Atten Percept Psychophys (2014) 76:1057–1068
DOI 10.3758/s13414-014-0643-1

The subjective experience of object recognition: comparing metacognition for object detection and object categorization

Julia D. I. Meuwese · Anouk M. van Loon · Victor A. F. Lamme · Johannes J. Fahrenfort

Published online: 20 February 2014

Abstract Perceptual decisions seem to be made automatically and almost instantly. Constructing a unitary subjective conscious experience takes more time. For example, when trying to avoid a collision with a car on a foggy road you brake or steer away in a reflex, before realizing you were in a near accident. This subjective aspect of object recognition has been given little attention. We used metacognition (assessed with confidence ratings) to measure subjective experience during object detection and object categorization for degraded and masked objects, while objective performance was matched. Metacognition was equal for degraded and masked objects, but categorization led to higher metacognition than did detection. This effect turned out to be driven by a difference in metacognition for correct rejection trials, which seemed to be caused by an asymmetry of the distractor stimulus: It does not contain object-related information in the detection task, whereas it does contain such information in the categorization task. Strikingly, this asymmetry selectively impacted metacognitive ability when objective performance was matched. This finding reveals a fundamental difference in how humans reflect versus act on information: When matching the amount of information required to perform two tasks at some objective level of accuracy (acting), metacognitive ability (reflecting) is still better in tasks that rely on positive evidence (categorization) than in tasks that rely more strongly on an absence of evidence (detection).

Keywords Metacognition · Object recognition · Categorization · Detection · Perception · Masking · Degrading · Consciousness

When driving in dense fog, and all of a sudden a car looms in front of you, you instantly brake or steer away in a reflex, before realizing that something is about to hit you or knowing what it is. Only moments later, you become aware that you have just successfully avoided being hit by a car. This illustrates that perceptual decisions can be automatic and, when under time pressure or when visibility is poor, often precede our subjective experience (Gregori-Grgic, Balderi, & De'Sperati, 2011; Jolij, Scholte, van Gaal, Hodgson, & Lamme, 2011). It takes time to fully process multiple sensory signals and experience them as a unitary representation of the world. And time is precious, especially when you are almost hit by a car on a foggy road.

The exact time course of object detection and object categorization is a subject of debate. Does one first detect that an object is there ("Something is right in front of me!"), and subsequently categorize what it is ("It's a car!")? Or are both categorization and detection the simultaneous outcome of the same process ("A car is right in front of me!")? Grill-Spector and Kanwisher (2005) proposed the latter, "as soon as you know it's there, you know what it is," on the basis of equal performance on both tasks (for equal exposure durations). However, others have since shown that detection might in fact be easier than categorization, at least when categories are more similar to each other (Bowers & Jones, 2008; Mack & Palmeri, 2010) and when stimuli are inverted or degraded instead of masked (Mack, Gauthier, Sadr, & Palmeri, 2008). They proposed that more visual information is required for

Electronic supplementary material The online version of this article (doi:10.3758/s13414-014-0643-1) contains supplementary material, which is available to authorized users.

J. D. I. Meuwese · A. M. van Loon · V. A. F. Lamme · J. J. Fahrenfort
Brain & Cognition, Department of Psychology and Cognitive Science Center Amsterdam (CSCA), University of Amsterdam, Amsterdam, The Netherlands

J. D. I. Meuwese (✉)
Department of Psychology, University of Amsterdam, Weesperplein 4, 1018 XA Amsterdam, The Netherlands
e-mail: julia.meuwese@gmail.com

categorization than for detection, which makes it more difficult when this information (or time) is limited.

However, these studies only investigated objective task performance. The subjective aspect of object recognition has not been part of this debate. Perceptual decisions may not tell the full story about perception, since they can be made at an onset for which stimuli are subjectively invisible or poorly visible. This is illustrated by the example above, but also by a condition described as "blindsight." Patients with a lesion to (part of) V1 can discriminate stimuli presented to their blind field above chance level, despite being subjectively blind (Weiskrantz, Warrington, Sanders, & Marshall, 1974). This condition can be mimicked in normal observers in various ways, such as by applying transcranial magnetic stimulation to V1 (Boyer, Harrison, & Ro, 2005), through binocular rivalry (Kolb & Braun, 1995), or by using a metacontrast mask (Lau & Passingham, 2006). This shows that subjective experience and objective performance can dissociate. So what about our subjective awareness of object detection and object categorization? When objective performance is equal, is the subjective access to these perceptual decisions equal as well? As soon as you know it's there, do you really *know* it is there, and do you really *know* what it is?

Here, we use metacognition to assess the subjective experience of object detection and categorization. Metacognition is becoming increasingly popular as a measure of subjective experience (see Fleming, Dolan, & Frith, 2012, for an overview), and is defined as the ability to have insight into the objective correctness of a response (Fleming, Weil, Nagy, Dolan, & Rees, 2010). This is measured by the correspondence between subjective confidence in the accuracy of an objective response and the actual, objective performance (Lau & Passingham, 2006). Several studies have shown that when objective performance is held constant, metacognition can vary between subjects (Fleming et al., 2010; Kanai, Walsh, & Tseng, 2010; Lau & Passingham, 2006). Perhaps differences in subjective experience between categorization and detection can reveal whether or not objective performance and subjective confidence are based on the same information.

We used two perceptual manipulations to decrease stimulus visibility (otherwise, objective performance and metacognition would be at ceiling): backward masking (Breitmeyer, 1984) and degrading (Genetti, Britz, Michel, & Pegna, 2010). Degrading is an effective, bottom-up way to manipulate stimulus visibility, because the input signal is degraded before it even enters the brain. Backward masking is thought to affect neural processing at a later stage, because the stimulus itself is presented without bottom-up interference. Only because of the mask (which is presented *after* stimulus presentation) is (recurrent) processing interrupted and objective discriminability perturbed (Fahrenfort, Scholte, & Lamme, 2007; Lamme & Roelfsema, 2000; Lamme, Zipser, & Spekreijse, 2002). Degrading has been shown to impair categorization more than detection, whereas masking affects performance equally on both tasks (Mack et al., 2008). But would this distinction also be reflected in differences in metacognitive ability (MA), the degree to which subjects have access to the correctness of their responses? In other words, does it matter for one's MA of detection and categorization whether objects are degraded or masked?

To answer these questions, we determined the amount of metacognition during stimulus detection and categorization for both degraded and masked objects, while keeping objective performance equal. The degree of degradation (manipulated by varying phase coherence) or masking strength (manipulated by varying stimulus duration) was adjusted for each subject, such that objective task performance was matched for both detection and categorization. This enabled us to compare subjective experience for the object recognition of degraded versus masked objects, without the confounding effects caused by differences in objective performance. By using multiple analyses of MA, we hoped to reveal a specific profile of metacognition for each manipulation, which will tell us more about the degree to which subjects have subjective access to the outcome of object detection and object categorization (see the Method section for a more detailed explanation of all metacognitive measures used).

Method

Subjects

A total of 51 subjects participated in this study (35 females, 16 males) for course credit or financial compensation. The subjects gave written informed consent before experimentation and had normal or corrected-to-normal vision. The experiment was approved by the Ethics Committee of the Psychology Department of the University of Amsterdam. Two subjects were excluded because we could not exclude the possibility that they were merely guessing on the categorization task (they scored less than 56 % correct on more than half of the blocks), and three subjects were excluded because they did not follow the instructions to use the whole range of confidence ratings (see the Confidence Ratings and Metacognition Scores section below). All analyses are based on the remaining 46 subjects (31 females, 15 males; 18 – 29 years of age, mean = 21.7 years, SD = 2.37).

General procedure

Subjects were randomly assigned to either the masked or the degraded condition. In both conditions, subjects performed a detection and a categorization task. First, they performed a short staircase procedure for both tasks separately, to determine the level of masking/degrading required to achieve a task

performance of 71 % correct. Then 12 interleaved blocks of the detection and categorization task were performed. For the detection task, subjects had to judge whether the masked/degraded stimulus contained an animal (cat, bird or fish) or no object at all (fully phase-scrambled versions of the images). For the categorization task an animal was always present, and they had to categorize it as belonging to the target category (randomly selected per block from cat, bird or fish), or to a distractor category. After every response, subjects had to rate their confidence about their response, on a scale from 1 (*not at all confident*) to 6 (*very confident*). These confidence ratings were linked with objective performance to calculate MA for detection and categorization. In between blocks, subjects could take a short break, after six blocks they took a longer break (5–10 min). To keep performance equal for both tasks across blocks, the percentage correct was monitored after each block, and the level of degrading/masking was adjusted to target 71 % accuracy.

Stimuli

The stimuli consisted of grayscale photos of animals from three categories: cats, birds, and fish. In total, 600 different images were used, 200 per category. The stimulus categories were matched for low-level image statistics; namely, the beta and gamma parameters of the Weibull function were fitted to the distribution of contrast values of the image, which has been shown to effectively balance the cortical responses in low-level visual areas to these images (Scholte, Ghebreab, Waldorp, Smeulders, & Lamme, 2009). Across tasks, every image was presented only once to each subject, and this was carefully randomized across subjects, such that each image appeared equally often in every task. For the categorization task, the distractor stimuli were taken from the same set as the target stimuli. For the detection task, the distractor stimuli were fully phase-scrambled versions of the target stimuli, such that the distractor contained no remaining object information (see Fig. 1b). These phase-scrambled images have a very similar profile, in terms of their low-level image statistics, to the target stimuli, since their second-order image statistics (the overall contrast and texture profile) were retained but their object information (luminance-defined edges) was removed. Such images are a commonly used control in the study of brain regions that are involved in the detection of objects (e.g., Malach et al., 1995; Op de Beeck, Baker, DiCarlo, & Kanwisher, 2006). The same stimuli were used in the masked and degraded conditions. In the degraded condition, the stimuli were degraded by varying degrees of phase scrambling, whereas in the masked condition, stimuli were masked by textured patterns with randomly oriented line elements. The thickness of these line elements varied randomly per mask (but within each mask line thickness was equal) (see Fig. 1a). In the degraded condition, "filler" stimuli were presented instead of the masks, which consisted of target stimuli that were scrambled on a pixel-by-pixel basis, resulting in homogeneous gray images. Stimuli were presented on a 60-Hz monitor (Dell, 35º × 22.5° of visual angle).

Detection and categorization task

The present detection and categorization paradigms were based on previous studies that had assessed objective performance (Grill-Spector & Kanwisher, 2005; Mack et al., 2008), with the addition of prompting a confidence rating on every trial in order to assess MA. Subjects performed both the detection and categorization tasks, in 12 interleaved blocks (six blocks of 60 trials per task, the order of which was counterbalanced across subjects). Before each block started, subjects were informed which question they had to answer during that block. At the start of the experiment, task performance was set at 71 % correct, to ensure equal performance within and across subjects, by a staircase procedure (see the Matching Objective Performance section for details).

During the detection task, the degraded or masked stimulus contained either an animal (cat, bird, or fish; target) or no animal (100 % phase-scrambled image; distractor). For every trial, subjects were asked "Was there an animal present?" They answered "yes" when the image contained an animal (50 % of the trials) and "no" when it contained a phase-scrambled texture with no object at all (50 % of the trials).

For the categorization task, the degraded or masked stimulus contained either a cat, bird, or fish. In each block, one category was randomly selected as the target category (such that each category was the target category for two out of the six blocks), and subjects had to categorize the animal as either a member of the target category or a member of a distractor category (divided 50–50 over the two remaining categories). For instance, during one block subjects could be asked "Was the animal a cat?," to which they should answer "yes" or "no." In all, 50 % of the trials had targets (images of animals from the target category) and 50 % had distractors (images of animals from the distractor categories [e.g., bird and fish]).

The subjects responded by first selecting their answer with the arrow keys, and then confirming with the space bar. They had to respond within 2 s; otherwise, they were informed that the trial was aborted because they had not responded in time. After each response, a confidence rating had to be provided; see below for details.

Masked and degraded conditions

Stimulus visibility was manipulated between subjects, by either pattern masking or degrading. Every trial lasted for 1,250 ms and started with a fixation cross against a black background (300 ms). Then, in the masked condition, the fixation cross was presented together with a textured pattern mask (200 ms). After that, the target or distractor stimulus was

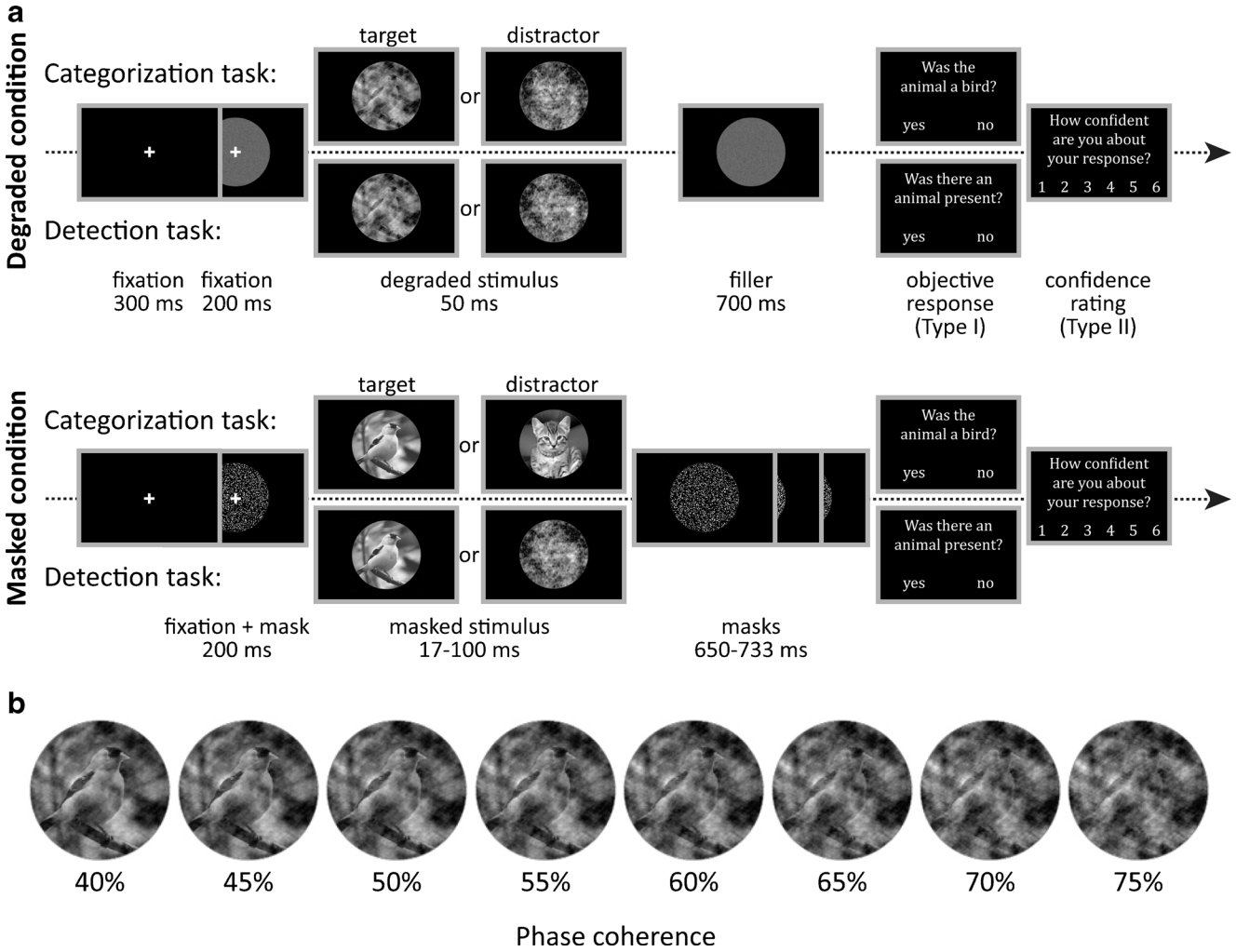

Fig. 1 Task design. **a** Subjects performed both the detection task and the categorization task in 12 interleaved blocks of 60 trials. Stimulus visibility was manipulated between subjects, by either pattern masking or degrading (phase scrambling). During the detection task, the degraded or masked stimulus contained either an animal (cat, bird, or fish; target) or a 100 % phase-scrambled image (distractor). On every trial, subjects were asked "Was there an animal present?" For the categorization task, a target category was randomly selected for each block (i.e., "bird"), and the stimuli consisted of a degraded/masked cat, bird, or fish. Subjects were asked whether the animal was a member of the target category (i.e., "Was the animal a bird?"). After the "yes"/"no" response was made on the detection and categorization tasks, subjects had to rate their confidence in the correctness of their response on a scale from 1 (*not at all confident*) to 6 (*very confident*). By linking confidence ratings with objective performance, metacognitive ability (MA) was calculated. **b** An image that is phase scrambled to different coherence levels: from left to right, 40 % to 75 % phase coherence, which was the range of phase coherence levels and step sizes used in degraded condition of the experiment

presented (for 17 – 100 ms, depending on the level required to score around 71 % correct, as determined by the staircase task; see below), followed by a series of textured pattern masks, consisting of random line elements of various widths. Six different masks were presented, the first five for 50 ms each; the final mask was presented until the end of the trial (i.e., for 400–483 ms, depending on the stimulus duration, since the total trial duration was fixed at 1,250 ms). In the degraded condition, noise was added to the stimuli by means of *phase scrambling* (changing their phase coherence levels). The stimuli were phase-scrambled to various degrees (depending on the level required to score around 71 % correct, as determined by the staircase procedure; see below). After the fixation cross (presented against a black background), the fixation cross was presented together with a "filler" neutral gray image (200 ms). Then the target or distractor stimulus was presented (50 ms), followed by a filler image (700 ms).

Matching objective performance

We aimed to keep performance equal for both tasks and across subjects, in order to ensure that objective performance could not confound MA. We chose a performance level of 71 % correct, because this ensured enough incorrect answers to calculate a reliable metacognition score, whilst leaving enough correct answers to keep subjects motivated.

To establish the level of degrading/masking at which each subject would achieve a performance level of 71 % correct, a

staircase procedure was performed at the start of the experiment. The staircase task consisted of 30 detection trials and 30 categorization trials (the order was counterbalanced). The staircase followed a "2-up, 1-down" rule: For every two correct answers, the level of degrading/masking was increased by one step size, and for every incorrect answer, the level was decreased by one step size. "One step size" equaled 5 % less/more phase coherence (degraded condition) or a 17-ms decrease/increase of the stimulus duration (masked condition). The starting point of the staircase task was set at a phase coherence level of 50 % (degraded condition) or a 67-ms stimulus duration (masked condition). The actual tasks started at the level of degrading/masking at which the staircase procedure ended (the final trial).

To keep performance equal for both tasks across blocks, the percentage correct was monitored after each block, and the level of degrading/masking was adjusted accordingly (separately for the categorization and detection tasks). Whenever performance deviated by more than 5 % from 71 % correct, the level was either decreased by one step size (when <66 % correct) or increased by one step size (when >76 % correct). This margin was chosen to prevent adjusting the level too often because of minor performance fluctuations. When performance was very high (> 86 % correct), the level was increased by two step sizes.

Confidence ratings and metacognition scores

On every trial, after subjects had made their response about stimulus presence or absence, they had to rate their confidence about the correctness of this response on a scale from 1 (*not at all confident*) to 6 (*very confident*). Subjects first selected their answer with the arrow keys, and then confirmed it by pressing the space bar. They had to respond within 3.5 s; otherwise, they were informed that the trial was aborted because they had not responded in time. Subjects were instructed to rate their confidence relative to the other stimuli in the task, using the whole range of the confidence scale (from 1 to 6; Fleming, Huijgen, & Dolan, 2012; Zylberberg, Barttfeld, & Sigman, 2012). This scale was chosen to accurately measure MA, since it is more sensitive than a discrete "high" or "low" response. Two subjects who did not follow these instructions had unusable MA scores, and were therefore excluded from further analysis (see the Subjects section).

Metacognition (or "Type II") performance was calculated by linking objective "Type I" performance with confidence ratings. Metacognitive Type II performance is high whenever a subject is confident about correct Type I responses (hits and correct rejections [CRs]) and not confident about incorrect Type I responses (misses and false alarms [FAs]). In other words, metacognition is high when a subject knows when he or she is objectively wrong or right. In the typical measure of metacognition (the "classic" measure), all Type I responses (hits, misses, CRs, and FAs) are included. However, this measure aggregates all Type I responses into correct (hits and CRs) and incorrect (misses and FAs) trials, thereby overlooking differences in metacognition for the reported absence (CRs and misses) versus presence (hits and FAs) of a stimulus. Kanai et al. (2010) introduced the SDI ("subjective discriminability of invisibility") measure, which only includes trials in which subjects reported stimulus absence (CRs and misses). They found that the SDI measure revealed selective differences between tasks that used attentional manipulations and tasks that used perceptual manipulations (affecting the visibility of the stimulus itself). The Kanai et al. study shows that metacognition is not a unitary measure that is equally influenced by different stimulus–response combinations across tasks. Therefore, we also used the SDI, and introduced SDI's counterpart, the SDV ("subjective discriminability of visibility") measure, which includes only trials in which subjects reported stimulus presence (hits and FAs), potentially revealing differences across tasks in metacognition between physical and perceived stimulus presence.

We constructed a receiver-operating characteristic (ROC) curve for each measure, by plotting the cumulative probability of confidence in correct trials (classic, hits and CRs; SDI, only CRs; SDV, only hits) against the cumulative probability of confidence in incorrect trials (classic, misses and FAs; SDI, only misses; SDV, only FAs; Kanai et al., 2010). Inflection points were plotted from high to low confidence. This meant that the first (leftmost) inflection point expressed the proportion (expressed as a fraction from 0 to 1; i.e., probability) of correct trials for which the highest confidence rating ("6") was given (*y*-axis) versus the percentage of incorrect trials for which a "6" rating was given (*x*-axis); the second inflection point represented the same for a confidence rating of "5" (yet cumulatively, so adding up to the probability of rating "6"); and so forth. Thus, an ROC curve above the diagonal meant that subjects had higher confidence for correct than for incorrect trials, which meant that metacognition was above what would be expected on the basis of chance alone. In contrast, when the ROC curve was equal to the diagonal, it meant that subjects' confidence ratings did not distinguish correct from incorrect responses. ROC curves were plotted for each condition and task separately, and for each different measure (classic, SDI, and SDV). Note that since ROC curves were calculated per Type I trial type (or a combination of those, for the classic measure), this measure was independent of the number of trials per trial type.

Finally, in order to further investigate the cause of any metacognitive differences between categorization and detection, we compared confidence ratings of the two tasks for each Type I response separately, by plotting an ROC curve of the cumulative probabilities of confidence for categorization (*y*-axis) against detection (*x*-axis; for hits, misses, CRs, and FAs separately). An ROC curve above the diagonal then meant that

confidence ratings for that particular Type I response were higher for categorization than for detection, and below the diagonal, responses were higher for detection.

To quantify a single metacognition score for each ROC, we calculated the "area under the curve" (AUC) value (Kanai et al., 2010; Szczepanowski & Pessoa, 2007; Wilimzig, Tsuchiya, Fahle, Einhäuser, & Koch, 2008). This value normally ranges from .5 (*no metacognition*) to 1 (*perfect metacognition*). For each measure, we compared the AUC values between groups (masked and degraded, for detection and categorization separately), using independent two-tailed *t* tests, and within subjects (detection vs. categorization, for each group separately) using paired two-tailed *t* tests. We also tested the AUC value of each categorization-versus-detection plot (for each Type I response separately) against .5, to test whether or not confidence ratings were equally distributed across tasks for each Type I response.

Results

Objective performance

Although performance was successfully matched between tasks [degraded group: detection, 72.8 % correct (SD = 4.7), categorization, 71.3 % correct (SD = 6.5), $t(1, 21) = 0.910$, p = .37; masked group: detection, 73.7 % correct (SD = 4.5), categorization, 72.7 % correct (SD = 3.1), $t(1, 23) = 0.862$, p = .40] and between conditions [detection, $t(1, 44) = –0.636$, p = .53; categorization, $t(1, 44) = – 0.961$, p = .34], we excluded blocks in which performance was lower than 56 % correct from further analysis, to exclude data that were the result of "guessing" or extreme fatigue (this happened mostly before the break and toward the end of the experiment, when subjects probably were tired or less motivated). In total, 6 % of the blocks were removed because of low performance (32 blocks, in 19 subjects).

After removal of these low-performance blocks, we tested whether the percentages correct were equal across tasks, to ensure that objective performance was still matched. Performance was successfully matched between tasks [degraded group: detection, 73.2 % correct (SD = 4.1); categorization, 73.2 % correct (SD = 4.8), $t(1, 21) = 0.043$, p = .97; masked group: detection, 74.1 % correct (SD = 4.3); categorization, 73.8 % correct (SD = 2.3), $t(1, 23) = 0.287$, p = .78] and between conditions [detection, $t(1, 44) = 0.723$, p = .47; categorization, $t(1, 44) = 0.612$, p = .54]. Also d', a criterion-free measure of objective performance (calculated from the z score of the hit rate minus the z score of the FA rate) was equal across tasks [within-subjects level: for the degraded group—detection, d' = 1.51 (SD = 0.43); categorization, d' = 1.38 (SD = 0.35), $t(1, 21) = 1.408$, p = .17; for the masked group—detection, d' = 1.45 (SD = 0.33); categorization, d' = 1.38 (SD = 0.19), $t(1, 23) = 0.915$, p = .37] and between conditions [detection, $t(1, 44) = –0.569$, p = .57; categorization, $t(1, 44) = 0.073$, p = .94].

For both tasks (categorization and detection), we also compared stimulus visibility, to exclude this as a possible confound in performance and/or metacognition scores. Note that we could not compare visibility between conditions, because the masking and degrading methods were very different and noncomparable on a quantitative, between-groups level. The levels of degrading/masking were equal for both tasks [degraded group: detection, 64.7 % phase coherence (SD = 8.1); categorization, 64.7 % phase coherence (SD = 4.3), $t(1, 21) = 0.014$, p = .99; masked group: detection, 29.3-ms stimulus duration (SD = 13.6); categorization, 32.3-ms stimulus duration (SD = 9.7), $t(1, 23) = –0.910$, p = .37], ruling out the confounding effects of visibility.

Metacognition

For both masking and degrading, metacognition was significantly higher for the categorization than for the detection task, according to the classic measure of MA (see Fig. 2). Subjects had better insight into their own performance when judging whether something was part of a certain category (categorization) than when judging whether an object was present (detection) [masked group: AUC difference = .05, $t(1, 23) = –4.603$, p = .0001; degraded group: AUC difference = .06, $t(1, 21) = –3.777$, p = .001]. To investigate the basis of this classic metacognition effect, we looked at SDI and SDV. The results revealed that the difference in classic metacognition was driven solely by the SDI measure (which included only CRs and misses) [masked group: AUC difference = .09, $t(1, 23) = – 4.496$, p = .0002; degraded group: AUC difference = .09, $t(1, 21) = –4.014$, p = .0006], rather than the SDV measure (including only hits and FAs) [masked group: AUC difference = .01, $t(1, 23) = – 0.523$, p = .61; degraded group: AUC difference = .01, $t(1, 21) = –0.486$, p = .63]. Thus, subjects had better insight into their performance for categorization than for detection, but only when reporting the absence of a target (CRs and misses), as measured by the SDI. This difference was still present in the classic measure (including all responses), but was driven by the SDI measure, since no difference was present for the SDV measure (hits and FAs).

In order to further investigate the difference between categorization and detection and the stronger result for the SDI than for the classic measure, we calculated the proportions of Type I responses per confidence rating separately for hits, misses, CRs, and FAs, and plotted detection against categorization (see Fig. 3). Whenever the resulting AUC differed from .5, this meant that confidence ratings were not equally distributed across the two tasks. This analysis allowed us to see whether the differences between detection and categorization

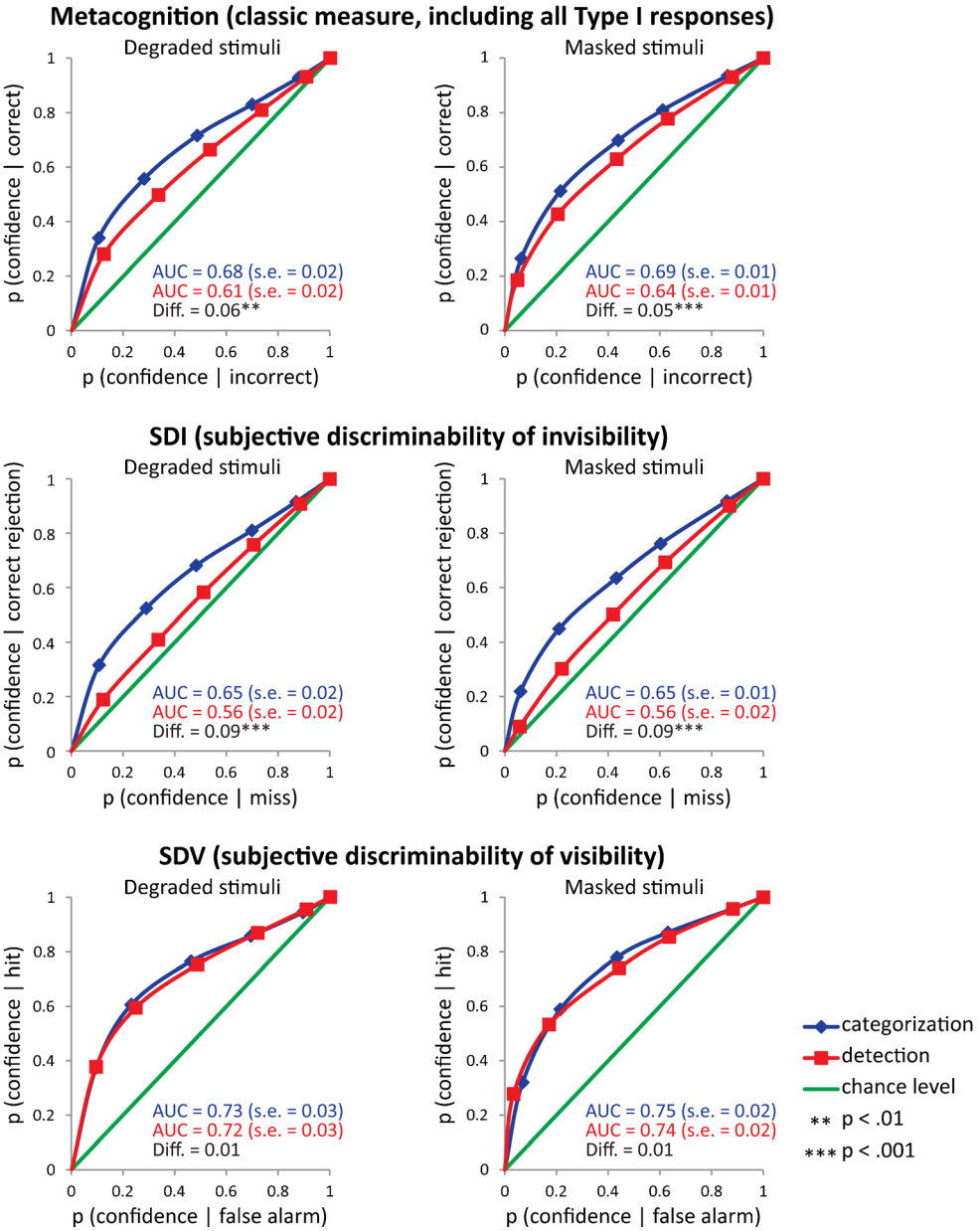

Fig. 2 Metacognition scores for detection and categorization tasks. Metacognitive measures reflect the access that subjects had to the correctness of their responses during categorization and detection. The "classic" measure includes all responses; "SDI" only includes Type I misses and correct rejections; and "SDV" includes only Type I hits and false alarms. Receiver-operating characteristic (ROC) curves were constructed by plotting the cumulative probabilities of confidence in correct versus incorrect trials. Inflection points are plotted from high to low confidence, such that an ROC curve above the diagonal means that subjects had higher confidence for correct than for incorrect trials, which means that metacognition was above chance level. We calculated the area-under-the-curve (AUC) value for each ROC curve (Kanai et al., 2010; Szczepanowski & Pessoa, 2007; Wilimzig et al., 2008), to quantify metacognitive ability (MA). For the classic and SDI measures, MA was significantly higher for the categorization than for the detection task

were specifically driven by a specific Type I stimulus–response combination. Indeed, the analysis revealed that the difference between categorization and detection was driven only by lower metacognition for Type I CRs in the detection task [masked group: AUC = .60, $t(1, 23) = –3.715$, $p = .001$; degraded group: AUC = .58, $t(1, 21) = –2.621$, $p = .016$]. In Fig. 3, one can see that, for instance, the proportion of CR responses for which the highest confidence rating had been given was higher in the categorization task than in the detection task (indicated by the leftmost inflection point). For all other Type I responses, confidence ratings were equally distributed across the two tasks [masked group: hits, AUC = .53, $t(1, 23) = –1.711$, $p = .101$; misses, AUC = .51, $t(1, 23) = –0.320$, $p = .75$; FAs, AUC = .51, $t(1, 23) = –0.249$, $p = .81$; degraded group: hits, AUC = .50, $t(1, 21) = 0.029$, $p = .977$; misses, AUC = .48, $t(1, 21) = 0.781$, $p = .44$; FAs, AUC = .48, $t(1, 21) = 0.959$, $p = .35$]. Thus, subjects were less confident when judging that something was not present (detection) than when judging that something was not a cat as opposed to another animal (fish or bird; categorization).

This result cannot have been caused by differences in either objective performance or the visibility of the masked/degraded objects. As we reported above, both visibility and objective performance did not differ between the tasks. Also, when visibility was calculated for every Type I response separately, no differences were found between detection and categorization for both conditions [degraded group: hits, $t(1, 21) = 0.032$, $p = .97$; misses, $t(1, 21) = 0.115$, $p = .91$; FAs, $t(1,$

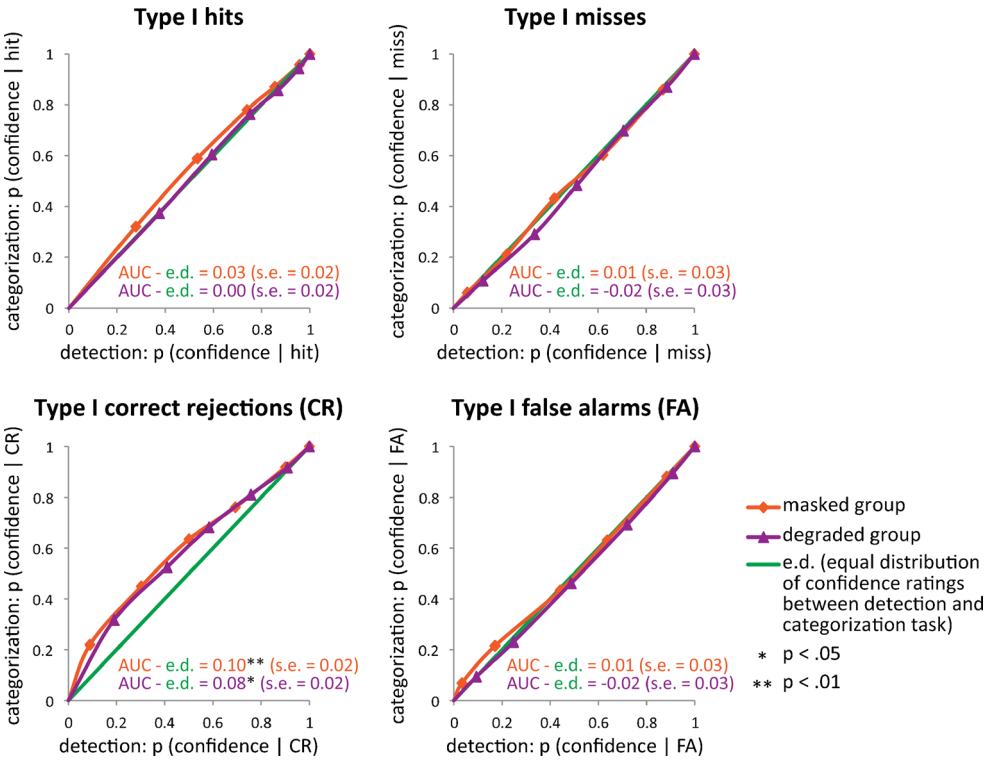

Fig. 3 Metacognition scores separated for all Type I responses, in categorization versus detection. In order to investigate the metacognitive differences between categorization and detection, we zoomed in on the confidence ratings for each Type I response separately. We plotted the cumulative probability of confidence for categorizaton against detection (for hits, misses, CRs, and FAs). An area-under-the-curve (AUC) value of .5 (depicted by the green "e.d." line) means that the confidence ratings were equally distributed across tasks for that Type I response. Only for the CRs were significantly higher confidence ratings provided in the categorization task. The confidence ratings for hits, misses, and FAs were equally distributed across tasks

21) = –0.115, p = .91; CRs, t(1, 21) = 0.118, p = .91; masked group: hits, t(1, 23) = –1.310, p = .20; misses, t(1, 23) = –0.472, p = .64; FAs, t(1, 23) = –0.431, p = .67; CRs, t(1, 23) = –0.836, p = .41].

Additionally, we were interested in the effect of the different perceptual manipulations on metacognition. Remarkably, we did not find any difference in metacognition between masking and degrading. For every measure and task, the MAs were the same for both masked and degraded stimuli: for the classic measure [detection task, t(1, 44) = 1.558, p = .13; categorization task, t(1, 44) = 0.548, p = .59], the SDI measure [detection task, t(1, 44) = –0.234, p = .82; categorization task, t(1, 44) = –0.099, p = .92], and the SDV measure [detection task, t(1, 44) = 0.583, p = .56; categorization task, t(1, 44) = 0.593, p = .56].

Reaction times (RTs)

However, decreased MA for the detection task might have been driven by a speed–accuracy trade-off. To investigate this possible confound, we performed an exploratory RT analysis (note that subjects were not requested to give speeded responses). We calculated the RTs for all correct responses, using the following constraints: Responses had to be made between 200 and 2,000 ms after being prompted (when no response was made after 2,000 ms, the trial was aborted). Also, trials in which a response switch occurred were discarded, to minimize noise in the RT data (since subjects confirmed their response with the space bar after selecting it with the arrow keys, the final response could differ from the initially selected response). Only a small percentage (5 %) of the trials were discarded, and the discarded trials were equally divided over tasks and groups. Just as in the full data set, we observed no between-task differences in percentage correct and d', and metacognition scores did not deviate from those for the full data set.

RTs were significantly longer for the categorization than for the detection task, in both groups [for the degraded group: detection, 1,032 ms (SD = 74); categorization, 1,071 ms (SD = 86), t(1, 21) = –4.972, p = .00006; for the masked group: detection, 1,037 ms (SD = 80); categorization, 1,064 ms (SD = 96), t(1, 23) = –2.807, p = .01]. No between-group differences in RTs were found [detection, t(1, 44) = –0.194, p = .85; categorization, t(1, 44) = 0.291, p = .77].

Since we found both longer RTs and higher metacognition for categorization than for detection, this raises the question whether these increased RTs could explain the higher metacognition scores in this task. However, we did not find any correlations between the RT difference and metacognition differences for categorization versus detection (classic, all rs < .11, ps > .64; SDI, all rs > –.17, ps > .43; SDV, all rs > –.31, ps > .15). We also calculated these correlations separately for all Type I responses, since the metacognition difference between tasks was driven by CRs. We only found *negative* correlations for the RT difference and CR metacognition difference (masked group, r = –.70, p = .001; degraded group,

$r = -.28$, $p = .2$; for the rest of the Type I responses: masked group—hits, $r = -.47$, $p = .02$; misses, $r = -.61$, $p = .002$; FAs, $r = .05$, $p = .8$; degraded group—hits, $r = -.16$, $p = .5$; misses, $r = -.44$, $p = .04$; FAs, $r = -.2$, $p = .4$). In general, the larger the RT difference, the smaller the metacognition difference between the two tasks (this effect was more pronounced in the masked group, but the direction was the same in both groups). Overall, RTs and metacognition seem to be inversely related (see also the supplementary materials for a correlation analysis in which trials were divided into bins of increasing RTs). Therefore, it seems unlikely that the longer RTs in the categorization task were directly related to increased metacognition scores for this task, and no speed–accuracy trade-off seems to have taken place for MA in the detection task.

Discussion

We found that when objective performance was matched for detection and categorization, access to the correctness of these perceptual decisions was not equal. MA was significantly higher for categorization than for detection, according to the "classic" measure (in which all Type I responses were included), and even more so for the more sensitive SDI measure ("subjective discriminability of invisibility," which only includes Type I correct rejections and misses). This difference in metacognition cannot be attributed to differences in the level of visibility of the masked/degraded objects, since visibility did not differ between tasks or manipulations. The fact that MA differed while objective performance was matched suggests that metacognition relies on different information in categorization versus detection, as compared to the information that is used to achieve some objective level of accuracy in these two tasks.

When we zoomed in on metacognition scores for each Type I response separately (hits, misses, correct rejections, and false alarms), it turned out that metacognition only differed between the two tasks for Type I correct rejections. Thus, metacognition is not higher for *all* categorization responses, which makes it unlikely that *all* stimuli were processed more deeply during categorization than during detection. These results show the importance of separating metacognition scores for different Type I responses, instead of aggregating them into a single measure (see also Kanai et al., 2010), because differences might in fact be due to a single Type I stimulus–response combination (in this case, correct rejections) rather than to an overall effect. But what does this metacognition difference for correct rejections mean?

Why would correct rejections cause a different pattern of metacognition between detection and categorization, whereas other stimulus–response combinations do not? With a correct rejection, a stimulus is correctly classified as being a distractor (e.g., in the categorization task, "no, this is not a cat"; in the detection task, "no, there is no animal present"). When it comes to this distractor stimulus, the two tasks are asymmetrical: A distractor does *not* contain any object-related information in the detection task (100 % phase-scrambled object), whereas it *does* contain object-related information in the categorization task (object from a nontarget category). Therefore, with categorization, when correctly classifying a stimulus as *not* belonging to the target category, it is likely that you would successfully categorize the distractor (i.e., "it is *not* a cat, but it *is* a fish"), resulting in high confidence/metacognition (see Supplementary Fig. 2). The distractor stimulus thus provides positive evidence for the presence of a distractor, and therefore a categorization correct rejection is more like a (distractor) "hit." In contrast, during detection the distractor only provides negative evidence, for the absence of a target. This is in line with findings from Zylberberg et al. (2012), which indicated that metacognition is driven only by positive evidence favoring the selected choice, and that metacognition is "blind" to evidence for the nonselected choice. Because there is no "positive" information in the case of target absence, metacognition is lower for detection distractors. We believe that this task/distractor asymmetry is not specific to our experimental design, but is a property inherent to the acts of categorizing and detecting. In real life, categorizing something as being "not A" is based on positive evidence that something is "B" instead, whereas deciding that A is "not present" relies only on negative evidence that no A-like object is present at all.

Interestingly, we showed that this inherent task asymmetry impacts MA, even when objective performance has been matched to control for any effect that it might have. Naturally, one could argue that extra information is present in a categorization task, in the form of evidence from the distractors. This should indeed impact accuracy specifically in the case of correct rejections, since these can be thought of as distractor "hits" in the categorization task. However, this effect was controlled for when performance was matched. Therefore, the information present in the distractors during the categorization task was no more effective than the noninformation present in the distractors during the detection task. Nevertheless, despite the matching performance, subjects were shown to have more information about the correctness of their decisions specifically when they were categorizing. The positive evidence that is contained in the distractor stimulus of the categorization task thus does not help to increase task performance, it only helps one once the categorization decision *has been made*, to feel more confident about it. This is not trivial. For example, this would not be predicted if one were to consider some artificial categorization algorithm. Why would such an algorithm have equal objective performance in two tasks, but more information about the correctness of its decisions in one task in particular? This suggests that something is special about the way the human decision-making machinery

rejects nontargets. If those nontargets have positive evidence, more information about the correctness of its decisions trickles through than is the case for negative evidence, despite the fact that both types of evidence have matching effects when considering objective performance levels. Apparently, the systems underlying metacognitive performance are biased toward positive evidence, which becomes apparent when comparing categorization with detection.

On the other hand, although objective performance levels were matched, we did find significantly longer RTs in the categorization task than in the detection task, for both groups. One could argue that RT is an objective measure as well, which implies that objective performance was in fact not completely matched. Despite equal stimulus visibility and performance for both tasks, the categorization task might have been more difficult. It should be noted, however, that RTs were not speeded; we used them as a post-hoc, exploratory measure to check whether increased metacognition in the categorization task was related to increased RTs. Namely, RT differences would play into this discussion if a speed–accuracy trade-off emerged, showing that longer RTs were correlated with higher accuracy and/or higher metacognition. Importantly, however, we observed the opposite of a speed–accuracy trade-off: If anything, the longer RTs in the categorization task were accompanied by lower objective performance and lower MA (see also Supplementary Fig. 1). Thus, although objective performance may not have been perfectly matched in terms of RTs, it is difficult to see how this could explain our results, because shorter RTs did not lead to lower performance in our task setup. Further research may reveal what happens if RTs are speeded and/or matched, although there is a limit to the number of variables that one can control completely.

Remarkably, different perceptual manipulations, degrading and masking, both had the same effect on the MA of detection and categorization. We did not find any difference in metacognition for masked as compared to degraded objects, for any of our measures (the classic and SDI measure, but also the SDV measure ["subjective discriminability of *visibility*," which only included Type I hits and false alarms]). Although degrading has been suggested to disrupt the objective performance of categorization more than masking (Mack et al., 2008), our results demonstrate that when objective performance is equal, no differences in subjective experience are observed between masked and degraded objects (for both categorization and detection). We did not include any neural measures in this experiment, but it would be interesting to investigate what would happen in terms of cortical processing when the objective and subjective performance of two perceptual manipulations was equal. Masking is thought to interfere with feedback processing, while leaving the initial feedforward activity intact (Di Lollo, Enns, & Rensink, 2000; Fahrenfort et al., 2007; Lamme et al., 2002). Degrading, although it is a bottom-up manipulation affecting the feedforward signals, has been shown to affect feedback processing, as well (Romeo, Arall, & Supèr, 2012). Therefore, we speculate that although early processing stages might differ, the neural "end results" in terms of feedback processing will be equal for masking and degrading, given a particular performance level. Insofar as the feedforward signals differ between masking and degrading, these signals do not seem to be subjectively accessible (VanRullen & Koch, 2003), in line with our observation that measures of metacognition for masked and degraded objects were equal, given that performance levels were matched.

Building on this, one may wonder what type of process MA relies on, given the apparent asymmetry between MA for detection and MA for categorization when performance is matched. One possibility is that MA relies more on recurrent signals than on feedforward signals, as compared to objective performance. It has been shown that both overt selection and the neural signals related to detection and categorization can be based on feedforward information only (Fahrenfort et al., 2012; VanRullen & Koch, 2003). Conscious categorization, on the other hand, has been shown to involve recurrent interactions (Fahrenfort et al., 2012; Koivisto, Railo, Revonsuo, Vanni, & Salminen-Vaparanta, 2011). If MA relies more heavily on these recurrent interactions, this might explain why MA increases when positive evidence exists, such as in categorization. The reasoning behind this would be that the absence of object information (as in detection distractors) would still result in a feedforward signal, but not in recurrent processing, whereas counterfactual information (as in categorization distractors) results in both a feedforward signal and recurrent processing. The implication of this would be that—given some objective level of performance—the amount of recurrent processing that takes place in categorization would be larger than the amount of recurrent processing taking place in detection, and that objective performance would rely more heavily on feedforward signals, whereas MA would depend more on recurrent interactions.

As a side note, the finding is in itself interesting that, with the same stimulus visibility (level of masking/degrading), objective performance was equal for the detection and categorization tasks (without being at ceiling—namely, on average, 73.6 % correct, $d' = 1.43$). This is in accordance with findings by Grill-Spector and Kanwisher (2005), who found that RTs as well as percentages correct are equal for both tasks (with the same stimulus duration). On the basis of these findings, Grill-Spector and Kanwisher put forward the controversial idea that categorization and detection require the same amounts of processing, and might even be the same mechanism. Mack et al. (2008) responded by showing that as soon as the stimuli are inverted or degraded, categorization performance and RTs become worse than those for detection, suggesting that the time courses of these tasks *can* be dissociated. Mack et al. (2008) stated that this is the case because

categorization requires more visual information than does detection, and information is limited when stimuli are degraded. Interestingly, here we showed that even for degraded stimuli, objective performance is equal for categorization and detection (when stimulus visibility is equal, stimuli are phase-scrambled to the same coherence level for these two tasks). This contradicts the findings of Mack et al. (2008), and fits with Grill-Spector and Kanwisher's account instead, although admittedly we did not determine this over a range of exposure times/coherence levels, as had been done in Mack et al.'s (2008) experiments. Moreover, our results showed a difference between categorization and detection in terms of the ability to reflect on accuracy: Metacognition is higher in categorization, showing that the processes involved in MA can conclude more on the basis of "positive" distractor evidence than on "negative" (absent) distractor evidence, suggesting that, at least in terms of MA, detection and categorization are not the same thing.

Conclusions

We have been the first to use metacognition to measure the subjective experiences during detection and categorization while keeping objective performance equal. We found higher metacognition for categorization than for detection, for both masked and degraded objects. Strikingly, this effect turned out to be driven by a difference in metacognition only for Type I correct rejections. This shows the importance of separating metacognition scores for different Type I responses, instead of aggregating them into a single measure (see also Kanai et al., 2010). We propose that this difference is caused by an asymmetry, in terms of the distractor stimulus, that is inherent to the tasks: The distractor does not contain any object-related information in the detection task (100 % phase-scrambled object), whereas it does in the categorization task (object from a nontarget category). With categorization, when correctly classifying a stimulus as not belonging to the target category, it is likely that one successfully categorizes the distractor (a "distractor hit"; i.e., "it is *not* a cat, but it *is* a fish"), resulting in high metacognition. This is in line with recent findings by Zylberberg et al. (2012) indicating that metacognition is driven only by positive evidence favoring the selected choice. Importantly, this inherent task asymmetry only impacted MA, since objective performance had been matched, which makes this a nontrivial finding. It tells us something about human decision making: Apparently, it is easier to be confident that something else *is* there than that something is *not* there, even when objective performance levels are the same. To draw nonspeculative conclusions about the reason why the human metacognitive system is biased toward positive evidence, additional research will be required.

Author note We thank Benedikt Aink, Samira Breukhoven, Michael van den Hoek, Cees Mudde, and Rianne Visser for their help with the data acquisition.

References

Bowers, J. S., & Jones, K. W. (2008). Detecting objects is easier than categorizing them. *Quarterly Journal of Experimental Psychology, 61,* 552–557. doi:10.1080/17470210701798290

Boyer, J. L., Harrison, S., & Ro, T. (2005). Unconscious processing of orientation and color without primary visual cortex. *Proceedings of the National Academy of Sciences, 102,* 16875–16879. doi:10.1073/pnas.0505332102

Breitmeyer, B. G. (1984). *Visual masking: An integrative approach*. New York, NY: Oxford University Press.

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,* 481–507. doi:10.1037/0096-3445.129.4.481

Fahrenfort, J. J., Scholte, H. S., & Lamme, V. A. F. (2007). Masking disrupts reentrant processing in human visual cortex. *Journal of Cognitive Neuroscience, 19,* 1488–1497. doi:10.1162/jocn.2007.19.9.1488

Fahrenfort, J. J., Snijders, T. M., Heinen, K., van Gaal, S., Scholte, H. S., & Lamme, V. A. F. (2012). Neuronal integration in visual cortex elevates face category tuning to conscious face perception. *Proceedings of the National Academy of Sciences, 109,* 21504–21509. doi:10.1073/pnas.1207414110

Fleming, S. M., Dolan, R. J., & Frith, C. D. (2012a). Metacognition: Computation, biology and function. *Philosophical Transactions of the Royal Society B, 367,* 1280–1286. doi:10.1098/rstb.2012.0021

Fleming, S. M., Huijgen, J., & Dolan, R. J. (2012b). Prefrontal contributions to metacognition in perceptual decision making. *Journal of Neuroscience, 32,* 6117–6125. doi:10.1523/JNEUROSCI.6489-11.2012

Fleming, S. M., Weil, R. S., Nagy, Z., Dolan, R. J., & Rees, G. (2010). Relating introspective accuracy to individual differences in brain structure. *Science, 329,* 1541–1543. doi:10.1126/science.1191883

Genetti, M., Britz, J., Michel, C. M., & Pegna, A. J. (2010). An electrophysiological study of conscious visual perception using progressively degraded stimuli. *Journal of Vision, 10*(14), 1–14. doi:10.1167/10.14.10

Gregori-Grgic, R., Balderi, M., & De'Sperati, C. (2011). Delayed perceptual awareness in rapid perceptual decisions. *PLoS ONE, 6,* e17079. doi:10.1371/journal.pone.0017079

Grill-Spector, K., & Kanwisher, N. (2005). Visual recognition as soon as you know it is there, you know what it is. *Psychological Science, 16,* 152–160. doi:10.1111/j.0956-7976.2005.00796.x

Jolij, J., Scholte, H. S., van Gaal, S., Hodgson, T. L., & Lamme, V. A. F. (2011). Act quickly, decide later: Long-latency visual processing underlies perceptual decisions but not reflexive behavior. *Journal of Cognitive Neuroscience, 23,* 3734–3745.

Kanai, R., Walsh, V., & Tseng, C. H. (2010). Subjective discriminability of invisibility: A framework for distinguishing perceptual and attentional failures of awareness. *Consciousness and Cognition, 19,* 1045–1057. doi:10.1016/j.concog.2010.06.003

Koivisto, M., Railo, H., Revonsuo, A., Vanni, S., & Salminen-Vaparanta, N. (2011). Recurrent processing in V1/V2 contributes to categorization of natural scenes. *Journal of Neuroscience, 31,* 2488–2492.

Kolb, F. C., & Braun, J. (1995). Blindsight in normal observers. *Nature, 377,* 336–338.

Lamme, V. A. F., & Roelfsema, P. R. (2000). The distinct modes of vision offered by feedforward and recurrent processing. *Trends in Neurosciences, 23,* 571–579. doi:10.1037/0096-3445.129.4.481

Lamme, V. A. F., Zipser, K., & Spekreijse, H. (2002). Masking interrupts figure–ground signals in V1. *Journal of Cognitive Neuroscience, 14,* 1044–1053.

Lau, H. C., & Passingham, R. E. (2006). Relative blindsight in normal observers and the neural correlate of visual consciousness. *Proceedings of the National Academy of Sciences, 103,* 18763–18768. doi:10.1073/pnas.0607716103

Mack, M. L., Gauthier, I., Sadr, J., & Palmeri, T. J. (2008). Object detection and basic-level categorization: Sometimes you know it is there before you know what it is. *Psychonomic Bulletin & Review, 15,* 28–35. doi:10.3758/PBR.15.1.28

Mack, M. L., & Palmeri, T. J. (2010). Decoupling object detection and categorization. *Journal of Experimental Psychology: Human Perception and Performance, 36,* 1067–1079. doi:10.1037/a0020254

Malach, R., Reppas, J. B., Benson, R. R., Kwong, K. K., Jiang, H., Kennedy, W. A., & Tootell, R. B. (1995). Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex. *Proceedings of the National Academy of Sciences, 92,* 8135–8139.

Op de Beeck, H. P., Baker, C. I., DiCarlo, J. J., & Kanwisher, N. G. (2006). Discrimination training alters object representations in human extrastriate cortex. *Journal of Neuroscience, 26,* 13025–13036.

Romeo, A., Arall, M., & Supèr, H. (2012). Noise destroys feedback enhanced figure–ground segmentation but not feedforward figure–ground segmentation. *Frontiers in Physiology, 3,* 274. doi:10.3389/fphys.2012.00274

Scholte, H. S., Ghebreab, S., Waldorp, L., Smeulders, A. W. M., & Lamme, V. A. F. (2009). Brain responses strongly correlate with Weibull image statistics when processing natural images. *Journal of Vision, 9*(4), 29.1–15. doi:10.1167/9.4.29

Szczepanowski, R., & Pessoa, L. (2007). Fear perception: Can objective and subjective awareness measures be dissociated? *Journal of Vision, 7*(4), 10. 1–17.

VanRullen, R., & Koch, C. (2003). Visual selective behavior can be triggered by a feed-forward process. *Journal of Cognitive Neuroscience, 15,* 209–217. doi:10.1162/089892903321208141

Weiskrantz, L., Warrington, E., Sanders, M., & Marshall, J. (1974). Visual capacity in the hemianopic field following a restricted occipital ablation. *Brain, 97,* 709–728.

Wilimzig, C., Tsuchiya, N., Fahle, M., Einhäuser, W., & Koch, C. (2008). Spatial attention increases performance but not subjective confidence in a discrimination task. *Journal of Vision, 8*(5), 7.1–10. doi:10.1167/8.5.7

Zylberberg, A., Barttfeld, P., & Sigman, M. (2012). The construction of confidence in a perceptual decision. *Frontiers in Integrative Neuroscience*, *6*:79, 1–10. doi:10.3389/fnint.2012.00079