a

✉ Michael L. Mack
michael.mack@utoronto.ca

[1] Department of Psychology, Vanderbilt University, Nashville, TN, USA

[2] Department of Psychology, University of Toronto, Toronto, ON, Canada

common assumption in this literature that visual pattern masks serve simply to erase visual input from further processing. Theories on backward masking posit that backward masks may, in fact, have more complex effects on visual processing (Breitmeyer & Ogmen, 2000; 2006; Smith et al., 2004). Rather than just stopping visual processing in its tracks after a specified duration, the barrage of visual input from a backward pattern mask may supplant the encoded perceptual representations of visual objects and disrupt downstream categorization processes (Ratcliff R & Rouder JN, 2000). If so, time course manipulations in past work may have altered perceptual encoding with distinct impacts across the time course for categorization at different levels of abstraction. At a minimum, it is prudent to examine the time course of visual categorization with alternative masking methods.

Here, we leverage OSM in a novel manner to characterize the time course of basic- and superordinate-level categorization. OSM arises from the brief presentation of an image surrounded by simple visual shapes, typically small squares or circles. The shapes acts as the cue to attend to (and usually make responses for) the central image. When the offset of the surrounding shapes is delayed relative to the central image, a masking effect is observed such that accuracy or speed for reporting some aspect of the central image is significantly reduced. Rather than disrupting the formation of perceptual representations, initial reports suggested OSM results from competitive dynamics between object-based representations of the cued object and the surrounding shapes (e.g., Di Lollo et al., 2000, Lleras & Moore, 2003, Moore & Lleras, 2005). Indeed, recent studies further support the notion that OSM disrupts higher-level representations of visual input (Harrison et al., 2016, Goodhew et al., 2013, Goodhew, 2017). For categorization, OSM reduces perceptual awareness and increases reaction times for categorization at the superordinate level (Koivisto et al., 2014). Thus, OSM offers an ideal paradigm for characterizing the time course of visual categorization without altering the encoding of visual information. By selectively disrupting higher-level representations at different times, OSM can interrogate how these representations may serve both superordinate- and basic-level categorization simultaneously and how they interact. Across two studies (one a replication) we find evidence for a competitive dynamic between rapid superordinate- and basic-level categorization.

Methods

Participants and materials

In Experiment 1, 34 participants (21 female, 13 males; 18–23 years old, average 19.1 years old) were recruited from the Vanderbilt University participant pool for course credit. One participant was excluded for failing to respond or responding faster than 150 ms on at least 15% of trials leaving a total sample size of 33. All participants reported normal or corrected-to-normal vision. The target sample size for Experiment 1 (N = 30) was based on sample sizes tested in prior studies of object categorization with backward masking (e.g., Mack & Palmeri, 2015). Procedures were approved by the Vanderbilt Institutional Review Board. Informed written consent was obtained before the experiment.

In Experiment 2, 81 participants (56 female, 25 males; 17-30 years old, average 18.8 years old, two participants did not report age) were recruited from the University of Toronto participant pool for course credit. Seven participants were excluded for failing to respond or responding faster than 150 ms on at least 15% of trials leaving a total sample size of 74. All participants reported normal or corrected-to-normal vision. The target sample size for Experiment 2 (N = 70) was based on two factors: 1) a power analysis of the change in masking effect observed in Experiment 1 (see Fig. 3; Cohen's f=0.387, α=0.05, β=0.05), which suggested a target sample size of 36, and 2) the expectation that less precise display equipment in Experiment 2 would add noise to our behavioral measures. Considering both factors, we doubled the target sample size of the power analysis (N = 72). Procedures were approved by the University of Toronto Research Ethics Board. Informed written consent was obtained before the experiment.

Stimuli were collected from online image searches for four basic-level categories (birds, dogs, planes, and cars) and two superordinate-level categories (animals and vehicles). Each category had 144 unique images and the superordinate-level categories did not include objects that could appear in the basic-level categories. Images with backgrounds were roughly cropped with the figure centered and scaled to 120 x 120 pixels (4.3°x4.3° of visual angle) then converted to greyscale. For Experiment 1, stimuli were presented on a 19" Sony Trinitron CRT monitor with a refresh rate of 120 Hz. For Experiment 2, stimuli were presented on a 24" AOC LCD monitor (G2460PQU) with a refresh rate of 120 Hz. In both experiments, participants were placed such that the visual array subtended 18.5° of visual angle. Both experiments were programmed and ran in Matlab using Psychtoolbox3 (Brainard, 1997; Kleiner et al., 2007; Pelli, 1997).

Procedure

The same general procedure was used for both experiments. Participants were tasked with verifying whether a label matched a cued image in a two-alternate forced-choice task

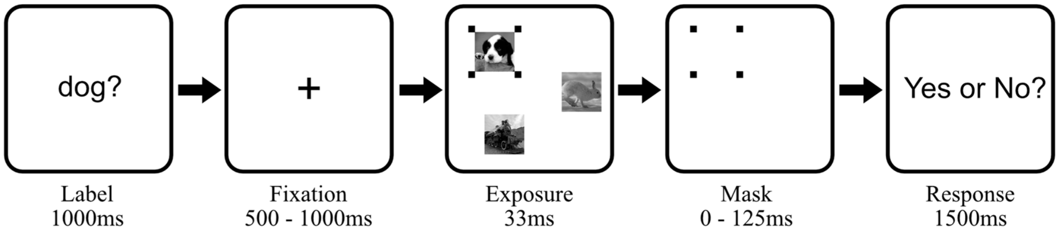

Fig. 1 Trial schematic. Images were presented equally spaced on an invisible circle with a radius of 200 px in a pseudo-random configuration

(Fig. 1). Each trial began with either a superordinate-level label (vehicle or animal) or a basic-level label (car, plane, dog, or bird) for 1000 ms. This was followed by a fixation cross for 500-1000 ms, sampled from a uniform distribution. The target and two distractor images were presented on an invisible circle with a radius of 200 pixels (7.3° of visual angle) at the center of the screen for 33 ms. Each image was equally spaced from one another. Four black 20 x 20 pixel squares (0.73°x0.73° of visual angle) were presented at the corners of the target image simultaneously with the stimulus onset. Critically, the offset of the black squares was delayed by 0, 17, 33, 50, 68, or 125 ms after the target offset. Mask offset times were randomized across trials. These black squares would act as the cue for the target and if the offset delay was greater than 0, also act as a mask. Participants were instructed to make a keyboard response as quickly and accurately as possible after stimuli onset indicating whether the label and the cued target matched with either yes or no (using the 1 or 2 key, respectively). No corrective feedback was given.

The study began with paper instructions followed by five practice trials. Then, participants performed six blocks of 96 trials each for a total of 576 trials. Each block consisted only of labels from one level of categorization and the blocks were grouped such that participants completed all blocks of one level of categorization before moving onto the other blocks with the order counterbalanced across participants. Participants were offered a short break between each block. Trial order within blocks was randomized for each participant. Each trial included a target image from one of the four basic-level categories and two distractor images, one from each superordinate-level category. Images were pseudo-randomly selected from their respective categories. The target images never repeated and distractor images repeated four times each. Half of the trials had images that matched the label and half did not.

Experiment 2 replicated Experiment 1 with a larger sample size using the same materials and followed similar procedures with the following exceptions: stimulus exposure was increased to 50 ms[1] and mask offset durations were limited to 0, 17, 33, 50, and 68 ms. We omitted the 125 ms offset to both focus on the mask offsets from Experiment 1 that demonstrated the first masking effects across categorization levels and to reduce the overall experiment duration. Additionally, the trial structure was reduced to four blocks (two blocks each of superordinate- and basic-level categorization) of 120 trials each. These paradigm changes provided a more efficient and feasible experiment that targeted a replication of Experiment 1's primary results in a larger sample of participants.

Analysis

We characterized categorization performance by estimating sensitivity (d′) according to standard signal detection theory methods (Green & Swets, 1966). Hits, the proportion of yes responses to trials in which the cued object matched the specified label, and false alarms, the proportion of yes responses to trial in which the cue object did not match the specified label, were calculated separately for each mask offset and category level (see Fig. 2). Our analyses focused on the masking effect for conditions with mask offsets longer than 0 ms. Specifically, we took the difference between d′ at mask offset of 0 ms versus all other mask offsets (see Fig. 3). We also characterized the speed of responses by analyzing response times for hits. Statistical analyses were conducted in R (version 4.0.2) with rstanarm (version 2.21.1) using Bayesian estimation of mixed effects linear regression. The existence of effects were evaluated by the probability of direction (*pd*) which

[1]Stimulus exposure was lengthened to best match the stimulus intensity of the CRT monitors used in Experiment 1. Although the LCD monitors used in Experiment 2 are gaming-style monitors with fast refresh rates (120 Hz) and response times (2 ms grey-to-grey), the amount of luminance change the monitors can provide with very brief exposures is limited. As such, we increased the exposure duration of the stimulus array to best match performance at the 0-ms mask offset condition.

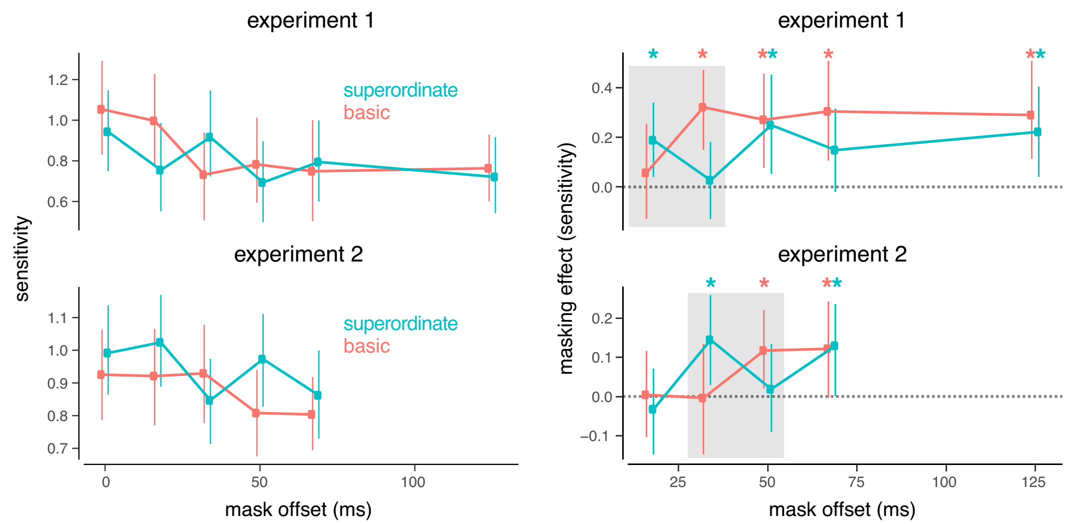

Fig. 2 Categorization d′ and masking effects. d′ (*left*) for Experiments 1 (*top*) and 2 (*bottom*) is plotted according to mask offset delay (ms) separately for superordinate- (*cyan*) and basic-level (*magenta*) categorization. The masking effect (*right*), calculated separately for each participant as the difference between d′ at mask offset of 0 ms versus all other mask offsets, is plotted according to the same conventions. A positive masking effect means performance was lower than performance at the 0-ms mask offset. Error bars represent bootstrapped 95% confidence intervals of the mean. *Asterisks* note significant (pd>0.95) masking effects relative to 0-ms mask offset. The *gray boxes* note time points used in the analysis presented in Fig. 3

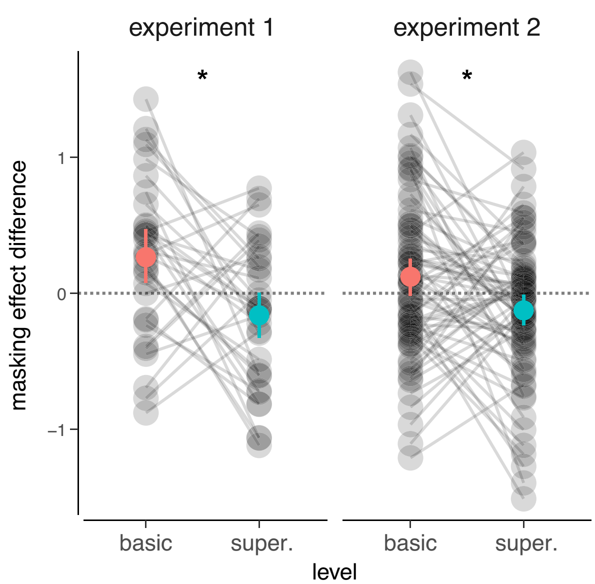

Fig. 3 Change in masking effect at short mask offsets. The change in masking effect between the shortest mask offsets showing significant masking effects (Experiment 1: 17 and 33 ms, Experiment 2: 33 and 50 ms) is plotted separately for basic- and superordinate-level (super.) categorization. Positive values correspond to an increase in masking effect with longer mask offsets, negative values correspond to a decrease in masking effects. Participant-specific masking effects are depicted with transparent *points* and *lines*. *Colored points* and *error bars* represent group means and bootstrapped 95% confidence intervals. *Asterisks* note significant differences (pd>0.95) between basic- and superordinate-level conditions

quantifies the proportion of posterior samples in the most probable direction (i.e., pd ranges from 0.5 to 1 with values closer to 1 for more likely effects). pd is akin to the frequentist p value and is best interpreted as the degree of evidence against a null effect (Makowski, Ben-Shachar, Chen, & Lüdecke, 2019). In addition to pd, we report median values (β) and 95% confidence intervals (CI) for each effect of interest.

Result

For Experiment 1, the mixed-effects linear regression of d′ revealed significant masking effects across both categorization levels. Figure 2 shows both the average d′ across mask offsets (left), as well as the masking effect (right) which was calculated as the difference in d′ between the 0-ms mask offset conditions and each other mask offsets (see also Table 1). Positive values of masking effect correspond to lower performance for a mask offset relative to the 0-ms mask offset baseline. For basic-level categorization, mask offsets at 33 ms (β=-0.312, CI=[-0.5,-0.144], pd=0.998), 50 ms (β=-0.269, CI=[-0.456,-0.091], pd=0.992), 68 ms (β=-0.302, CI=[-0.488,-0.124], pd=0.994), and 125 ms (β=-0.29, CI=[-0.462,-0.091], pd=0.994) led to significantly lower d′ than in the 0-ms mask offset condition. For superordinate-level categorization, the pattern differed: mask offsets at 17 ms (β=-0.188, CI=[-0.363,-0.011], pd=0.957), 50 ms (β=-0.247, CI=[-0.429,-0.071], *pd*=0.984), and 125 ms (β=-0.221, CI=[-0.388,-0.039], *pd*=0.98) resulted in significantly lower d′ than in the 0-ms mask offset condition. When compared directly, the masking effect was reliably higher for basic- than superordinate-level categorization at 33-ms mask offset (*pd*=0.993).

For Experiment 2, the mixed-effects linear regression of d′ revealed similar effects as Experiment 1, but with critical mask offsets shifted towards longer durations (Fig. 2, bottom row). For basic-level categorization, d′ was significantly lower than baseline for mask offsets at 50 ms (β=-0.117, CI=[-0.219,-0.001], *pd*=0.953) and 68 ms (β=-0.121, CI=[-0.234,-0.019], *pd*=0.963). For superordinate-level categorization, masking effects were observed for mask offsets at 33 ms (β=-0.145, CI=[-0.248,-0.034], *pd*=0.983) and 68 ms (β=-0.13, CI=[-0.225,-0.011], *pd*=0.973). The masking effect was reliably higher for superordinate- than basic-level categorization at 33-ms mask offset (*pd*=0.975).

For both experiments, mixed-effects linear regression models of hit response times revealed no significant masking effects (Experiment 1: all *pd*s<0.94; Experiment 2: all *pd*s<0.78) nor significant differences between levels of abstraction (Experiment 1: *pd*=0.89; Experiment 2: *pd*=0.94).

Both experiments demonstrated a similar overall pattern of masking effects, but with a temporal shift towards longer offsets in Experiment 2. Given that the two experiments included data from different participant pools and utilized different experimental equipment, such a shift in the baseline effect may be expected. As such, we performed a post hoc analysis to temporally align the masking effect patterns by focusing on the earliest mask offsets within each experiment that demonstrated a significant masking effect relative to the 0ms mask offset (see gray boxes in Fig. 2). In Experiment 1, this corresponded to mask offsets at 17 and 33 ms; for Experiment 2, this was the mask offsets at 33 and 50 ms. In particular, we were interested in the apparent trade-off in performance between superordinate- and basic-level categorization due to selective masking effects. To examine this trade-off, we calculated the difference in masking effects between the two mask offsets for the two levels of categorization (Fig. 3). In both experiments, an increased masking effect for basic-level categorization was coupled with a decreased masking effect for superordinate-level categorization (Experiment 1: β=-0.427, CI=[-0.625,-0.196], *pd*>0.999; Experiment 2: β=-0.245, CI=[-0.392,-0.093], *pd*=0.995. In both experiments, a majority of participants showed this trade-off in categorization performance (Experiment 1: 23/33, 70%; Experiment 2: 49/74, 66%).

Discussion

Here, we leveraged an OSM paradigm to target higher-level processes that link encoded visual information to object knowledge without the introduction of a spatially overlapping mask. In doing so, we find that superordinate-level categorization performance is significantly affected at earlier time points during the time course of categorization with a masking effect on basic-level categorization appearing later, similar to previous reports with backward masking (Mack & Palmeri, 2015; Macé et al., 2009). Additionally, in a novel departure from prior work, we demonstrate that the masking effect on superordinate-level categorization is eliminated at the first time point exhibiting an OSM effect on basic-level categorization. Importantly, this masking effect trade-off between superordinate- and basic-level categorization was replicated in a second study that targeted the earliest time window of categorization in a much larger sample of participants.

In characterizing the time course of categorization with an OSM paradigm, our findings are largely consistent with previous work. That superordinate-level categorization was impacted by OSM at earlier time points converges well with the rich literature on ultra-rapid superordinate categorization. Indeed, several decades of research point to important behavioral and neural signatures of fast categorization processes for superordinate categories (Thorpe et al., 1996; VanRullen & Thorpe, 2001b; 2001a; Fabre-Thorpe, 2011; Vanmarcke et al., 2016), with evidence showing a distinct speed advantage for superordinate- relative to basic-level categorization (Macé et al., 2009). Reconciling these findings with the classic basic-level advantage (Rosch et al., 1976) has revealed important empirical factors like blocking trials by category level and brief exposure durations that lead to faster superordinate decisions (Mack & Palmeri, 2015), both of which were used in the current study. Thus, our findings that superordinate-level categorization was the first to demonstrate disruption from OSM offers further evidence for such rapid superordinate-level categorization processes.

Basic-level categorization did show a strong masking effect in both studies; however, this effect was delayed relative to the masking effect in superordinate-level categorization. Surprisingly, the disruption to basic-level categorization was momentarily coupled with the elimination of the masking effect for superordinate-level categorization. This simultaneous sparing of superordinate-level categorization, notably occurring after a clear masking effect earlier in the time course, is not well explained in the context of leading current theories. Biologically-inspired models that

Table 1 Mean d′ values (95% CI) for each mask offset and category level in the two experiments

		mask offset (ms)					
	level	0	17	33	50	68	125
E1	basic	1.05 (.247)	0.997 (.238)	0.732 (.232)	0.783 (.218)	0.749 (.258)	0.763 (.179)
	superordinate	0.943 (.216)	0.755 (.226)	0.916 (.219)	0.694 (.209)	0.794 (.199)	0.721 (.191)
E2	basic	0.925 (.141)	0.921 (.152)	0.929 (.160)	0.808 (.131)	0.804 (.116)	
	superordinate	0.992 (.135)	1.02 (.134)	0.847 (.134)	0.973 (.143)	0.862 (.145)	

formalize feed-forward mechanisms of visual categorization (Serre et al., 2007) offer compelling descriptions of rapid categorization. Similarly, recent studies with this class of model suggest advantages for superordinate- versus basic-level categorization may arise due to key differences in representational discriminability within and between categories (Sofer et al., 2015). These findings converge with decades of research on cognitive models of categorization that posit distinct advantages at different levels of abstraction based on the time course of perceptual encoding (Cohen & Nosofsky, 2003; Lamberts, 2000) and similarity in multidimensional representational spaces (Nosofsky, 1986; 1988; Nosofsky & Palmeri, 1997; Palmeri, 1999; Mack & Palmeri, 2011; Love et al., 2004). Although both classes of models offer compelling accounts for speeded categorization decisions, neither explicitly define a competitive dynamic between levels of abstraction. One possible explanation is that the activation of representations in category knowledge are partly modulated by inhibition across levels of abstraction. When considered alongside the well-established centrality of basic-level categories in conceptual knowledge (e.g., Rosch et al., 1976; (Richler et al., 2011), 2011), a potential account emerges: Activation of basic-level representations may inhibit activation of an object's categories at different levels. If masking disrupts the activation of basic-level category representations, inhibition across levels would lessen and weakly activated superordinate-level category representations may have a greater impact on category decisions potentially leading to the type of masking effect recovery we see in the current work. This account is speculative, but the role of inhibition across category levels offers a compelling hypothesis for future work.

Another potential avenue for reconciling the current findings with extant cognitive theories is the emerging fruitful approach of combining convolutional neural networks with cognitive models (Annis et al., 2020). Recent work has demonstrated the distinct ability of recurrent deep neural networks to characterize representations throughout the ventral visual system (Kietzmann et al., 2019; Kar et al., 2019). By leveraging the ability to form representations over time, these networks can instantiate a representational bottleneck wherein category-specific information is most relevant first for superordinate-level features then giving way to basic-level features. We speculate that a recurrent network that experiences "masking" as disruption to recurrent connections across its layers may demonstrate a similar competitive dynamic favoring one categorization level over another as observed in our behavioral experiments. Integrating the representational dynamics of these sophisticated models of biological vision with well-established models of category learning and decision making (Love et al., 2004; Nosofsky & Palmeri, 1997; Usher & McClelland, 2001) could provide insight into how representations within and across different layers of the visual hierarchy influence category decisions (Annis et al., 2020) and how masking may shift the balance of decision evidence between different levels of abstraction.

Backward masking paradigms have played a major role in investigating the time course of object categorization, but other paradigms have offered important insights. One such example is a variant of the signal-to-respond paradigm in which participants verify that a visually presented object belongs to a category by making a yes or no response after a specified interval denoted by a signal cue (e.g., an auditory tone; Rogers & Patterson, 2007, Mack et al., 2009). In this paradigm, the interval between object presentation and response signal is manipulated to characterize performance across the time course of categorization. Importantly, no visual backward mask is used thus eliminating concerns with interfering visual processing of the visual mask stimulus. However, by dictating when responses are to be made, the signal-to-respond paradigm may in fact additionally interfere with or cut short processes related to decision making and response execution rather than simply limiting visual perception. In contrast, OSM potentially offers a more precise tool for isolating processes that link visual representations to object knowledge (Goodhew et al., 2013; Goodhew, 2017).

One notable methodological difference in the current study to related prior work (Bacon-Macé et al., 2005; Bacon-Macé et al., 2007; Fabre-Thorpe, 2011; Macé et al., 2009; Mack & Palmeri, 2011; 2015; Thorpe et al., 1996) is that target stimuli were presented peripheral to fixation and attention likely had to shift to the target to enable perceptual encoding of visual features important for categorization. Spatially distributed attention is considered a central aspect of the OSM paradigm (but see Filmer et al. 2015), thus was likely key to revealing the masking effects we observed. However, this initial attention shift did impact overall categorization performance–relative to prior work that used the similar stimuli presented centrally with brief exposure durations and backward masking (Mack & Palmeri, 2015), categorization performance in the current studies at the baseline condition (i.e., 0-ms mask offset) was notably lower. It may be that the attention shift in the OSM paradigm weakens the initially encoded perceptual representation. Although such weakening may make successful categorization depend more on the type of top-down processes OSM is thought to disrupt (Di Lollo et al., 2000; Goodhew et al., 2013; Goodhew, 2017), the impact of attention shifts in categorization should be further explored. Looking ahead, a combined approach with studies employing backward masking, OSM, and signal-to-respond methods with the same participants categorizing the same set of objects may offer the best opportunity for comprehensively characterizing the time course of object categorization.

In conclusion, the current findings are, to our knowledge, the first empirical evidence suggestive of a competitive dynamic between levels of abstraction during visual object categorization. Although future studies are necessary to validate and expand on these findings, the current results present a notable challenge for current theories of visual categorization.

Acknowledgements The authors thank Meg Schlichting for helpful discussions and Melisa Gumus, Marlie Tandoc, and Ian Mahoney for help with data collection and discussions of findings.

Funding This research is supported by a Natural Sciences and Engineering Research Council (NSERC) Discovery Grant to MLM (RGPIN-2017-06753), Canada Foundation for Innovation and Ontario Research Fund (36601) to MLM, and the Brain Canada Future Leaders in Canadian Brain Research Grant to MLM.

Declarations

Ethics approval Procedures were approved by the Vanderbilt University Institutional Review Board and the University of Toronto Research Ethics Board for Experiment 1 and Experiment 2, respectively.

Consent to Participate Informed written consent was obtained before the experiments in accordance with the guidelines of the Vanderbilt University Institutional Review Board and the University of Toronto Research Ethics Board for Experiment 1 and Experiment 2, respectively.

Consent for Publication Informed consent for publications was obtained in accordance with the guidelines of the Vanderbilt University Institutional Review Board and the University of Toronto Research Ethics Board for Experiment 1 and Experiment 2, respectively. In accordance with this consent, all data included in this work includes no identifying information.

Conflict of Interest The authors declare no conflict of interest.

Open Practices Statement The data and materials for both experiments are available on OSF (https://osf.io/rbv23/) and neither experiment was preregistered.

References

Annis, J., Gauthier, I., & Palmeri, T. J. (2020). Combining convolutional neural networks and cognitive models to predict novel object recognition in humans *Journal of Experimental Psychology, Learning: Memory, and Cognition*. https://doi.org/10.1037/xlm0000968

Bacon-Macé, N., Kirchner, H., Fabre-Thorpe, M., & Thorpe, S. J. (2007). Effects of task requirements on rapid natural scene processing: From common sensory encoding to distinct decisional mechanisms. *Journal of Experimental Psychology: Human Perception and Performance*, *33*(5), 1013–1026. https://doi.org/10.1037/0096-1523.33.5.1013

Bacon-Macé, N., Macé, M. J.-M., Fabre-Thorpe, M., & Thorpe, S. J. (2005). The time course of visual processing: Backward masking and natural scene categorisation. *Vision Research*, *45*(11), 1459–1469. https://doi.org/10.1016/j.visres.2005.01.004

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, *10*(4), 433–436. https://doi.org/10.1163/156856897X00357

Breitmeyer, B. G., & Ogmen, H. (2000). Recent models and findings in visual backward masking: a comparison, review, and update. *Perception & Psychophysics*, *62*(8), 1572–1595. https://doi.org/10.3758/BF03212157

Breitmeyer, B. G., & Ogmen, H. (2006). Visual masking: Time slices through conscious and unconscious vision. https://doi.org/10.1093/acprof:oso/9780198530671.001.0001

Cohen, A. L., & Nosofsky, R. M. (2003). An extension of the exemplar-based random-walk model to separable-dimension stimuli. *Journal of Mathematical Psychology*, *47*(2), 150–165. https://doi.org/10.1016/S0022-2496(02)00031-7

Di Lollo, V., Enns, J. T., & Rensink, R. A. (2000). Competition for consciousness among visual events: The psychophysics of reentrant visual processes. *Journal of Experimental Psychology: General*, *129*(4), 481–507. https://doi.org/10.1037/0096-3445.129.4.481

Enns, J. T., & Di Lollo, V. (1997). Object substitution: a new form of masking in unattended visual locations. *Psychological Science*, *8*(2), 135–139. https://doi.org/10.1111/j.1467-9280.1997.tb00696.x

Fabre-Thorpe, M. (2011). The characteristics and limits of rapid visual categorization. *Frontiers in Psychology* 2. https://doi.org/10.3389/fpsyg.2011.00243

Filmer, H. L., Mattingley, J. B., & Dux, P. E. (2015). Object substitution masking for an attended and foveated target. *Journal of Experimental Psychology. Human Perception and Performance*, *41*(1), 6–10. https://doi.org/10.1037/xhp0000024

Goodhew, S. C. (2017). What have we learned from two decades of object-substitution masking? Time to update: Object individuation prevails over substitution. *Journal of Experimental Psychology: Human Perception and Performance*, *43*(6), 1249–1262. https://doi.org/10.1037/xhp0000395

Goodhew, S. C., Pratt, J., Dux, P. E., & Ferber, S. (2013). Substituting objects from consciousness: a review of object substitution masking. *Psychonomic Bulletin & Review*, *20*(5), 859–877. https://doi.org/10.3758/s13423-013-0400-9

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

Harrison, G. W., Rajsic, J., & Wilson, D. E. (2016). Object-substitution masking degrades the quality of conscious object representations. *Psychonomic Bulletin & Review*, *23*(1), 180–186. https://doi.org/10.3758/s13423-015-0875-7

Johnson, K. E., & Mervis, C. B. (1997). Effects of varying levels of expertise on the basic level of categorization. *Journal of Experimental Psychology: General*, *126*(3), 248–277. https://doi.org/10.1037/0096-3445.126.3.248

Kar, K., Kubilius, J., Schmidt, K., Issa, E. B., & DiCarlo, J. J. (2019). Evidence that recurrent circuits are critical to the ventral stream's execution of core object recognition behavior. *Nature Neuroscience*, *22*(6), 974–983. https://doi.org/10.1038/s41593-019-0392-5

Kietzmann, T. C., Spoerer, C. J., Sörensen, L. K. A., Cichy, R. M., Hauk, O., & Kriegeskorte, N. (2019). Recurrence is required to capture the representational dynamics of the human visual system. *Proceedings of the National Academy of Sciences*, *116*(43), 21854–21863. https://doi.org/10.1073/pnas.1905544116

Kleiner, M., Brainard, D., & Pelli, D. (2007). What's new in Psychtoolbox-3? In Perception 36 ECVP abstract supplement.

Koivisto, M., Kastrati, G., & Revonsuo, A. (2014). Recurrent processing enhances visual awareness but is not necessary for fast categorization of natural scenes. *Journal of Cognitive Neuroscience*, *26*(2), 223–231.

Lamberts, K. (2000). Information-accumulation theory of speeded categorization. *Psychological Review*, *107*(2), 227–260. https://doi.org/10.1037/0033-295X.107.2.227

Lleras, A., & Moore, C. M. (2003). When the target becomes the mask: Using apparent motion to isolate the object-level component of object substitution masking. *Journal of Experimental Psychology: Human Perception and Performance*, *29*(1), 106–120. https://doi.org/10.1037/0096-1523.29.1.106

Love, B. C., Medin, D. L., & Gureckis, T. M. (2004). SUSTAIN: A network model of category learning. *Psychological Review*, *111*(2), 309–332. https://doi.org/10.1037/0033-295X.111.2.309

Macé, M. J.-M., Joubert, O. R., Nespoulous, J.-L., & Fabre-Thorpe, M. (2009). The time-course of visual categorizations: You spot the animal faster than the bird. *PLoS ONE*, *4*(6), e5927. https://doi.org/10.1371/journal.pone.0005927

Mack, M. L., & Palmeri, T. (2010). The speed of categorization: a priority for people? *Journal of Vision*, *10*(7), 988–988. https://doi.org/10.1167/10.7.988

Mack, M. L., & Palmeri, T. J. (2011). The timing of visual object categorization. *Frontiers in Psychology* 2. https://doi.org/10.3389/fpsyg.2011.00165

Mack, M. L., & Palmeri, T. J. (2015). The dynamics of categorization: Unraveling rapid categorization. *Journal of Experimental Psychology: General*, *144*(3), 551–569. https://doi.org/10.1037/a0039184

Mack, M. L., Wong, A. C.-N., Gauthier, I., Tanaka, J. W., & Palmeri, T. J. (2009). Time course of visual object categorization: Fastest does not necessarily mean first. *Vision Research*, *49*(15), 1961–1968. https://doi.org/10.1016/j.visres.2009.05.005

Makowski, D., Ben-Shachar, M. S., Chen, S. H. A., & Lüdecke, D. (2019). Indices of effect existence and significance in the Bayesian framework. *Frontiers in Psychology* 10. https://doi.org/10.3389/fpsyg.2019.02767

Mandler, J. M., Bauer, P. J., & McDonough, L. (1991). Separating the sheep from the goats: Differentiating global categories. *Cognitive Psychology*, *23*(2), 263–298. https://doi.org/10.1016/0010-0285(91)90011-C

Moore, C. M., & Lleras, A. (2005). On the role of object representations in substitution masking. *Journal of Experimental Psychology: Human Perception and Performance*, *31*(6), 1171–1180. https://doi.org/10.1037/0096-1523.31.6.1171

Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General*, *115*(1), 39–57. https://doi.org/10.1037/0096-3445.115.1.39

Nosofsky, R. M. (1988). Exemplar-based accounts of relations between classification, recognition, and typicality. *Journal of Experimental Psychology: Learning Memory, and Cognition*, *14*(4), 700–708. https://doi.org/10.1037/0278-7393.14.4.700

Nosofsky, R. M., & Palmeri, T. J. (1997). An exemplar-based random walk model of speeded classification. *Psychological Review*, *104*(2), 266–300. https://doi.org/10.1037/0033-295X.104.2.266

Palmeri, T. J. (1999). Learning categories at different hierarchical levels: A comparison of category learning models. *Psychonomic Bulletin and Review*, *6*, 495–503. Retrieved from https://doi.org/10.3758/BF03210840

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*(4), 437–442. https://doi.org/10.1163/156856897X00366

Ratcliff R, & Rouder JN (2000). A diffusion model account of masking in two-choice letter identification. *Journal of Experimental Psychology: Human Perception and Performance*, *26*(1), 127–140. Retrieved from https://doi.org/10.1037/0096-1523.26.1.127

Richler, J. J., Gauthier, I., & Palmeri, T. J. (2011). Automaticity of basic-level categorization accounts for labeling effects in visual recognition memory. *Journal of Experimental Psychology: Learning Memory and Cognition*, *37*(6), 1579–1587. https://doi.org/10.1037/a0024347

Rogers, T. T., & Patterson, K. (2007). Object categorization: Reversals and explanations of the basic-level advantage. *Journal of Experimental Psychology: General*, *136*(3), 451–469. https://doi.org/10.1037/0096-3445.136.3.451

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, *8*(3), 382–439. https://doi.org/10.1016/0010-0285(76)90013-X

Serre, T., Wolf, L., Bileschi, S., Riesenhuber, M., & Poggio, T. (2007). Robust object recognition with cortex-like mechanisms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *29*(3), 411–426. https://doi.org/10.1109/TPAMI.2007.56

Smith, P. L., Ratcliff, R., & Wolfgang, B. J. (2004). Attention orienting and the time course of perceptual decisions: Response time distributions with masked and unmasked displays. *Vision Research*, *44*(12), 1297–1320. Retrieved from https://doi.org/10.1016/j.visres.2004.01.002

Sofer, I., Crouzet, S. M., & Serre, T. (2015). Explaining the timing of natural scene understanding with a computational model of perceptual categorization. *PLOS Computational Biology*, *11*(9), e1004456. https://doi.org/10.1371/journal.pcbi.1004456

Tanaka, J. W., & Taylor, M. (1991). Object categories and expertise: is the basic level in the eye of the beholder? *Cognitive Psychology*, *23*(3), 457–482. https://doi.org/10.1016/0010-0285(91)90016-H

Thorpe, S. J., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature*, *381*(6582), 520–522. https://doi.org/10.1038/381520a0

Usher, M., & McClelland, J. L. (2001). The time course of perceptual choice: The leaky, competing accumulator model. *Psychological Review*, *108*(3), 550–592. https://doi.org/10.1037/0033-295X.108.3.550

Vanmarcke, S., Calders, F., & Wagemans, J. (2016). The time-course of ultrarapid categorization: The influence of scene congruency and top-down processing. *i-Perception*, *7*(5), 204166951667338. https://doi.org/10.1177/2041669516673384

VanRullen, R., & Thorpe, S. J. (2001a). Is it a bird? Is it a plane? Ultra-rapid visual categorisation of natural and artifactual objects. *Perception*, *30*(6), 655–668. https://doi.org/10.1068/p3029

VanRullen, R., & Thorpe, S. J. (2001b). The time course of visual processing: From early perception to decision-making. *Journal of Cognitive Neuroscience*, *13*(4), 454–461. https://doi.org/10.1162/08989290152001880