**Two objects or one? Similarity rather than complexity determines objecthood when
resolving dynamic input**

Stephanie C. Goodhew¹, Mark Edwards¹, Hannah L. Boal¹, Jason Bell²

¹Research School of Psychology, The Australian National University

²School of Psychology, The University of Western Australia

Word count: (main text): 6,413

Corresponding Author: Stephanie C. Goodhew

Address: Research School of Psychology (Building 39)

The Australian National University, Canberra, 0200

Email: stephanie.goodhew@anu.edu.au

Running head: Two objects or one?

Abstract

The human brain is continuously confronted with dynamic visual input, and from this it must infer whether input belongs to a single versus multiple object identities across time. Object substitution masking (OSM), in which perception of a target stimulus is impaired by a temporally-trailing four-dot mask, reflects a failure to segment the target and mask as discrete objects. According to Bouvier and Treisman (2010), OSM only occurs for targets that require binding multiple separate features (e.g., colour and orientation) in order to be identified. In contrast, a target that represents a unique feature is thought to be impervious to masking. Here, however, we show that a single orientation target (a Gabor) is susceptible to masking with an orientation-discrimination task, but only when the mask is similar in orientation to the target. That is, target-mask similarity, rather than target complexity determines masking. A re-examination of Bouvier and Treisman's results show that they can be explained within this target-mask similarity perspective. This means that the similarity of two objects determines whether they will be integrated or segmented across time, rather than the complexity of one of the objects in isolation.

Keywords: segmentation; object substitution masking; attention; feature binding; visual masking; temporal object segmentation; object updating.

The human brain faces a fundamental challenge in visual processing: inferring whether input arises from a single object continuing and changing across time, versus multiple distinct objects, occupying the same location at different points in time. For example, when walking down a crowded street, the same object (e.g., a particular person) can change their appearance markedly due to changes in viewpoint and location, and conversely different objects can occupy the same location at different points in time (e.g., two different people at the same location between successive glances). The brain therefore has the challenging task of determining, in the face of dynamic input, whether visual input from a given location belongs to a continuous object identity through time, or multiple separate objects. We call this process of carving up ongoing input across time into distinct object identities *temporal object segmentation*. Here we investigated the claim that object-complexity determines whether two objects are integrated into a single object representation or segmented as distinct objects (Bouvier & Treisman, 2010), and to anticipate, we show that this is not the case, and instead, the apparent effect of complexity can be attributed to similarity of the two, to-be-segmented objects.

Object substitution masking (OSM) represents an ideal paradigm with which to measure temporal object segmentation in the laboratory. In OSM, a target stimulus (e.g., a Landolt C) is presented briefly, with four small dots (the mask) on each corner of an imaginary square centred on the target (Di Lollo, Enns, & Rensink, 2000; Enns & Di Lollo, 1997; for a review, see Goodhew, Pratt, Dux, & Ferber, 2013). Target-like distractors are typically presented simultaneously with the target (but see Argyropoulos, Gellatly, Pilling, & Carter, 2013). Participants' task is to identify (e.g., the location of the gap in the C) or detect the target stimulus, and their accuracy in this task serves as measure of target perception. If the target, mask, and distractors all disappear simultaneously (simultaneous mask offset), then target perception is unimpaired. If, however, the four-dot mask is displayed after the

target and distractors have disappeared (delayed mask offset), then target perception is substantially impaired. That is, the temporally-trailing four-dots disrupt the perception of, or “mask” the target (Di Lollo, et al., 2000; Enns & Di Lollo, 1997).

At its core, OSM reflects a failure of temporal object segmentation. That is, the masking is a consequence of the brain treating two objects (1. the target surrounded by four dots and 2. the four-dots alone) as a single, continuing object. Since the mask is presented second and for longer than the target, its representation prevails in conscious awareness (Goodhew, Pratt, Dux, & Ferber, 2013), although the target can be recovered with prolonged mask exposure, as this allows time to consolidate the lingering target representation as a distinct object (Goodhew, Dux, Lipp, & Visser, 2012; Goodhew, Visser, Lipp, & Dux, 2011). Evidence for the intrinsic object-segmentation basis underlying OSM is that cues that encourage segmenting the target and mask as distinct objects, such as them appearing in distinct colours (Moore & Lleras, 2005), or luminance polarities (Luiga & Bachmann, 2008), reduces OSM. Further evidence that OSM taps fundamental temporal-object segmentation processes arises from its temporal dynamics. That is, when the mask is presented after a blank interval instead of temporally-continuous with the target display, then masking occurs when the interval between the target display and trailing mask is short, such that it is conducive to apparent motion, meaning that the trailing mask is perceived as the same object that was present during the target display. In contrast, target perception is unimpaired (masking eliminated) when a physically-identical mask is presented after a longer interval that is not conducive to apparent motion, such that the mask and target are perceived as distinct objects (Lleras & Moore, 2003). Altogether then, this suggests that OSM reflects the brain’s inference about single versus multiple objects across time, rather than image-level degradation.

Further evidence for object-level effects in OSM is that preview of placeholder stimuli (that do not indicate the identity of the target, e.g., rings that subsequently became Landolt Cs) at the locations of targets and distractors prior to the onset of the target array attenuates masking (Gellatly, Pilling, Carter, & Guest, 2010; Guest, Gellatly, & Pilling, 2012). Similarly, preview of mask objects (at multiple locations, so not predictive of the spatial location of the target) prior to the onset of the target reduces masking (Neill, Hutchison, & Graves, 2002), but only when the previewed masks and the masks visible during the target array are perceived as continuous objects (Lim & Chua, 2008). Finally, OSM is modulated by the temporal resolution of the visual system at the time the target and mask are presented. That is, when the relative contribution of magnocellular neurons (one of the major classes of visual cells which have superior temporal resolution (Chen et al., 2007; Derrington & Lennie, 1984; Livingstone & Hubel, 1988; Maunsell et al., 1999)), is enhanced, masking is reduced (Goodhew, Gozli, Ferber & Pratt, 2013), whereas when the contribution of these cells are actively diminished, masking is exacerbated (Goodhew, Boal, & Edwards, 2014). Altogether then, this supports the notion that OSM reflects a failure of temporal object segmentation.

While there has been some debate about whether this process in OSM implicates two distinct object representations that vie for conscious representation (object substitution) versus a single representation that initially represents the target which is updated to reflect the mask (object updating) (for a review of this debate, see Goodhew et al., 2013), at the level of conscious perception, one or both of these processes ultimately result in a failure to consciously perceive two objects when they appear in close spatiotemporal proximity. That is, they fundamentally probe the failure to segment objects through time for conscious representation. Critically, this means that these theoretical accounts unanimously implicate object-level interactions and modifications in this process of OSM.

Further evidence for the role of object-level interactions in OSM is that it appears that the to-be-identified target needs to consist of a combination of features in order to be susceptible to masking. To demonstrate this, Bouvier and Treisman (2010) used stimuli that consisted of two overlapping bars, one vertical and one horizontal (arranged in a plus-sign). One of the bars was coloured, and the other was white. Six such elements appeared in each array, and four white dots appeared around one of the items designating it as the target. The four dots either offset simultaneously or 300ms later. Participants' task was to identify both the colour and orientation of the non-white bar (dual task). The orientation identification task required binding of multiple features, since identifying the orientation of the coloured bar depends on combining colour and orientation features to a particular object in space, whereas the colour identification task did not, since there was only one non-white colour in the stimulus, identification of this colour could occur in parallel, without binding the identified colour to a particular object (Treisman & Gelade, 1980). The authors found that orientation identification was subject to OSM (poorer performance for delayed relative to simultaneous mask offset), whereas colour was not (equivalent performance for delayed and simultaneous mask offset conditions). A control condition eliminated the white bar from the stimuli, such that now identifying orientation no longer required binding. Here, there was no masking for either colour or orientation. This means that masking was only observed when the response to correctly identify the target required a conjunction of multiple independent features. From this, Bouvier and Treisman (2010) concluded that targets that are characterised by unique features are not susceptible to OSM, such that a necessary condition for maskability is that the target consists of multiple features that require binding.

Pilot data from a recent experiment we conducted (designed to answer a different theoretical question) appeared to produce evidence that is, at least *prima facie*, consistent with this. In this pilot experiment we found that a Gabor target (a sine-wave of a given

frequency in a Gaussian envelope, see Figure 2), which could be oriented to the left or right of vertical and participants' task was to identify its orientation, was resistant to masking by four-black dots, despite a Landolt C target being susceptible to masking under identical conditions. This suggests one of two intriguing possibilities: given that cells in the primary visual cortex are sufficient to resolve the orientation of a Gabor (De Valois, Albrecht, & Thorell, 1982; Hubel & Wiesel, 1968, 1974), or, in more cognitive terms, orientation is a unique feature in the target, it could mean that this simplicity renders such stimuli invulnerable to masking. Whereas objects whose perception necessarily requires higher object-levels brain areas, or binding of multiple features (such as required to identify the gap in a Landolt C) (Treisman & Souther, 1985), are susceptible to masking. This would be consistent with Bouvier and Treisman's (2010) assertion that the need for feature binding is a necessary condition for maskability. However, there is another contrasting possible explanation for the absence of masking for the Gabor. That is, given that the four-dot mask we used consisted of four black dots, it could be that differences in low-level visual properties between these dots of the mask and Gabor facilitated segmentation of the target and mask as discrete objects, thereby protecting against masking. What low-level cue would drive this segmentation? It has been established that colour (Moore & Lleras, 2005) and luminance polarity (Luiga & Bachmann, 2008) serve as segmentation cues in the face of dynamic input in close spatiotemporal proximity, such that the brain can recognise them as distinct objects and therefore masking is reduced when the target and mask are different colours or luminance polarities to one another. Neither of these cues was available here. A Gabor target is a sine-wave in a Gaussian envelope. This means that it has a specific spatial frequency, and therefore has a narrow-band spatial-frequency profile. The Gaussian envelope around the Gabor means that it has soft edges – that is, there is no abrupt change in luminance from the edge of the stimulus to the background, instead, it fades out gently.

(These two properties are also related: the softness of the edges helps to keep the narrow-band spatial-frequency profile, whereas abrupt edges entail a broader spatial frequency bandwidth). Gabors also have a specific orientation (e.g., vertical). In contrast, a black dot has abrupt or hard edges (immediate change in luminance from edge of object to the background), and is broadband both in terms of its spatial frequency profile (Graham, 1989) and its orientation (i.e., non-specific) (see Figure 1 for an illustration of orientation specificity).

Figure 1. Demonstration of orientation-specificity (i.e., narrowband) of a line (left), versus orientation non-specificity of a circle (i.e., broadband) (right). For the line, a single orientation is present. A Gabor is akin to the line, as a single orientation value is represented. In the circle, all possible orientations are present. Although a Landolt C is not a perfect circle (due to the gap), it is akin to a circle in that it is broadband because all orientation values are represented. While the gap may introduce some subtle degree of orientation information in Fourier space, it is far more broadband than a Gabor, meaning that the Cs and Gabors differ markedly in their orientation specificity.

This means that the combination of a Gabor target with black dots constituting the four-dot mask would have differed in their spatial frequency profile, orientation specificity, as well as the severity of their luminance edge profile. Dissimilarities in any one or more of these cues may have allowed the brain to segment the target and mask as distinct objects through time, thereby thwarting masking, since these properties are effective spatial segmentation cues for static object texture segmentation (Graham, Sutter, & Venkatesan, 1993; Klymenko & Weisstein, 1986; Landy & Bergen, 1991). If this is so, then under particular conditions (i.e., when the mask is more similar to it), a Gabor target would indeed be susceptible to masking. Such a result would challenge Bouvier and Treisman's conceptualisation that a target which can be identified by a single feature is not subject to OSM.

Figure 2. Top row: example of left-oriented Gabor targets. Bottom row: example of right-gap Landolt C targets. Left to right: four-dot mask consists of: broadband black dots, narrowband (4cpd) mini-Gabor dots, and (0cpd) Gaussian blob dots.

Experiment 1

To disentangle the different target complexity versus segmentation possibilities, we fully crossed target complexity (Gabor versus Landolt C) with mask type (Gabor, Black Dot, or Gaussian Dot mask). The target Gabor's spatial frequency was 4 cycles per degree (cpd) and the standard deviation of the Gaussian envelope was 0.2° of visual angle (i.e., narrow-band spatial frequency profile centred at 4cpd), it had a specific orientation, and had soft edges due to its Gaussian envelope. Target Gabors could be oriented to the left or to the right of vertical (by 22.5°). Landolt C targets consisted of a circle with a gap (making it a C-shape), with the hard edges that accompany such geometric shapes, and non-specific orientation¹. The four-dots constituting the mask could consist of Black Dots (hard edge, broadband spatial frequency), Gabors ($.07^\circ$ SD Gaussian envelope creating a soft edge, narrowband spatial frequency² at 4cpd), or Gaussian dots (soft edge, narrowband spatial frequency at 0cpd). It was not possible to have a Gaussian target, as it contains no discriminable information (see Figure 2). If the resistance of the Gabor to masking in our pilot data was driven by the target's minimal stimulus complexity as Bouvier and Treisman (2010) would predict, then we should see an absence of masking for the Gabor targets, irrespective of mask type (and masking should be observed for all Landolt C targets, irrespective of mask type). In contrast, if dissimilarity in low-level properties were enhancing segmentation and thereby preventing masking, then masking should depend on the similarity between the target and mask, rather than target type in isolation. Specifically, if dissimilarity of spatial frequency and lack of overlap in the spatial frequency spectra provides a segmentation cue, then the absence of masking should be specific to the Gabor-target with Black Dot mask condition (and the Landolt-C target with Gabor-mask condition). Alternatively, if similarity in edge-abruptness facilitates object segmentation in this context, then the absence of masking should be specific to the Gabor-target Black Dot mask condition, as well as the Landolt-C target with either the Gabor or Gaussian Dot mask. Finally, if orientation specificity drives segmentation, then the

absence of masking should be observed in the Gabor-target Black Dot mask condition and Gabor-target Gaussian-mask conditions, and the Landolt-C target Gabor-mask conditions.

Method

Participants. Twenty participants (15 female) were recruited via The Australian National University's SONA site (electronic sign-up system) and completed the experiment in exchange for course credit or pay. Participants' mean age was 21.15 years ($SD = 3.7$), and nineteen of the twenty participants were right-handed. All provided written informed consent prior to participation.

Stimuli and apparatus. Stimuli were presented on a cathode-ray tube (CRT) gamma-corrected monitor running at a 75Hz refresh rate. Viewing distance was fixed at 44cm with a chinrest. Stimuli were programmed in Matlab using the Psychophysics Toolbox (Brainard, 1997). The background was set to grey. On each trial, one target (signalled by the unique presence of the mask surrounding it) was presented simultaneously with five distractors that were the same category (Gabor, Landolt C) as the target (i.e., set-size six displays), but distractors were randomly determined to be either left or right oriented/gap location. Target and distractor stimuli were arranged in an imaginary circle with a radius of 9.3° . All targets (both Gabor and Landolt-Cs) had a diameter of 1.4° , and the diameter of all mask types was 0.5° , with 0.2° of separation between dots and targets.

Gabors were sine-waves in a Gaussian envelope centred on a mean spatial frequency of 4cpd, presented at 100% contrast. Target and distractor Gabors were oriented 22.5° to either the left or to the right of vertical (defined from the top of the stimulus). Mask Gabors were always vertically oriented. Landolt-Cs were black, with a gap either on the left or the right of the object. It was a true Landolt C, because stroke-width and the size of the gap were one-fifth of its size. Black Dot masks were black circles with hard edges. Gaussian Dot

masks were sine waves in a Gaussian envelope (identical to that for the Gabor masks) but centred on a mean spatial frequency of 0cpd.

Procedure. Each trial consisted of the following sequence: A fixation cross was displayed for 1000ms, then the target array (target, mask, and distractors) for 40ms, followed by either a blank screen (simultaneous mask offset) or the four-dot mask alone for 200ms (delayed mask offset). The fixation cross was presented until the participant responded. The inter-trial interval (during which the screen was blank grey) was 1000ms. This is a typical OSM procedure (see e.g., Goodhew et al., 2013).

Participants' task was to identify whether the target was oriented to the left or right of vertical (Gabor targets) or whether the gap in the target was on the left or right (Landolt C targets) by pressing the arrow keys on a standard keyboard. Responses were unspeeded. Target-type was blocked (order counterbalanced across participants). Within each block, one-third of trials (100 trials) were assigned to each mask type (Black Dot, Gabor Mask, Gaussian Dot), and of each of these, half (50 trials) were simultaneous offset trials and half were delayed offset trials, and the order of each of these combination of variables was randomised for each participant. On half of trials the target was left (oriented, gap) and half it was right.

Prior to each experimental block (Gabor, Landolt-C), participants completed a practice block of 12 trials, which started with slowed-down presentation times and progressively sped up. If necessary, participants repeated this practice block until they scored a minimum of 75% correct before progressing to the experimental block for that target type. The experiment consisted of 600 trials (300 per target type). Rest breaks were scheduled halfway through each target type, as well as at the changes between blocks, the length of which was at the discretion of the participant.

Results & Discussion

Data from one participant were excluded as this participant did not successfully meet the 75% performance criterion on the practice block after several attempts, and when run on the experiment itself performed at chance level (55% average). Target identification accuracy scores for the other nineteen participants were submitted to a 2 (target type) by 3 (mask type) by 2 (mask duration) repeated-measures ANOVA. All values reported are with the Greenhouse-Geisser correction for sphericity. This revealed that there was no significant main effect of target type ($F < 1$, $\eta_p^2 = .001$). Crucially, this means that the Gabors and Landolt-Cs were equally discriminable. There was, however, a significant main effect of mask duration, $F(1,18) = 48.55$, $p < .001$, $\eta_p^2 = .730$, such that target identification accuracy was on average greater with simultaneous mask offset ($M = 80.6\%$), than with delayed mask offset ($M = 73.7\%$). This demonstrates that our paradigm was successful in inducing OSM. There was also a significant main effect of mask type, $F(1.37, 24.7) = 8.63$, $p = .004$, $\eta_p^2 = .324$, such that target identification accuracy was greatest with Gaussian Dot mask ($M = 80.5\%$), followed by the Black Dot mask ($M = 76.3\%$), and lowest with the Gabor mask ($M = 74.6\%$). However, this main effect was qualified by subsequent interactions. The interaction between target type and mask type was significant, $F(1.96, 35.28) = 19.9$, $p < .001$, $\eta_p^2 = .525$, as was the interaction between target type and mask duration, $F(1,18) = 7.27$, $p = .015$, $\eta_p^2 = .288$. This interaction was driven by the fact that there was greater masking (difference in accuracy between simultaneous and delayed mask offset) for the Landolt C ($M = 10.1\%$) than for the Gabor targets ($M = 3.7\%$). Strikingly, however, there was also a significant three-way interaction among target type, mask type, and mask duration, $F(1.49, 26.90) = 5.70$, $p = .014$, $\eta_p^2 = .241$. This means that the effectiveness of the mask in obscuring target perception was not simply a function of mask or target type, but rather different mask types were differently effective for the different target types.

Figure 3. Masking magnitudes (differences in % correct between simultaneous and delayed mask offset conditions) for each of the target and mask-type combinations. Masking magnitudes that are statistically reliably different to zero are flagged with an asterisk. Error bars depict standard errors with the correction for within-subjects design described in (Cousineau, 2005) applied.

In order to follow-up this three-way interaction, we performed repeated measures *t*-tests comparing accuracy in the simultaneous versus delayed mask offset conditions for each combination of target and mask type. This test is essentially showing whether the masking magnitude for that combination of target and mask type (plotted in Figure 3) was reliably greater than zero. This revealed that for the Gabor targets, only the Gabor mask was effective ($p = .004$), whereas neither the Black Dot ($p = .421$) [replicating our pilot data] nor the Gaussian Dot mask ($p = .193$) were effective. The opposite was true for the Landolt-C targets: both the Black Dots ($p < .001$) and the Gaussian Dots ($p < .001$) were effective masks, whereas the Gabor mask was not ($p = .051$).

Mask:	Black-Dot	Black-Dot	Gabor	Gabor	Gaussian	Gaussian
MaskOffset	Simul	Delay	Simul	Delay	Simul	Delay
Gabor target:	81.2	79.2	75.5	69.0	80.6	77.9
Landolt-C target	79.3	65.7	80.1	73.8	86.8	76.5

Table 1. Percent correct values for simultaneous and delayed mask offset conditions for each mask type with each target type (Gabor vs Landolt C). Simul = simultaneous target and mask offset, Delay = delayed mask offset (by 200ms).

Here we were interested in disentangling the roles of stimulus complexity versus similarity in temporal object segmentation. The results demonstrate that OSM magnitude (our measure of temporal object segmentation) depends on target-mask similarity, rather than target complexity. That is, it was not the case that Gabor targets were invariantly impossible to mask and that Landolt Cs were always maskable. Instead, strikingly, it was clearly the

match between the target and mask type that was the critical determinant of masking, and therefore (the failure of) segmentation processes. This means that the absence of masking with a Gabor target and Black Dot mask that motivated the current study was a consequence of dissimilarity between the target and mask, providing segmentation cues and therefore protecting against masking. That is, target and mask similarity, rather than target complexity, was the clear determinant of masking, such that masking occurred when the target and mask were similar but not when they were dissimilar.

The results appear to suggest that dissimilarity in orientation-specificity was the primary driver of the presence or absence of masking. If this is so, then we could infer that orientation-specificity was serving as a cue to temporal object segmentation. In other words, the Gabor-target with Gabor-mask shared the same specificity of orientation information, leading the brain to determine that they belonged to the same object across time, resulting in a loss of target perception. When the mask instead had non-specific (i.e., broadband) orientation information (black dot, Gaussian dot), this facilitated segmentation of these as distinct from the Gabor target. When the Landolt-C was the target, with its non-specific orientation, the two masks with corresponding non-specific orientations were effective in impairing target perception, indicating that the brain deemed them as belonging to the same object as there was no segmentation cues available. Whereas the Gabor-mask that had distinct specific orientation information was treated as distinct from the Landolt C. Thus, orientation-specificity appears to facilitate object segmentation through time.

Edge-abruptness and breadth of spatial frequency profile are two intertwined low-level properties. Here, however, it did not appear that either served as an effective temporal object segmentation cue. It did not appear that edge-abruptness was a strong segmentation cue in its own right, as both Gabors and Gaussians have soft edges, yet these masks produced different effects for the different targets. That is, the Gabor mask was effective for the Gabor

target, whereas the Gaussian Dot was not, and in contrast, the Gaussian Dot mask was effective for the Landolt C target, whereas the Gabor mask was not. For the same reason, breadth of spatial frequency profile was not uniquely serving as cue to object segmentation across time, as both Gabors and Gaussian are narrowband or spatial-frequency-specific stimuli, but as described above, they behaved differently in interaction with the soft-edge Gabor versus hard-edge Landolt-C targets, whereas they should have had similar effects if specificity in spatial frequency was a segmentation cue. Of course, here orientation had a tighter bandwidth than spatial frequency, so it is perhaps unsurprising that orientation appeared to be the most effective segmentation cue. So while spatial frequency and edge properties were not effective temporal object segmentation cues in this experiment, there may be other contexts in which they are. Most critically, similarity rather than complexity appeared to be the primary determinant of masking.

Experiment 2

The results of Experiment 1 were consistent with the notion that similarity in orientation specificity determined maskability. This conclusion, however, was drawn from comparisons among stimuli that differed in multiple ways, and our definition of “similarity” was qualitative rather than quantitative. The purpose of Experiment 2, therefore, was to provide converging evidence for the role of similarity rather than complexity in OSM. To do this, we focussed on a single dimension of similarity (orientation), and parametrically varied the quantitative similarity between the target and mask along this dimension. We chose orientation as the crucial dimension because a) the results of Experiment 1 suggested that similarity in orientation-specificity was the key factor in determining objecthood in the face of dynamic input, and b) orientation similarity does not depend on the size of the stimuli, whereas a property such as spatial frequency is affected by the rather small size of the mask (Graham, 1989). If similarity is indeed the primary determinant of masking, then orientation-

similarity should interact with mask duration, such that the greatest masking is produced when the target and mask are similar and least masking when they are dissimilar.

Method

Participants. Eighteen participants (10 female) were recruited via The Australian National University's SONA site (electronic sign-up system) and completed the experiment in exchange for course credit or pay. Participants' mean age was 21.22 years ($SD = 2.3$), and 17 of the 18 participants were right-handed. All provided written informed consent prior to participation.

Stimuli and apparatus. Stimuli and apparatus were identical to Experiment 1, with the following exceptions. Target and mask stimuli were now always Gabors. As in Experiment 1, target Gabors were always oriented either to the left or right of vertical, and we instead varied the angular offset difference ($\Delta\theta$) between the target and mask to be one of: 0° , 22.5° , 45° , 67.5° , or 90° . The mask could be offset by the given $\Delta\theta$ in either direction from target (equi-probable). For example, $\Delta\theta = 0$ means that the target and mask would have identical orientations, whereas $\Delta\theta = 90$ means that the target and mask would have orthogonal orientations.

Procedure. The procedure was identical to Experiment 1, except that now there were 720 experimental trials, allowing 72 trials for each of the 10 conditions (five $\Delta\theta$ s by two mask durations).

Results & Discussion

Data from two participants were excluded from the analysis because their accuracy in one or more simultaneous mask offset conditions was at or below chance. The remaining datasets were submitted to a $5 (\Delta\theta) \times 2$ (mask duration) repeated-measures ANOVA, with the

Greenhouse-Geisser correction for sphericity. This revealed a significant main effect of $\Delta\theta$ on target identification accuracy, $F(2.18, 32.65) = 13.02$, $p < .001$, $\eta_p^2 = .465$, as well as a significant main effect of mask duration, $F(1,15) = 45.33$, $p < .001$, $\eta_p^2 = .751$. Moreover, there was also a significant interaction between $\Delta\theta$ and mask duration, $F(2.67, 40.06) = 17.73$, $p < .001$, $\eta_p^2 = .542$.

Figure 4. Masking magnitudes (differences in % correct between simultaneous and delayed mask offset conditions) for each angular offset (degrees) in orientation ($\Delta\theta$) between the target and mask Gabors. Masking magnitudes that are statistically reliably different to zero are flagged with an asterisk. Error bars depict standard errors with the correction described in (Cousineau, 2005) applied.

As can be seen from Figure 4, masking magnitude scaled as a function of $\Delta\theta$, our quantitative measure of target-mask similarity, such that the greatest masking is apparent when the target and mask have zero or small $\Delta\theta$, and masking tapered off to non-existent as $\Delta\theta$ increases. To confirm this, repeated-measures *t*-tests revealed that the large masking effect at $\Delta\theta = 0^\circ$ (18.0%) was significant, $t(15) = 5.37$, $p < .001$, as was the masking effect (16.1%) at $\Delta\theta = 22.5^\circ$, $t(15) = 9.84$, $p < .001$. The smaller masking effect (4.5%) at $\Delta\theta = 45^\circ$ was also significant $t(15) = 2.38$, $p = .031$. In contrast, neither the masking effect at $\Delta\theta = 67.5^\circ$ (-2.3%) nor at $\Delta\theta = 90^\circ$ (1.6%) were significant ($ts \leq 1.62$, $ps \geq .127$).

$\Delta\theta$	0	0	22.5	22.5	45	45	67.5	67.5	90	90
Mask	Simul	Delay	Simul	Delay	Simul	Delay	Simul	Delay	Simul	Delay
Offset										
	74.6	56.6	77.7	61.6	77.5	73.0	74.0	76.3	77.0	75.3

Table 1. Percent correct values for simultaneous and delayed mask offset conditions for angular offset in orientation between the target and mask ($\Delta\theta$). Simul = simultaneous target and mask offset, Delay = delayed mask offset (by 200ms).

To summarise, here we held constant target complexity and systematically varied the orientation-similarity between the target and mask, and this had a potent impact on masking

magnitude. This is converging evidence with our results from Experiment 1 that similarity, rather than complexity, dictates masking magnitude in OSM.

General Discussion

The purpose of the present experiments was to test the competing explanations of target complexity versus target-mask similarity as determinants of temporal object segmentation. We used OSM as our measure of temporal object segmentation, such that masking magnitude is inversely related to segmentation. In Experiment 1, we found that simple Gabor targets were indeed maskable, but only when the mask was also a Gabor (no masking with either the Black Dot or Gaussian Dot masks). With the more complex Landolt C targets, masking also appeared to depend on similarity, in that the Gabor mask was ineffective with the Landolt C target, whereas strong and reliable masking was observed with the Black Dot and Gaussian Dot masks. Experiment 2 provided convergent evidence that similarity rather than complexity is crucial in OSM, as here we had a constant simple Gabor target, and systematically varied and quantified the similarity between the target and mask in terms of orientation ($\Delta\theta$). Here, target-mask similarity powerfully determined masking magnitude, such that strong masking was obtained when the target and mask were highly similar, and no masking was obtained when they were orthogonal in orientation. These results tell us that orientation can serve as a cue to temporal object segmentation, in addition to colour (Moore & Lleras, 2005) and luminance polarity (Luiga & Bachmann, 2008). In other words, when the brain is bombarded with dynamic input including many stimuli in close spatiotemporal proximity, it draws on the heuristic specifying that stimuli that share similar orientation profiles are likely to belong to the same object. Moreover, the present results demonstrably refute the claim that target complexity is a prerequisite for masking.

At first glance, the fact that a Gabor target is subject to OSM might appear to undermine the object-updating mechanisms theorised to underlie OSM (Goodhew et al., 2013; Lleras & Moore, 2003; Luiga & Bachmann, 2008; Moore & Lleras, 2005), and contradict the conclusion that a single-feature stimulus (in this case, orientation) is not susceptible to OSM (Bouvier & Treisman, 2010). However, on closer consideration, the exclusion of Gabors from the category of ‘object’ is not well-supported by either emerging neurophysiological evidence, or a closer examination of the cognitive findings, in particular Bouvier and Treisman’s (2010) work. Instead, we argue that for a variety of reasons a Gabor should indeed be considered an object, and this then means that our results can be reconciled with the expansive body of evidence implicating object-level interactions in OSM. Furthermore, we propose a revision of Bouvier and Treisman’s (2010) theory that encompasses target-mask relationships rather than complexity of the target in isolation, which can explain our results as well as those obtained by Bouvier and Treisman.

Cells in V1 are exquisitely tuned to orientation information, and therefore *can* resolve the orientation of a Gabor (De Valois, et al., 1982; Hubel & Wiesel, 1968, 1974). This does not necessarily mean, however, that V1 is the endpoint of its processing. With the proliferation of re-entrant connections and the role of feedforward-feedback interactions in core aspects of visual processing (e.g., Bar et al., 2006; Boehler, Schoenfeld, Heinze, & Hopf, 2008; Bullier, 2001; Cudeiro & Sillito, 2006; Fahrenfort, Scholte, & Lamme, 2007; Kveraga, Boshyan, & Bar, 2007; Lamme & Roelfsema, 2000; Mumford, 1991; Pascual-Leone & Walsh, 2001; Ro, Breitmeyer, Burton, Singhal, & Lane, 2003; Salin & Bullier, 1995; Wyatte, Curran, & O'Reilly, 2012), such an assertion is difficult to support. Indeed, recent evidence suggests that orientation information is represented in lateral occipital cortex (LOC), an object-selective cortical area (Silson et al., 2013). This means that even though V1 cells are specialised for orientation encoding, this is clearly not the only region in which

orientation information is encoded. With this in mind, it is problematic to assume that the Gabor target is “simple” or “early” from a neurophysiological perspective. Instead, it is likely to require integration of information from multiple sources in this brain. If this integration process were considered to be the hallmark of whether a stimulus is perceptually treated as an object (as opposed to an isolated feature), it seems reasonable to infer that a Gabor is treated as an “object” in terms of perceptual processing. From this it is therefore logical why a Gabor is susceptible to the object-updating processes that underlie OSM, under conditions of high target-mask similarity that create confusion between the target and mask identities. Moreover, it reconciles our results with the body of work described earlier demonstrating object-level effects in OSM.

According to Bouvier and Treisman (2010) a stimulus is only sufficiently complex to warrant object-level status, and therefore the operations of OSM, if it requires the binding of multiple independent features, such as colour and orientation. In contrast, when the target can be correctly identified via a unique feature (e.g., a unique colour), it is resistant to OSM. In order to correctly identify the orientation of a Gabor stimulus (as was observers’ task in both of our experiments), only a single feature (orientation) needed to be processed. In other words, the target that reflects a value (e.g., left-of-vertical) along a single feature dimension (orientation). Therefore, Bouvier and Treisman’s perspective predicts that Gabors should not be subject to masking with an orientation-discrimination task. Yet here clearly they are. This means that our results are problematic for Bouvier and Treisman’s (2010) conclusions.

It is possible that Bouvier and Treisman’s (2010) framework for determining object-status, and therefore susceptibility to OSM, can be redeemed if we add a critical modification: considering the relationship between the target and mask. That is, these authors focussed exclusively on the complexity of information required to identify the target in isolation. Here, however, we have seen how the relationship between the target and mask has

a powerful influence on masking. Analogously, in real-world vision, objects do not appear in a vacuum, and so one of the challenges in object perception is differentiating an object from nearby objects, rather than just resolving that object in isolation. Thus, we propose that the difficulty of differentiating the target from the mask may determine susceptibility to masking, rather than the complexity of processing required to resolve the target in isolation. Reconsidering Bouvier and Treisman's (2010) design, it is possible that the pattern of results that they observed can also be explained with reference to this relationship. The absence of masking these authors observed when participants' response was to identify the coloured bar in the target could have been driven by the fact that the four-dot mask around the target was always white, providing a clear colour segmentation cue between the coloured part of the stimulus and the mask when the two bars were overlaid. In the control condition, in which just the coloured bar was presented with the white bar omitted, participants' orientation discrimination responses were now impervious to masking. Importantly, with this change there was now orientation-specificity in the target (i.e., a single orientation present), and no orientation specificity in the mask. Given the present results, it seems likely that this difference in orientation specificity may have served as a segmentation cue in Bouvier and Treisman's (2010) experiment, protecting these targets from masking. Finally, the presence of masking when identifying the orientation of the combined colour-orientation stimulus could reflect that in this case both the target and mask were orientation non-specific: there were multiple orientations present in the two-oriented-bar stimulus, and a dot is orientation-non-specific. Extrapolating from the orientation-specificity segmentation observed in our results, it seems likely that this similarity could have resulted in masking.

In conclusion, the presence or absence of OSM is dependent on the similarity between the target and mask objects, rather than target complexity per se. This means that even a simple, unique-feature target is susceptible to masking if the mask provides similar,

competing information along this feature dimension. This challenges Bouvier and Treisman's (2010) conclusion that binding of multiple features in order to identify the target in isolation is a necessary condition for OSM. However, a revised version of this framework, which takes into account the requirement to differentiate the target and mask as distinct objects, rather than just identify the target in isolation, can explain this result. Since OSM is an established metric of temporal object segmentation, this means that the brain's ability to resolve discrete objects representations in the face of dynamic input is affected by similarity, rather than complexity. That is, two objects are more likely to be integrated into a single, continuous object representation in time if they are more similar, and more likely to be segmented if they are more dissimilar. In contrast, the complexity of processing required to identify one of those objects does not appear to influence this segmentation process.

Notes

1. Note that non-specific orientation means that its orientation cannot be resolved by cells in V1. Although the C could be said to be upright relative to canonical C-shape representations, its orientation in the psychophysical sense is broadband or non-specific. See Figure 1 for an illustration of orientation-specificity.
2. Spatial-frequency specificity covaries with stimulus size, and thus the Gabors of the mask would have been slightly less specific in their spatial frequency than the Gabor targets, but most importantly much more specific than the Black Dots.

Acknowledgements

This research was supported by an Australian Research Council (ARC) Discovery Early Career Research Award (DE140101734) awarded to S.C.G., ARC Discovery Grant (DP110104553) awarded to M.E. and ARC Discovery Grant (DP110101511) awarded to J.B. The authors thank Rebecca Le and Reuben Rideaux for assistance with the data collection. Correspondence regarding this study should be addressed to Stephanie Goodhew (stephanie.goodhew@anu.edu.au), Research School of Psychology, The Australian National University.

References

- Argyropoulos, I., Gellatly, A., Pilling, M., & Carter, W. (2013). Set size and mask duration do not interact in object-substitution masking. *Journal of Experimental Psychology: Human Perception and Performance*, 39(3), 646-661. doi: 10.1037/a0030240
- Bar, M., Kassam, K. S., Ghuman, A. S., Boshyan, J., Schmidt, A. M., Dale, A. M., . . . Halgren, E. (2006). Top-down facilitation of visual recognition. *Proceedings of the National Academy of Sciences*, 103(2), 449-454. doi: 10.1073/pnas.0507062103
- Boehler, C. N., Schoenfeld, M. A., Heinze, H. J., & Hopf, J. M. (2008). Rapid recurrent processing gates awareness in primary visual cortex. *Proceedings of the National Academy of Sciences*, 105(25), 8742-8747. doi: 10.1073/pnas.0801999105
- Bouvier, S., & Treisman, A. (2010). Visual feature binding requires reentry. *Psychological Science*, 21(2), 200-204. doi: 10.1177/0956797609357858
- Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, 10(4), 433-436. doi: 10.1163/156856897X00357
- Bullier, J. (2001). Feedback connections and conscious vision. *Trends in Cognitive Sciences*, 5(9), 369-370. doi: 10.1016/S1364-6613%2800%2901730-7
- Chen, C.-M., Lakatos, P., Shah, A. S., Mehta, A. D., Givre, S. J., Javitt, D. C., & Schroeder, C. E. (2007). Functional Anatomy and Interaction of Fast and Slow Visual Pathways in Macaque Monkeys. *Cerebral Cortex*, 17(7), 1561-1569. doi: 10.1093/cercor/bhl067
- Cousineau, D. (2005). Confidence intervals in within-subject designs: A simpler solution to Loftus and Masson's method. *Tutorial in Quantitative Methods for Psychology*, 1(1), 42-45.
- Cudeiro, J., & Sillito, A. M. (2006). Looking back: Corticothalamic feedback and early visual processing. *Trends in Neurosciences*, 29(6), 298-306. doi: 10.1016/j.tins.2006.05.002

- De Valois, R. L., Albrecht, D. G., & Thorell, L. G. (1982). Spatial frequency selectivity of cells in macaque visual cortex. *Vision Res*, 22(5), 545-559.
- Derrington, A. M., & Lennie, P. (1984). Spatial and temporal contrast sensitivities of neurones in the lateral geniculate nucleus of the macaque. *Journal of Physiology*, 357, 219-240.
- 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. doi: 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. doi: 10.1111/j.1467-9280.1997.tb00696.x
- Fahrenfort, J. J., Scholte, H. S., & Lamme, V. A. F. (2007). Masking disrupts reentrant processing in human visual cortex. *Journal of Cognitive Neuroscience*, 19(9), 1488-1497. doi: 10.1162/jocn.2007.19.9.1488
- Gellatly, A., Pilling, M., Carter, W., & Guest, D. (2010). How does target duration affect object substitution masking. *Journal of Experimental Psychology: Human Perception and Performance*, 36(5), 1267-1279. doi: 10.1037/a0018733
- Goodhew, S. C., Boal, H. L., & Edwards, M. (2014). A magnocellular contribution to conscious perception via temporal object segmentation. *Journal of Experimental Psychology: Human Perception and Performance*, 40(3), 948-959. doi: 10.1037/a0035769
- Goodhew, S. C., Dux, P. E., Lipp, O. V., & Visser, T., A. W. (2012). Understanding recovery from object substitution masking. *Cognition*, 122(3), 405-415. doi: 10.1016/j.cognition.2011.11.010

- Goodhew, S. C., Gozli, D. G., Ferber, S., & Pratt, J. (2013). Reduced temporal fusion in near-hand space. *Psychological Science*, 24(6), 891-900. doi: 10.1177/0956797612463402
- 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. doi: 10.3758/s13423-013-0400-9
- Goodhew, S. C., Visser, T., A. W., Lipp, O. V., & Dux, P. E. (2011). Competing for consciousness: Prolonged mask exposure reduces object substitution masking. *Journal of Experimental Psychology: Human Perception and Performance*, 37(2), 588-596. doi: 10.1037/a0018740
- Graham, N., Sutter, A., & Venkatesan, C. (1993). Spatial-frequency- and orientation-selectivity of simple and complex channels in region segregation. *Vision Research*, 33(14), 1893-1911. doi: 10.1016/0042-6989(93)90017-Q
- Graham, N. V. S. (1989). *Visual Pattern Analyzers*. New York: Oxford University Press.
- Guest, D., Gellatly, A., & Pilling, M. (2012). Reduced OSM for long duration targets: Individuation or items loaded into VSTM? *Journal of Experimental Psychology: Human Perception and Performance*, 38(6), 1541-1553. doi: 10.1037/a0027031
- Hubel, D. H., & Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *The Journal of physiology*, 195(1), 215-243.
- Hubel, D. H., & Wiesel, T. N. (1974). Sequence regularity and geometry of orientation columns in the monkey striate cortex. *J Comp Neurol*, 158(3), 267-293.
- Klymenko, V., & Weisstein, N. (1986). Spatial frequency differences can determine figure-ground organization. *Journal of Experimental Psychology: Human Perception and Performance*, 12(3), 324. doi: 10.1037/0096-1523.12.3.324

- Kveraga, K., Boshyan, J., & Bar, M. (2007). Magnocellular projections as the trigger of top-down facilitation in recognition. *Journal of Neuroscience*, 27(48), 13232-13240. doi: 10.1523/jneurosci.3481-07.2007
- Lamme, V. A. F., & Roelfsema, P. R. (2000). The distinct modes of vision offered by feedforward and recurrent processing. *Trends in Neurosciences*, 23(11), 571-579. doi: 10.1016/S0166-2236(00)01657-X
- Landy, M. S., & Bergen, J. R. (1991). Texture segregation and orientation gradient. *Vision Research*, 31(4), 679-691. doi: 10.1016/0042-6989(91)90009-T
- Lim, S. W. H., & Chua, F. K. (2008). Object substitution masking: When does mask preview work? *Journal of Experimental Psychology: Human Perception and Performance*, 34(5), 1108-1115. doi: 10.1037/0096-1523.34.5.1108
- Livingstone, M., & Hubel, D. (1988). Segregation of form, color, movement, and depth: Anatomy, physiology, and perception. *Science*, 240(4853), 740-749. doi: 10.1126/science.3283936
- 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. doi: 10.1037/0096-1523.29.1.106
- Luiga, I., & Bachmann, T. (2008). Luminance processing in object substitution masking. *Vision Research*, 48(7), 937-945. doi: 10.1016/j.visres.2008.01.001
- Maunsell, J. R., Ghose, G. M., J.A., A., McAdams, C. J., Boudreau, C. E., & Noerager, B. D. (1999). Visual response latencies of magnocellular and parvocellular LGN neurons in macaque monkeys. *Visual Neuroscience*, 16, 1-14.

- 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. doi: 10.1037/0096-1523.31.6.1171
- Mumford, D. (1991). On the computational architecture of the neocortex: I. The role of the thalamo-cortical loop. *Biological Cybernetics*, 65(2), 135-145. doi: 10.1007/BF00202389
- Neill, W. T., Hutchison, K. A., & Graves, D. F. (2002). Masking by object substitution: Dissociation of masking and cueing effects. *Journal of Experimental Psychology: Human Perception and Performance*, 28(3), 682-694. doi: 10.1037/0096-1523.28.3.682
- Pascual-Leone, A., & Walsh, V. (2001). Fast backprojections from the motion to the primary visual area necessary for visual awareness. *Science*, 292(5516), 510-512. doi: 10.1126/science.1057099
- Ro, T., Breitmeyer, B., Burton, P., Singhal, N. S., & Lane, D. (2003). Feedback contributions to visual awareness in human occipital cortex. *Current Biology*, 11, 1038-1041.
- Salin, P., & Bullier, J. (1995). Corticocortical connections in the visual system: Structure and function. *Physiological Review*, 75, 107-154.
- Silson, E. H., McKeefry, D. J., Rodgers, J., Gouws, A. D., Hymers, M., & Morland, A. B. (2013). Specialized and independent processing of orientation and shape in visual field maps LO1 and LO2. *Nature Neuroscience*, 16(3), 267-269. doi: 10.1038/nn.3327
- Treisman, A., & Souther, J. (1985). Search asymmetry: A diagnostic for preattentive processing of separable features. *Journal of Experimental Psychology: General*, 114(3), 285-310. doi: 10.1037/0096-3445.114.3.285
- Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, 12(1), 97-136. doi: 10.1016/0010-0285(80)90005-5

Wyatte, D., Curran, T., & O'Reilly, R. (2012). The limits of feedforward vision: Recurrent processing promotes robust object recognition when objects are degraded. *Journal of Cognitive Neuroscience*, 24(11), 2248-2261. doi: 10.1162/jocn_a_00282