White Rose Research Online

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk
Universities of Leeds, Sheffield and York

Deposited via The University of York.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/id/eprint/120672/

Version: Accepted Version

Article:
Quinlan, Philip Thomas, Yue, Yue and Cohen, Dale (2017) The processing of images of biological threats in visual short-term memory. Proceedings of the Royal Society B: Biological Sciences. ISSN: 1471-2954

https://doi.org/10.1098/rspb.2017.1283

Reuse
Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

UNIVERSITY OF LEEDS
University of Sheffield
UNIVERSITY of York

PROCEEDINGS OF
THE ROYAL SOCIETY B
BIOLOGICAL SCIENCES

The processing of images of biological threats in visual short-term memory

Journal:	*Proceedings B*
Manuscript ID	RSPB-2017-1283.R1
Article Type:	Research
Date Submitted by the Author:	n/a
Complete List of Authors:	Quinlan, Philip; The University of York, Psychology Yue, Yue; Liverpool Hope University, Psychology Cohen, Dale; The University of North Carolina at Wilmington, Psychology
Subject:	Cognition < BIOLOGY
Keywords:	Visual short-term memory, memory for biological threats, emotional memory enhancement, emotional distinctiveness
Proceedings B category:	Neuroscience & Cognition

SCHOLARONE™
Manuscripts

Running Head: Biological threats and VSTM

The processing of images of biological threats in visual short-term memory

Philip T. Quinlan

The University of York, UK

Yue Yue

Liverpool Hope University, UK

Dale J. Cohen

The University of North Carolina at Wilmington, USA

Author Note

Please address any correspondence regarding the work to Philip Quinlan, Department of Psychology, The University of York, Heslington York, North Yorkshire, YO10 5DD, UK. Email: philip.quinlan@york.ac.uk. We would like to express our sincere thanks to David Watson for help with using MATLAB.

Abstract

The idea that there is enhanced memory for negatively, emotionally charged pictures was examined. Performance was measured under rapid, serial visual presentation (RSVP) conditions in which, on every trial, a sequence of six photo-images was presented. Briefly after the offset of the sequence, two alternative images (a target and a foil) were presented and participants attempted to choose which image had occurred in the sequence. Images were of threatening and non-threatening cats and dogs. The target depicted either an animal expressing an emotion distinct from the other images, or the sequences contained only images depicting the same emotional valence. Enhanced memory was found for targets that differed in emotional valence from the other sequence images, compared to targets that expressed the same emotional valence. Further controls in stimulus selection were then introduced and the same emotional distinctiveness effect obtained. In ruling out possible visual and attentional accounts of the data, an informal dual route topic model is discussed. This places emphasis on how visual short-term memory reveals a sensitivity to the emotional content of the input as it unfolds over time. Items that present with a distinctive emotional content stand out in memory.

Keywords: Visual short-term memory; memory for biological threats; emotional memory enhancement; emotional distinctiveness.

1. Introduction

Within evolutionary psychology it is claimed that the adaptive usefulness of being able to detect rapidly the emotional content of the current visual scene serves prey in their effective avoidance of predators. Within cognitive psychology such ideas have been explored in terms of the *fear response hypothesis* [1, 2]. The hypothesis states that humans have evolved a *fear system* [2], which is rapidly and automatically elicited by the presence of threat in the immediate environment. Its prime purpose is to produce an automatic early warning signal that alerts the observer to the threat.

Much of the empirical support for the fear response hypothesis has come from studies of speeded visual search. In the typical case, the participant must respond 'Same' as quickly and as accurately as possible if all the images in a search display depict instances of the same kind and 'Different' if one of the images is taken from a distinct kind. The very rapid detection of the presence of the 'threatening' singleton (termed *pop-out*), such as a snake image amongst images of flora [3, 4] has been taken to support the fear response hypothesis. The fact that a non-threatening singleton does not pop-out from a background of images of threatening instances has also been taken to suggest a form of attentional bias towards threatening stimuli [3]. Since the original paper was published [3], many other studies have followed up on the findings, and in 2013, Quinlan [5] reviewed a total of 18 cases comprising 44 experiments.

The present work addresses a related but different set of issues concerning the processing of visual images of biological threats. Central here is how the emotional content of an image determines how well it is remembered. In the same way that arguments can be made about the adaptive usefulness of the rapid visual detection of threat, it can also be argued that keeping track of the location and nature of a threat in the immediate vicinity are

also key to the avoidance of predation. Mechanism that are held responsible for retaining visual information over brief intervals are collectively discussed in terms of visual short-term memory (VSTM). Two assumptions about this kind of memory system are key, namely, (i) that the store is of very limited capacity [6], and, (ii) that the contents of the store are maintained for only a matter of seconds (see e.g., [7]). On the understanding that the memory system under consideration holds information for only a matter of seconds, the images that were presented for retention in the present experiments were presented very rapidly.

Harris and Pashler [8] addressed whether memory enhancement for negatively charged pictures might be due to selective rehearsal of, or, rumination about these kinds of images. An RSVP paradigm was used in which, on each trial, a sequence of five photo-images was presented with each image only being exposed for 250 ms. The task was to provide, at test, one-word descriptors of each of the pictures. The rate of presentation of the images in the sequence was taken to be so rapid as to make any form of rehearsal unlikely/impossible. Testing occurred either immediately or following a filled delay in which participants had to engage in counting backwards in sevens for 20 s. Hence any form of picture rehearsal/rumination was circumvented. The critical picture sequences contained four neutral photo-images and one negatively charged image and the results were clear in showing a robust recall advantage for the negative pictures relative to the neutral pictures. As these effects occurred in the absence of either selective rehearsal or rumination, the claim was that this evidence was indicative of "enhancement of storage due to emotional content" ([8] p. 196).

The results of [8] accord with other sentiments that, at a psychological level, emotional stimuli are special. For instance, in line with the arguments made on the basis of visual search data, claims have also been made [9] that there is an attentional bias towards emotional stimuli such that these, preferentially, enter into, and, are preferentially, maintained

in VSTM. Indeed, attentional factors are repeatedly called upon to explain emotional memory enhancement in the context of the RSVP paradigm and these are examined here. For instance, in [10] Malijkovic and Martini provided evidence that the rate of information accrual varied across different image types when RSVP techniques were used. In their experiment, sequences of ten images were used such that the duration of the images varied from 13 ms to 4 s. Sequences contained a mixture of images with neutral, positive, or negative valence. Participants were then given an old/new recognition task and judged whether test images had been included in the original sequence. Following a detailed psychophysical method, it was concluded that, relative to neutral and positive images, the processing of negative images was distinctive. For negative images, the rate of information accrual accelerated as the presentation duration of the image increased. This was taken to be indicative of attentional mechanisms such that "image features most diagnostic of negative content attract most of the early processing resources" (p. 6). Mean luminance, contrast, and spatial frequency differences within and across the images were also examined [10], but there were no systematic links between these kinds of low-level visual statistics and performance in the RSVP tasks. The eventual account rests on assumptions about the functional significance of attention to image features that are diagnostic of negative content, but having ruled out low-level visual features, it is unclear what the "diagnostic features" actually are.

From the cited studies, an overall impression is that the memorial processes associated with negative images are special. Some might resist this conclusion on the grounds of other evidence showing a general emotional VSTM enhancement for both positive and negatively charged images relative to neutral images [11]. However, in that study the positive images were erotic images so whether the data underscore a general principle about the operation of VSTM remains to be seen. Nonetheless, in cases where there is enhanced memory for negatively charged pictures, selective rehearsal/rumination has been ruled out as

an explanation by [8]. The favoured account is one in which negatively charged image content, 'quickly enhances memory storage or accessibility' (p. 197). This account dovetails with that in [10] where it is argued that negatively-charged images are, in some sense, prioritized when they are presented. Negatively charged images gain privileged access to VSTM, and, as a consequence, accrue enhanced storage. Fundamentally, attentional factors are called upon to explain these effects.

There is, however, an alternative interpretation. In the picture sequences used in [8] we may refer to the negatively charged picture as *the target* or *the isolate* and draw parallels with other experiments on the, so-called, *isolation effect* (or *von Restorff effect*). Since the work of von Restorff (see [12]) it is well-established that memory for a distinctive item, in an otherwise homogenous list of items, is considerably better than it is for an item in a wholly homogenous list. Although the effect was originally examined in the context of perceptual distinctiveness, it is now well established that similar effects can be obtained when distinctiveness is defined in terms of semantics (see e.g., [13, 14]).

The semantic isolation effect in [8], shows that memory for a negatively charged isolate (the target) is enhanced when tested against other neutral (background) items. What remains unknown, however, is whether such an effect emerges when the valence of the target and the background items is reversed. Is it simply the case that VSTM favours negatively charged targets, in this case biological threats, or, is it that the system is sensitive to a distinctive threat image when tested against a background of items of a neutral valence? Such questions are pertinent given the claims about attentional bias towards threatening images that have arisen in the visual search literature [3]. If the reason for enhanced memory for negatively, emotionally charged pictures is due to some kind of attentional bias towards such stimuli then the expectation is that the isolation effect should only obtain when a 'threaten' isolate is situated amongst a background of neutral images. A similar effect ought not to arise

if a neutral isolate occurs amongst a background of threatening images: from the visual search literature, it has been shown that non-threatening singletons do not pop-out. These predictions are tested in the following experiments.

2. Experiment 1

The RSVP paradigm adopted here is illustrated in Figure 1. Once the sequence was presented and following a brief delay, memory was probed with the presentation of two images: a *target* that had occurred in the sequence and a *foil* that had not. Participants attempted to recognise the target. In using very rapid presentation with immediate test, it is assumed (following [8]), that neither selective rehearsal nor rumination (i.e., strategic intervention) can be used to explain performance in the tasks. It is also notable that in the paradigm used in [8] the participants were being tested on their memory for all the images that had been presented and they were being tested in a way that invoked verbal recoding. In the present case, memory for a single image was being probed and the task did not depend on verbal processing: The target image was re-presented at test. On these grounds, we assume that performance in the experiments provides a direct reflection of visual, and not, verbal memory processes.

(a) Method

(i) Participants

Thirty-two participants were tested in total. They were all students at the University of York. Testing was carried out over a fixed period during term time and the eventual sample size was set at the number of individuals who could be tested within this two-week period. Participants were either paid for taking part or were provided with course credit. Having fixed the sample size at 32, this then operated as the sample size for the other experiments.

(ii) Stimuli and design

The stimuli were taken from [15]. There were four sets of images each comprising 48 instances. Two image sets were of cats and two were of dogs and, within these sets, half of the images were of non-threatening animals and half were of threatening images. The non-threatening images depicted domesticated animals. Every threatening image depicted a snarling animal. The importance of images of snarling animals can be traced back to Darwin [16] who, in his discussion of the expression of rage, described how the bearing of teeth in a snarl is a signal of a preparedness to attack. Predominantly the threatening images were of wild animals, but a small number of the images of dogs depicted snarling attack dogs. In [15] steps were taken to attempt to control for the general pose of the animals across the threatening and non-threatening cases in a bid to control for any gross visual differences across these image sets. In addition, in [15] independent ratings of the threat of images were taken and showed that the threatening images were rated more threatening than the non-threatening images.

The basic sequence of events on a trial comprised a RSVP of six centrally presented images followed by a 2AFC screen in which an upper and lower pair of images was presented. Each of the individual images measured 4^o x 6^o. The 2AFC images were respectively located 4^o degrees above and below the screen centre. One of the images (the target) had occurred in the sequence and the other (the foil) had not. The up/down position of the target was determined at random prior to the start of each trial.

Across the trials the target occurred equally often at each sequence position. For ease of exposition, the items other than the target that were included in the memory sequence will be referred to as *background images*. Central interest is with three factors, namely, *category* defined relative to two levels – same vs. different, that is, whether the biological category of

the target (cat vs. dog) matched with that of the background images; *type of target* – non-threatening vs. threatening; and, *type of background image* – non-threatening vs. threatening. The design encapsulated a factorial combination of these three factors. In brief, the experimental trials comprised all possible combinations of type of target and type of background image and whether the target and background images were taken from the same biological category.

Given the constraints of a particular trial, the background images were chosen at random from the 48 for that particular image set. In addition, the target and foil were chosen at random from the same particular image set. As a consequence, it was assumed that any simple verbal labelling strategy would be rendered useless. For instance, merely, noting (and labelling) a 'wild cat' amongst 'domesticated cats' would be of little help as the 2AFC screen contained two images of 'wild cats'.

(iii) Procedure

Each participant was tested individually, in a quiet testing room at The University of York. Participants sat a table in front of a chin rest fixed 57 cm from a computer screen (model ProNitron 17/200). The experiment was controlled by a PC running an E-Prime script [17]. Responses were made via a keyboard such that pressing '1' indicated that the upper image was the target and pressing '2'indicated that the lower image was the target. Participants were able to respond with whichever hand they felt most comfortable with.

An initial block of 6 practice trials was administered prior to 5 blocks of 96 experimental trials. Participants were instructed that accuracy rather than speed was important and to guess if they were unsure. They were also instructed not to spend too long on any trial, as there were many trials. They were also initially forewarned about the

'threatening nature of some of the images' and were told not to continue, after the practice block, if they were finding the images uncomfortable.

(b) Results and discussion

The data were scored and broken down according to three factors, category, type of target and type of background image. For a given participant, each cell in this design gave rise to a total of 60 observations hence the scores that were analysed were items correct out of 60. Figure 2 provides a graphical illustration of the summary data.

The scores were entered into a three-way repeated measures ANOVA in which category (i.e., same vs. different), type of target (i.e., non-threatening vs. threatening), and, type of background image (i.e., non-threatening vs. threatening) were entered as fixed factors and participants was entered as a random factor. The analysis revealed statistically significant main effects of category, $F(1, 31) = 5.89$, $p < .05$, *partial* $\eta^2 = .160$, type of target, $F(1, 31) = 6.26$, $p < .05$, *partial* $\eta^2 = .168$, and, type of background image, $F(1, 31) = 4.32$, $p < .05$, *partial* $\eta^2 = .122$. In addition, the type of target x type of background image interaction also reached statistical significance, $F(1, 31) = 25.00$, $p < .001$, *partial* $\eta^2 = .446$. None of the remaining interactions were statistically reliable; $F < 1.0$, for the category x type of target interaction; $F(1, 31) = 1.99$, $p > .05$, *partial* $\eta^2 = .060$, for the category x type of background image interaction; and, $F < 1.0$, for the category x type of target x type of background image interaction.

In sum, the main effect of category revealed that participants were more accurate in their reports when the target and the background images were taken from different biological categories than when they were taken from the same category. This particular result will be returned to later. The more important effect is the statistically reliable type of target x type of background image interaction. This revealed that participants were much more accurate in

reporting the target when the valence of the type of target and type of background image differed than when they matched. In other words, targets were much better reported in cases where they signalled a distinctive valence than when they did not. Figure 2 clearly reveals the nature of this two-way interaction.

It is true that the data do show emotional memory enhancement for negatively charged images – a result that replicates the findings of [8]. Memory for a threatening isolate in a sequence of neutral images is enhanced relative to when only neutral images are used. However, this effect must be considered in the context of the further finding that there is a similar memory advantage for a neutral isolate contained in a sequence of threatening images. The conclusion, therefore, is that it is not threat content *per se* that is the critical factor, but the match of the emotional content of the target image to its background. Isolates that signal a distinctive emotional valence are better remembered than are items that have the same valence as the other items. This particular finding casts doubt on the more generally received idea; namely, that an attentional bias towards negatively emotionally charged images explains enhanced memory for such images.

2. Experiment 2

In order to explore the replicability of the pattern of effects a follow-on experiment was undertaken in which the trials were blocked by type of background image. That is, testing was divided into two parts such that, in one part all of the background images were threatening, and, in a separate part, all of the background images were neutral.

Since the seminal work of Schneider and Shiffrin [18] it is well known that performance varies dramatically according to whether the task is defined in terms of consistent vs. varied mappings. In their seminal RSVP studies, consistent mappings were cases where both the target and non-target categories remained constant over trials. In

contrast, under varied mapping constraints, the target and non-target categories interchanged randomly over trials. The data were interpreted in terms of the distinction between automatic and controlled attentional processes. Performance in the varied mappings conditions reflected a much more strategic form of processing than did performance with consistent mappings.

It was of interest to see whether performance in the current tasks is differentially sensitive to the composition of the trials within a testing session. In the verbal memory literature, it has been argued that attention can be focused better in blocked than in mixed conditions, "because an a priori-determined selection set reduces trial-by-trial uncertainty." ([19], p. 314). On these grounds, participants were now tested under constant mapping conditions. On the understanding that there is an attentional bias towards negatively emotionally charged images, it was predicted that the blocking manipulation ought to heighten this bias. That is, with these testing conditions it was predicted that a greater isolation effect for the threatening vs. non-threatening isolates would occur.

In contrast to this, the data from a follow-on experiment (see Supplementary Materials) revealed a very similar pattern of effects to those reported in Experiment 1. Critically, the same type of target x type of background image interaction recurred: participants were much more accurate in their target reports when the emotional valence of the target and background images differed than when they were the same. As the nature of this interaction did not change from the first experiment, then the evidence is not consistent with the idea that the isolation effects primarily reflect the operation of attentional processes. The same kind of interaction arose across both constant mapping and varied mapping conditions (cf. [18]).

The blocking manipulation used in the follow-on experiment was reused in Experiment 2 reported here. A remaining concern was whether or not the critical type of

target x type of background image interaction truly reflects the operation of mechanisms that are sensitive to the emotional content of the images or whether the threat content of the images is carried by distinctive low-level visual characteristics (cf. [10]). In order to examine this, a gist description of each of the images was derived using the techniques described in [20]. This technique entails converting each image first into a grey scale luminance 'pixel' matrix and then into a vector that carries information about local image luminance and spatial frequency. Inter-image similarity can be computed from these gist descriptors by simple linear correlation between the corresponding vectors for the images. A high vector correlation indicates a high-level of image similarity. These correlative techniques were deployed in the analysis of the 2AFC data.

Care was also taken to reduce any other confounding differences between the threatening and non-threatening images used previously. In Experiment 1, the non-threatening images were of domesticated cats and dogs whereas, predominantly the threatening images were of wild cats and dogs. Steps were now taken to source new non-threatening images so that the same pose and species was matched across threatening and non-threatening pairs of images. In this way, the aspiration was to arrive at a set of images in which the only systematic difference across the threatening and non-threatening cases was the emotional valence conveyed.

Ratings of both threat and the arousal were collected for the complete set of images used in Experiment 2 (see Table 1). The threat ratings were entered into a two-way repeated measures ANOVA in which valence (threatening vs. non-threatening) and animal (cat vs. dog) acted as fixed factors and participants acted as a random factor. Only the main effect of valence, $F(1, 23) = 670.03$, $p < .001$, *partial* $\eta^2 = .967$, reached statistical significance; $F(1, 23) = 1.28$, $p > .05$, *partial* $\eta^2 = .053$ for the main effect of animal; $F < 1.0$ for the valence x

animal interaction. A similar analysis of the arousal ratings revealed that only the main effect of arousal reached statistical significance, $F(1, 23) = 5.37$, $p < .05$, *partial* $\eta^2 = .183$; $Fs < 1.0$ for the main effect of animal and the arousal x animal interaction. In sum, the threatening images were rated as being highly threatening and more threatening than the non-threatening images. In addition, the threatening images were rated as being more arousing than the non-threatening images. The exact same patterns of differences were observed for the cat and the dog image sets.

(a) Method

Aside from minor details, the same methods to those used in Experiment 1 were used again. Testing was divided into two parts in which the trials were blocked by kind of background image (either threatening or non-threatening). Each part consisted of one short block of 6 practice trials followed by 5 blocks of 48 experimental trials. The order of the parts was counterbalanced across participants and the two parts were administered in a single testing session that lasted approximately 45 mins.

(i) Participants

The experiment was run at the University of York and new sample of 32 volunteers was tested.

(ii) Stimuli

The same threatening images were used as before but a new set of 96 non-threatening images was sourced. A non-threatening counterpart was sourced for each threatening image by matching the depicted species and pose. For every image a gist description vector was computed using the Matlab code provided in [20] (see, http://people.csail.mit.edu/torralba/code/spatialenvelope/). Each image was passed through a

series of Gabor filters spanning eight orientations and four spatial frequencies giving rise to 32 filtered images. Each of these 32 images was then divided into a 4 x 4 grid and pixel intensities were averaged within each grid. This method transforms a given image into gist descriptor vector of length 512 (i.e., 4 x 4 x 32).

(b) Results and discussion

The data were analysed as before. In this case, only the type of target x type of background image interaction, $F(1, 31) = 16.64$, $p < .001$, *partial* $\eta^2 = .348$, reached statistical significance, all other tests $F < 1.0$. As can be seen from Figure 3, the nature of the interaction is as described before, participants were considerably better in reporting targets when the valence of the target and the background images differed than it was the same.

(i) Subsidiary analysis

The previous datasets only contained records of the type of target and the type of background. In pursuing the more detailed analysis concerning possible visual confounds, each participant's dataset now contained a record of every image that had been presented on every trial. Initially, a correlative analysis was carried out for every experimental trial for every participant. That is, every pairwise (Pearson) correlation was computed between the target gist descriptor and the five other image gist descriptors in its trial sequence and the average correlation for that trial was recorded. An assumption here was that the correlation between any two image vectors indexed the amount of visual similarity between the corresponding images. The visual distinctiveness of any given target was therefore indexed by the corresponding average correlation – a low average correlation therefore would signal a highly distinctive target.

The overall dataset now contained information about every single trial in the experiment coding the participant, category, type of target, and type of background image factor as before. Additional codes captured whether the trial response was correct or not, and, what the average image correlation for that trial was. This dataset was then analysed using a mixed-model ANOVA with participants as a random factor (using R, The R Core Team, [21]). In this analysis, the average image correlation was treated as a continuous variable and entered first into the analysis. The analysis revealed that the only reliable predictor in the model was the type of target x type of background image interaction, $F(1, 124) = 47.50$, $p < .0001$; all of the other predictors, $F < 1.0$.

In sum, this analysis replicates the key type of target x type of background image interaction found previously, but does so when the visual similarity between the target and its background items is taken into account. In this regard, the results replicate the findings of Malijkovic and Martini [10] in showing that low level visual characteristics of the images appear not to determine how effectively they are remembered.

3. General discussion

Previous evidence reported in [8] suggests that there is enhanced memory for negatively emotionally charged pictures without selective rumination. Here we show that this is only a very partial characterization of exactly how emotional content is dealt with by VSTM. Although we have repeatedly demonstrated enhanced memory for images of biological threats without selective rumination, we have also shown that there is enhanced memory for any picture whose emotional valence is different from its background. That is, non-threatening images were better reported when they were presented in the context of threatening background images than when they were presented in the context of other non-threatening images.

The key finding is that images are better remembered when they *differ* in emotional valence from the other to-be-remembered (*tbr*) images with which they are presented than when they share the same valence. A central conclusion therefore is that the mechanisms that underpin performance in the task are highly sensitive to the emotional valence of the visual input but in ways that are not captured by claims about negative emotional memory enhancement. Moreover, the effects cannot readily be explained by visual encoding mechanisms because we have shown that the visual salience of the targets fails to account for the basic pattern of findings (see Experiment 2). In turn, the findings do not support an attentional bias towards negatively charged images because (i) the effects depend on both the type of target and the type of background images, and, (ii) the effects are, essentially, identical across consistent mapping (Experiment 2) and varied mapping (Experiment 1) conditions. On these grounds, we take it that the locus of the effects resides elsewhere.

It is important to note that the effects reported here accord well with other semantic isolation effects in the literature [12, 13, 14]. In the present experiments, the isolate images are defined relative to the emotional valence of the other images in the sequence. In this regard, it is not so much semantic category knowledge that is influential, but the threat content of the images themselves. This is particularly true of the images in the Experiment 2 where the only systematic factor to discriminate the isolate from its background is threat content: Memory for an isolate image was enhanced when its emotional valence stood out from its background. There is no simple visual attentional account of these findings. Alternative accounts might be sought in terms of item maintenance. For instance, it might be argued that the effects of emotional content reflect some form of privileged item maintenance for items of a distinctive emotional valence in the visual short-term store. The data are consistent with such a view, but it is perhaps more fruitful to think in terms of other models of memory that provide detailed and ready accounts of the findings.

Of particular interest are the ideas put forward by Steyvers and Griffiths [22] in their account of the semantic isolate effect described in [13] for memory for word lists. In discussing processes of word memory, they considered *dual route topic models*. By way of explanation, they discussed a distinction between gist and verbatim information such that the human memory system is seen to be differentially sensitive to both. Gist information, essentially, captures a gross summary of the content of the material, for instance, that the list of tbr words contained the names of vegetables. In contrast, verbatim information captures the identities of the actual tbr items (i.e., the list contained 'carrot', 'onion', 'swede', 'cabbage', 'kale').

According to the dual route topic model, a given item can be encoded via a *topic route* or a *special word route*. If there is a coherent context for encoding an item in a particular way, then this will define the current topic route (this defines the current gist). If there is no such context or if an item stands out from the current context, then the special word route comes into play (and this captures verbatim information). The manner in which information about a word is stored in memory reflects the route by which the item was encoded. In the dual route topic model described in [22], scale values are propagated on two distributions for each word - a topic distribution and a special word distribution. In cases of a coherent context, all words will score highly and evenly on the topic distribution but only minimally on the special word distribution. In contrast, in cases where an otherwise homogenous list of words contains an isolate, then the isolate will score particularly highly on the special word distribution whereas the other items will score minimally on this distribution. This salient score on the special words distribution then underpins the enhanced memory for the isolate at test. Although only a summary of the account is provided here, Steyvers and Griffiths [22] did provide qualitative fits with the semantic isolation effect

reported by [13] when they explored the properties of a mathematical account of item encoding and retrieval.

Such a dual route model provides a useful framework for thinking about the current effects. The model can be applied on the assumption that the topic generated by the image sequence on a given trial captures the valence of the images, that is, the topic captures the predominant emotional content of the background images. We may also posit both topic and verbatim (special image) distributions. As a consequence, when a sequence contains an isolate, this image score highly on the special image distribution and this underpins the memory enhancement for these kinds of items. There is no reason to think that simple extensions to the basic mathematical model could not be undertaken to encompass the present data. However, such an exercise is beyond the scope of the present work.

What is key is that the dual route topic model provides a ready explanation of the pattern of results found here. A foundational assumption, though, is that the emotional content of the images – the emotional gist –is recovered with great speed. Such an idea accords well with recent dual pathway account of the human visual system put forward by Wolf and colleagues [23]. In this account, a division is drawn between selective and a non-selective processing pathways. It is via the selective pathway that "candidate objects must be individually selected for recognition" (p. 77), whereas it is via the non-selective pathway that global scene semantics are rapidly derived. Our data accord with the operation of such a non-selective pathway and go further in suggesting that, in turn, VSTM is critically sensitive to how scene semantics change over time. As Wolf et al. stated, the operation of this non-selective pathway is *preattentive* in nature (p. 81) and therefore outside of executive control. On the understanding that the non-selective pathway recovers the emotional content of an unfolding scene, then the present data show that it is not threat *per se* that is special but any

stimulus that differs from the predominant emotional content. As a consequence, we claim that images that deviate in emotional content from the current context stand out in memory.

The central finding here is that it is emotional distinctiveness rather than emotional *per se* that is important in VSTM. Appeals to item distinctiveness are common in theories of short-term memory [24]. Here the claim, that emotionally distinctive items stand out in memory, only makes sense only on the basis of the sorts of gist processing mechanisms that we have discussed. We see these as being different from attentional mechanisms.

It would be remiss not to mention the effects of category that were found. Looking across all of the experiments, the effects are not expressed consistently and in Experiment 2 there were no such effects. Further work is therefore needed to unpack how such conceptual information is retrieved and maintained over the short-term (cf. [25, 26]). The current data do, however, suggest that the recovery and storage of categorical identity reflects mechanisms different from those that are concerned with processing emotional content.

It would also be remiss not to mention the topic of face processing, given that a given facial expression can convey a biological threat. The data from visual search studies are equivocal (see [27], Chapt. 16), as there is evidence both for the rapid detection of an angry facial expression (an anger superiority effect [28]) and evidence for the rapid detection of a happy facial expression (a happiness superiority effect [29]). However, there is more convincing evidence suggesting a VSTM advantage for angry expressions over less negative emotions [30, 31]. Currently work is under way to examine these effects in the context of the RSVP procedures used here. Given the importance of being able to interpret and keep track of social signals in conversation, it would be surprising to find that facial expression processing in VSTM is not special.

(a) Additional considerations

In the foregoing, much has been made in distinguishing 'attentional' from 'memorial' processes. Whether such a simple distinction is helpful is moot especially given the intricacies of the interactions between cortical and sub-cortical mechanisms in the brain [32]. Nonetheless, in using the term "attentional" here, reference is to possible effects arising because of stimulus selection at encoding. We claim that we have ruled out perceptual distinctiveness accounts of the effects as considered in detail in Experiment 2. We have also considered the degree to which the effects reflect an attentional bias towards negatively emotionally charged stimuli: the blocking manipulation in Experiment 2 did not alter the central findings reported in Experiment 1. On these grounds, we do not accept that the effects we report are, in these ways, attentional in nature.

We have, however, argued that the key visual difference between the threatening and non-threatening images is the presence of the snarl and it is this that marks out the isolate. We may therefore ask about the degree to which attention to this characteristic is at play. From the visual search literature, it is well established that it is far easier to register the presence of a critical feature than its absence (see for instance, the data on search asymmetries [33]). On these grounds, it is predicted that an isolate defined by the presence of a snarl would be remembered better than an isolate defined by its absence, if the effects were solely attentional in nature. However, the data show no such asymmetry. We, therefore, conclude that although the presence of a snarl conveys the threat content of an image it is not attention *per se* to this visual characteristic that is key: What is key is how the threat content of the images affects memory.

Other ideas about selective attention to items as maintained in VSTM have begun to emerge in the literature (see e.g., [34]). In this way, the simple distinction between attentional and memorial processes becomes blurred. The extant data [34] suggest that in some circumstances (i.e., when four simple coloured geometrical shapes are presented once every

500 ms in sequence) participants can exert some control over memory for the items. In contrast, here we have tested some traditional ideas regarding attention as a process of stimulus selection during encoding see [35] and have found these to be wanting. Our current position is that the present data sit most comfortably with extant ideas about the operation of preattentive mechanisms and memory over the short term. Future work could usefully examine the degree to which visual memory for emotionally-charged materials can be determined by subjective control.

We are also mindful of the issues over whether 'emotional' effects arise because of the valence of a stimulus or its arousing nature [36]. We note that our threatening images have been rated as being both more threatening than the non-threatening images but also that the threatening images have been rated as being more arousing than our non-threatening images. Clearly further research could address this issue directly by systematically combining these two factors in new experiments. What we can conclude from the present findings, however, is that it is not the level of threat or level of arousal of the isolate that is key, but whether the isolate deviates from the other items in terms of its content.

4. Conclusions

Here we report a novel set of findings that show that VSTM is particularly sensitive to rapidly unfolding scene semantics - items that convey a contrasting emotional valence to the background stand out in memory. In the past, it has been shown that there is emotional memory enhancement for negatively charged images, but here we show that it is distinctive emotional valence that is key not threat *per se*. In adopting a dual route topic account from the verbal memory literature, we cast light on a basic characteristic of how memory in general operates over brief intervals. We offer an account that implicates memorial rather than attentional processes.

Ethics. Ethical approval for the experimental work was obtained from the Departmental Ethical Committees of the Department of Psychology at The University of York and of the Department of Psychology at Liverpool Hope University prior to data collection.

Data accessibility. The data are available on the Open Science Framework - osf.io/73tmf

Author contributions. P.Q. as the principal investigator instigated this work and was primarily responsible for the data collection, generation of the E-prime scripts and production of the manuscript. Y.Y. provided further help with E-prime and was primarily responsible for data collection in the supplementary experiment. D.C provided further help and advice and was responsible for the subsidiary statistical analyses reported in Experiment 2. All authors contributed to the final write-up.

References

1. Öhman A. Distinguishing unconscious from conscious emotional processes: Methodological considerations and theoretical implications. In: Dalgleish T, Powers MJ, editors. Handbook of cognition and emotion. Chichester, England: Wiley; 1999. p. 321-52.

2. Öhman A, Mineka S. Fears, phobias, and preparedness: Toward an evolved module of fear and fear learning. Psychol Rev. 2001;108:483-522. (http://dx.doi.org/10.1037//0033-295X.108.3.483)

3. Öhman A, Flykt A, Esteves F. Emotion drives attention: Detecting the snake in the grass. J Exp Psychol Gen. 2001;130:466-78. (http://dx.doi.org/10.1037//0096-3445.130.3.466)

4. Tipples J, Young AW, Quinlan PT, Broks P, Ellis AW. Searching for threat. Q J Exp Psychol. 2002;55A:1007-26. (http://dx.doi.org/DOI:10.1080/02724980143000659)

5. Quinlan PT. The visual detection of threat: A cautionary tale. Psychon Bull Rev. 2013;20:1080-101. (http://dx.doi.org/10.3758/s13423-013-0421-4)

6. Cowan N. The magical number 4 in short-term memory: A reconsideration of mental storage capacity. Behav Brain Sci. 2001;24:87-185. (http://dx.doi.org/10.1017/S0140525X01003922)

7. Gorgoraptis N, Catalao RFG, Bays PM, Husain M. Dynamic updating of working memory resources for visual objects. J Neurosci. 2011;31:8502-11. (http://dx.doi.org/10.1523/JNEUROSCI.0208-11.2011)

8. Harris CR, Pashler H. Enhanced memory for negatively emotionally charged pictures without selective rumination. Emotion. 2005;5:191-9. (http://dx.doi.org/10.1037/1528-3542.5.2.191)

9. Reinecke A, Rinck M, Becker ES. Spiders crawl easily though the bottleneck: Visual working memory for negative stimuli. Emotion. 2006;6:438-49. (http://dx.doi.org/10.1037/1528-3542.6.3.438)

10. Malijkovic V, Martini P. Short-term memory for scenes with affective content. J Vis. 2005;5:215-29. (http://dx.doi.org/10.1167/5.3.6)

11. Lindström BR, Bohlin G. Emotion processing facilitates working memory performance. Cogn Emot. 2011;25:1196-204. (http://dx.doi.org/10.1080/02699931.2010.527703)

12. Hunt RR. The subtlety of distinctiveness: What von Restorff really did. Psychon Bull Rev. 1995;2:105-12. (http://dx.doi.org/10.3758/BF03214414)

13. Hunt RR, Lamb CA. What causes the isolation effect? J Exp Psychol Learn Mem Cogn. 2001;27:1359-66.

14. Bireta TJ, Mazzei CM. Does the isolation effect require attention? Mem Cognit. 2016;44:1-14. (http://dx.doi.org/10.3758/s13421-015-0538-y)

15. Yue Y, Quinlan PT. Appraising the role of visual threat in speeded detection and classification tasks. Front Psychol. 2015;6:755. (http://dx.doi.org/10.3389/fpsyg.2015.00755)

16. Darwin C. The expression of the emotions in man and animals. London, UK: James Murray; 1873.

17. Schneider W, Eschman A, Zuccolotto A. E-Prime user's guide. Pittsburgh, PA: Psychology Software Tools Inc.; 2002.

18. Schneider W, Shiffrin RM. Controlled and automatic human information processing: I. Detection, search and attention. Psychol Rev. 1977;84:1-66. (http://dx.doi.org/10.1037/0033-295X.84.1.1)

19. Bentin S, Moscovitch M, Nirhod O. Levels of processing and selective attention effects on encoding in memory. Acta Psychol (Amst). 1998;98:311-41.

20. Oliva A, Torralba A. Modeling the Shape of the Scene: A Holistic Representation of the Spatial Envelope. Int J Comput Vis. 2001;42:145-75.

21. Team TRC. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2014.

22. Steyvers M, Griffiths TL. Rational analysis as a link between human memory and information retrieval. In: Chater N, Oaksford M, editors. The Probabilistic Mind: Prospects from Rational Models of Cognition. Oxford, UK: Oxford University Press; 2008. p. 329-49.

23. Wolfe JM, Võ MLH, Evans KK, Greene MR. Visual search in scenes involves selective and nonselective pathways. Trends Cogn Sci. 2011;15:77-84. (http://dx.doi.org/10.1016/j.tics.2010.12.001)

24. Brown GDA, Neath I, Chater N. A temporal ratio model of memory. Psychol Rev. 2007;114. (http://dx.doi.org/10.1037/0033-295X.114.3.539)

25. Potter MC. Short-term conceptual memory for pictures. J Exp Psychol Hum Learn. 1976;2:509-22. (http://dx.doi.org/10.1037/0278-7393.2.5.509)

26. Potter MC, Staub A, O'Connor DH. Pictorial and conceptual representation of glimpsed pictures. J Exp Psychol Hum Percept Perform. 2004;30:478-89. (http://dx.doi.org/10.1037/0096-1523.30.3.478)

27. Quinlan PT, Dyson BJ. Cognitive Psychology. Harlow, England: Pearson; 2008.

28. Öhman A, Lundqvist D, Esteves F. The face in the crowd revisited: A threat advantage with schematic stimuli. J Pers Soc Psychol. 2001;80:381-96. (http://dx.doi.org/10.1037//0022-3514.80.3.381)

29. Juth P, Lundqvist D, Karlsson A, Öhman A. Looking for foes and friends: Perceptual and emotional factors when finding a face in the crowd. Emotion. 2005;4:379-95. (http://dx.doi.org/10.1037/1528-3542.5.4.379)

30. Jackson MC, Linden DEJ, Raymond JE. Angry expressions strengthen the encoding and maintenance of face identity representations in visual working memory. Cogn Emot. 2014;28:278-97. (http://doi.org/10.1080/02699931.2013.816655)

31. Jackson MC, Wu C-Y, Linden DEJ, Raymond JE. Enhanced visual short- term memory for angry faces. J Exp Psychol Hum Percept Perform. 2014;35:363-74. (http://doi.org/10.1037/a0013895)

32. Pessoa L, Adolphs R. Emotion processing and the amygdala: from a 'low road' to 'many roads' of evaluating biological significance. Nat Rev Neurosci. 2010;11:773-82. (http://dx.doi.org/10.1038/nrn2920)

33. Wolfe JM. Asymmetries in visual search: An introduction. Percept Psychophys. 2001;63:381-9.

34. Hu Y, Hitch GJ, Baddeley AD, Zhang M, Allen RJ. Executive and Perceptual Attention Play Different Roles in Visual Working Memory: Evidence From Suffix and Strategy Effects. J Exp Psychol Hum Percept Perform. 2014;40:1665-78. (http://dx.doi.org/10.1037/a0037163)

35. Quinlan PT. On the use of the term 'attention'. In: Mitchell CJ, Le Pelley ME, editors. Attention and associative learning From brain to behaviour. Oxford, UK: OUP; 2010. p. 217-43.

36. Kensinger EA, Corkin S. Memory enhancement for emotional words: Are emotional words more vividly remembered than neutral words? Mem Cognit. 2003;31:1169–80. (http://dx.doi.org/10.3758/BF03195800)

37. Bakeman R, McArthur D. Picturing repeated measures: Comments on Loftus, Morrison, and others. Behav Res Methods Instrum Comput. 1996;28:584-9. (http://dx.doi.org/10.3758/BF03200546)

Figure Captions

Figure 1. A schematic representation of the sequence of events on a trial. Some of the photo-images are not the originals due to copyright reasons.

Figure 2. Mean levels of accuracy for the conditions of interest in Experiment 1. Error bars represent 95% confidence intervals as defined in [37]. Diff = Target is from a different category to that of the background images. Same = Target and background images are from the same category. NT-B = Non-threatening background images, T-B = Threatening background images, NT-T = Non-threatening target, T-T = Threatening target.

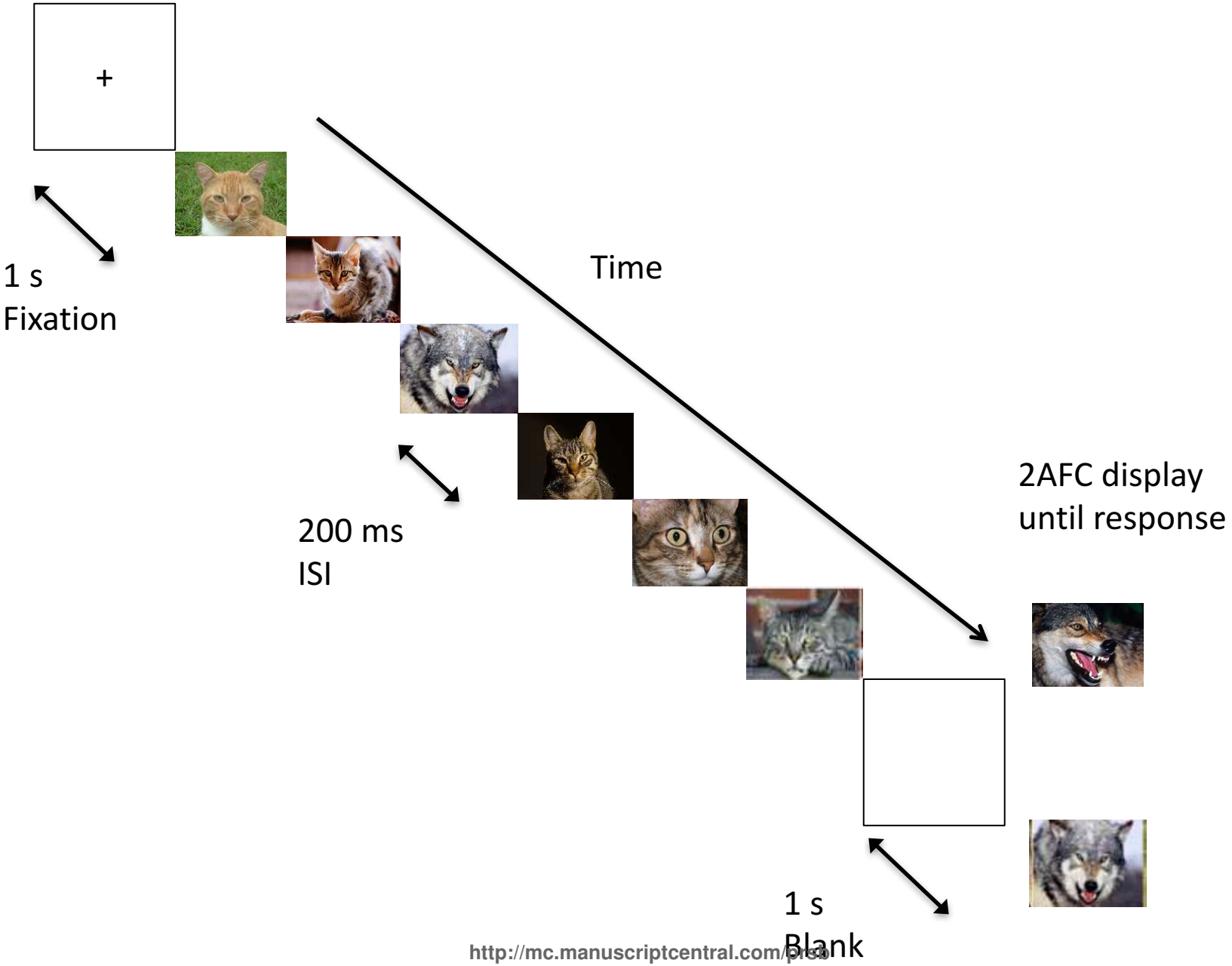
Figure 3. Mean levels of accuracy for the conditions of interest in Experiment 2. Error bars represent 95% confidence intervals as defined in [37]. Diff = Target is from a different category to that of the background images. Same = Target and background images are from the same category. NT-B = Non-threatening background images, T-B = Threatening background images, NT-T = Non-threatening target, T-T = Threatening target.

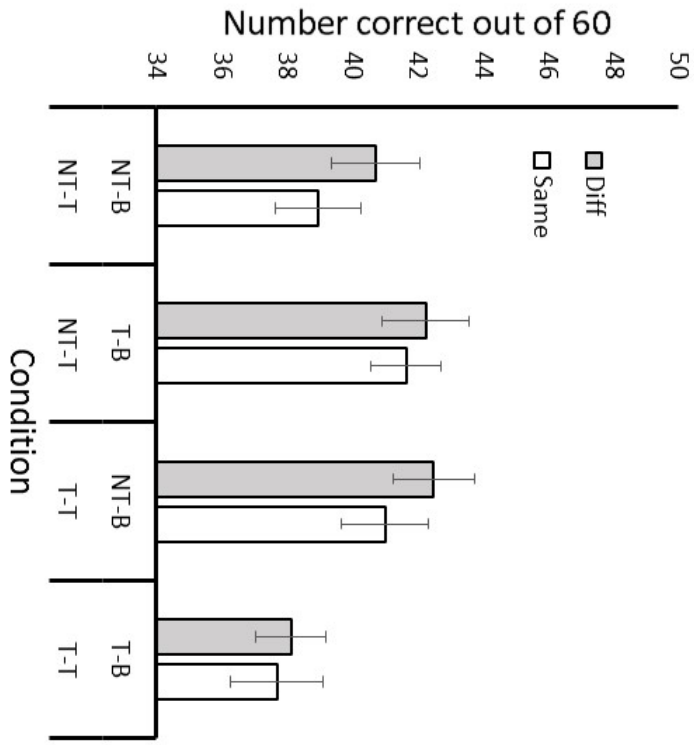
Table Caption

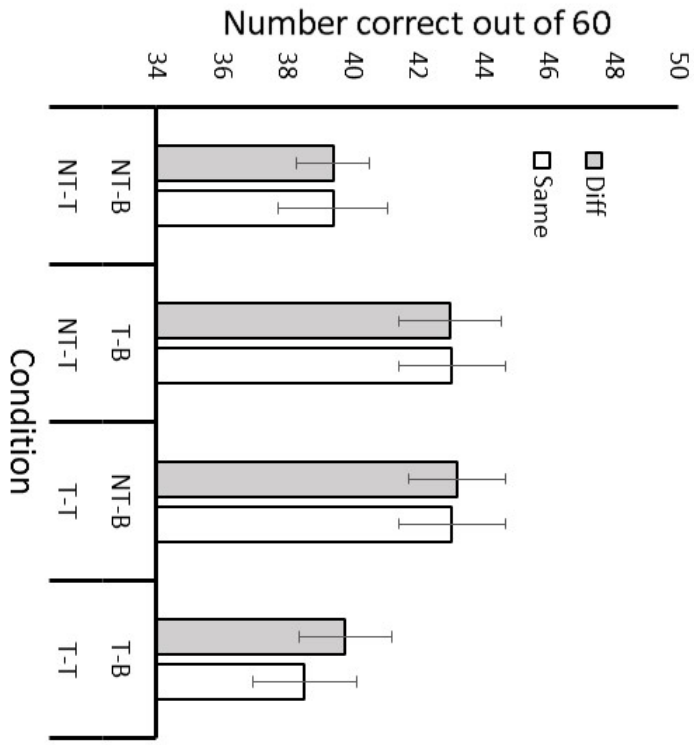
Table 1. Ratings of threat and arousal for the image sets used in Experiment 2. All ratings were taken on a 1-7 Likert scale. SD stands for standard deviation.

Biological Category	Threat rating		Arousal Rating	
	Mean	SD	Mean	SD
		'Threatening' Images		
Dogs	5.77	0.67	4.17	1.73
Cats	5.87	0.78	4.20	1.84
		'Non-threatening' Images		
Dogs	1.55	0.74	3.00	1.41
Cats	1.71	0.76	3.01	1.37

Table 1.


+
1 s
Fixation
Time
200 ms
ISI
2AFC display
until response
1 s
Blank


Number correct out of 60
34
36
38
40
42
44
46
48
50
Diff
Same
Condition
NT-B
NT-T
T-B
NT-T
NT-B
T-T
T-B
T-T


Number correct out of 60
34
36
38
40
42
44
46
48
50
Diff
Same
Condition
NT-B
NT-T
T-B
NT-T
NT-B
T-T
T-B
T-T