

Do semantically consistent scenes sharpen perceived object representations across different cultures?

Ahamed Miflah Hussain Ismail, Justin Samuel Wibowo, Bryan Leong Qi Zheng, and
Alejandro J. Estudillo

Abstract

Perceptual representations of objects are modulated by the semantic consistency between the object and its context, such as the background scene in which it is embedded. In this registered report, we questioned the universality of this semantic consistency effect (SCE) by comparing it across two cultures that are conventionally considered to differ in their tendencies to incorporate context in perception and cognition: White-British and Malaysian-Chinese participants. We measured SCEs by comparing the perceived sharpness of objects embedded in semantically consistent and inconsistent scenes. Malaysian-Chinese participants demonstrated semantic consistency effects akin to those observed in the literature, but White-British participants experienced no effect. This cultural difference was not mediated by individual differences in cognitive styles (holistic versus analytical tendencies in visual processing) and/or self-construals that are commonly believed to bias cognitive styles. We speculate that fundamental differences in predictive coding may account for the observed cultural differences, but more research is required to confirm this hypothesis.

Keywords: *objects, scenes, perception, semantic consistency, culture, thinking styles*

Introduction

Humans are adept at forming perceptual representations of objects and making perceptual decisions about them. It is a remarkably rapid process, where the presence of an object (e.g., animal) in a natural scene can be detected by viewing a scene just for 20 milliseconds (ms; Fize et al., 2005). With an additional 10 ms, people can even assign a basic-level semantic label to the detected object, like classifying the picture of an animal as a dog (Wiebel et al., 2013). These findings highlight a remarkable feat of the human visual system, and empirical work conducted in this area has tremendously improved our understanding of object representations in the brain. However, the ecological validity of many of these studies can be questioned. While humans typically encounter objects in meaningful contexts—within specific times and places, alongside preceding and co-occurring perceptual information—many previous studies have largely overlooked the role of this contextual information in shaping object representations. We are not claiming that these studies present objects separated from their context (e.g., the background scene), but rather they fail to consider how the context in which objects are embedded influence the representation of such objects.

Object-scene consistency effects

The background scene in which an object appears is a form of context, and regularities in scenes play a crucial role in object processing. One such regularity is that objects (e.g., flowers) frequently occur in specific categories of scenes (e.g., gardens), which gives rise to the semantic consistency between an object and a background scene. This semantic consistency is well-known to affect various aspects of object processing. In this regard, Palmer's (1975) initial findings prompted scientists to objectively measure the influence of

scenes on object processing. Palmer presented (and primed viewers with) line drawings of scenes which were followed by line drawings of objects. Participants classified objects more accurately when they were primed with semantically consistent scenes (e.g., a loaf of bread preceded by a kitchen) than inconsistent scenes (e.g., a mailbox preceded by a kitchen). Following this, several studies have extensively quantified the influence of context and have generally demonstrated an advantage for semantic consistency. When viewers are tasked with classifying objects, semantic consistency often facilitates classification (Biederman et al., 1982; Davenport & Potter, 2004). Similarly, when the task is to locate an object of a specific category, viewers are also quicker if objects are placed in their most naturally likely locations (i.e., a semantic consistency in spatial locations within scenes) than when they are not (Malcolm & Henderson, 2010). Moreover, the deployment of overt attention (fixations) is also guided by the same consistency (Castelhano & Henderson, 2007). In brief, when tasks involve processing categorical attributes of objects, semantic consistency appears to result in improved performance.

Conversely, when tasks do not involve explicit classification and rely heavily on perceptual analyses of objects, the effects of semantic consistency are highly task-and stimulus-dependent. On the one hand, semantic consistency can incur a cost or a benefit depending on the task at hand. For instance, when participants have to detect if an object in a scene has changed (e.g., from being present in the scene to being absent) by pointing to the spatial location of the change, semantic consistency incurs a cost (LaPointe et al., 2013). There is also a cost when perceptually matching an isolated object to one that was presented previously in a scene (Ortiz-Tudela et al., 2023; Spaak et al., 2022). However, instead of locating the change or performing a perceptual matching task, if participants have to identify which object changed (e.g., whether it was a glass that disappeared), semantic consistency

facilitates (LaPointe et al, 2013). Another example of task-dependency is that, in the absence of an explicit task to search for an object, effects of semantic consistency on gaze deployment have produced mixed results. While some find that gaze is attracted by semantically consistent objects, others find attraction to inconsistent objects, but it is somewhat clear that viewers fixate longer on semantically inconsistent than consistent objects (see Cornelissen & Melissa (2017) for experimental findings and Wu et al. (2014) for a brief overview of the relevant literature). On the other hand, semantic consistency can enhance or suppress perceptual representations of objects depending on the reliability of the object's visual signal. For example, both behavioural and neuroimaging evidence demonstrate that semantic consistency increases the perceived sharpness of objects when the object's signal is unreliable (i.e., blurry), but it decreases the perceived sharpness when the signal is reliable (i.e., not blurry) (Brandman & Peelen, 2017; Rossel et al., 2022; Rossel et al., 2023).

Semantic consistency effects, regardless of whether it is beneficial or costly, are often explained by invoking predictive coding theories which provide a compelling account (Rao and Ballard, 1999; Peelen et al., 2024). Here, the gist extracted from a scene activates expectations (i.e., prior knowledge) we hold about such scenes, and these expectations are formed from exploiting regularities in the environment through perceptual experiences (Bar, 2004; Peelen et al., 2024). The timescale of such experiences sometimes extends beyond the life experiences of an individual, as we see that some expectations that match regularities of the environment are hardwired in the human brain (Furmanski & Engel, 2000; Girshick et al., 2011). Regardless of their origin, the activation of expectations (e.g., by the gist of a presented scene) results in predictions about what objects should be present in the scene. The disparity between the predicted signal (e.g., expected object) and the actual sensory signal (e.g., the presented object) is conventionally referred to as the *prediction error*, whereby

semantic consistency between a presented object and its background scene would lead to smaller prediction errors.

As highlighted above, the consequences of smaller prediction errors (i.e., high semantic consistency) are highly task dependent, but predictive coding can explain consequences that are both advantageous as well as costly. For instance, when the prediction error is low for objects that meet the predictions, it may allow the visual system to easily integrate—and therefore, process—those objects (Bar, 2004; Peelen et al., 2024). Integration would be difficult for objects that do not fit the predictions which result in greater prediction error. This could explain why semantic consistency facilitates classification (Biederman et al., 1982; Davenport & Potter, 2004) and sharpens perceptual representations when the signal is unreliable (Rossel et al., 2022, 2023). Alternatively, we may allocate greater processing capacity and prioritize objects that result in greater prediction errors than those that result in smaller errors, presumably an attempt to resolve the prediction error in the former case (Peelen et al., 2024). This could explain why in some cases, perceptual representations are relatively salient for objects that are semantically inconsistent with scenes (Spaak et al., 2022; Rossel et al., 2023), as well as why gaze is sometimes attracted by inconsistent objects, and why people spend more time viewing inconsistent objects (see Wu et al., 2014).

While semantic consistency effects take many facets, we were curious whether such effects are uniform across individuals, and more importantly, whether predictive coding based on prior expectations would operate uniformly. When considering potential sources of individual differences, we believe culture is an appropriate candidate to examine, an individual characteristic that semantic consistency studies do not really mention when reporting their

participant sample. One possibility is that individuals from different cultures hold different expectations based on their distinctiveness in long-term visual experience, and these differences can contribute to differential perception (e.g., Tse & Cavanagh, 2000). The other possibility that we were interested in is that individuals from different cultures may share similar expectations but may differ in the extent to which those expectations are employed to guide perception.

Culture as a modulator

Several studies have shown cross-cultural differences about the modulatory effect of context in visual processing. For example, individuals from Eastern cultures (e.g., Japanese) are often influenced by contextual information to a greater extent than those from Western cultures (e.g., Americans), when judging the perceived length or orientation of a line (target) that appears within a square-shaped frame surrounding it (context) (Ji et al., 2000; Kitayama et al., 2003; but see Willey and Liu, 2022 for contradicting evidence). Such cultural differences are not limited to the contextual effect of low-level attributes (i.e., a square) on the perception of another low-level attribute (i.e., a target line). For instance, some studies have demonstrated that classification of expressions conveyed by a target face is altered by its spatial context (e.g., the background scene and emotions expressed by individuals surrounding the target) in Easterners more than Westerners (Masuda et al., 2008; Ito et al., 2013). This is an example of cultural differences in the effect of a high-level semantic attribute in the context (e.g., emotion of surrounding individuals) on the interpretation of the same attribute in the target. What we do not know is whether cultures differ in the influence of high-level attributes of context (e.g., semantic meaning of a scene) on the perception of

145 lower-level attributes of a target object (e.g., the perceived sharpness of objects), specifically
146 when predictive coding operates through top-down effects.

147

148 These cultural differences in integrating the target with the context obviously cannot arise
149 from a nominal distinction between cultures. Accordingly, one of the popularly scrutinized
150 explanations of these cultural differences originate from the relationship between social
151 structures archetypal of cultures and cognition. Traditionally, Western cultures are considered
152 individualistic (e.g., American and many European cultures) and Eastern cultures are
153 considered collectivistic (e.g., Japanese, Chinese and other East-Asian cultures), and based on
154 national-level indicators, this segregation still remains to be true (see Pelham et al., 2022).
155 Individualism places greater emphasis on the individual as an entity separated from the rest of
156 the society, while collectivism emphasises on the individual in relation to the society.
157 Individuals raised in different social structures differ in their self-construals, with
158 individualism promoting independent self-construals and collectivism promoting inter-
159 dependent self-construals (Markus and Kitayama, 1991a; Matsumoto, 1999).

160

161 The idea that social structures and resulting self-construals may influence cognition is not
162 new (see Markus & Kitayama, 1991b) and has also been proposed to explain cultural
163 differences in susceptibility to context. Differences in social structures and self-construals
164 foster different types of thinking, and therefore modulates perception and cognition too
165 (Hong et al., 2000; Ji & Yap, 2016; Markus & Kitayama, 1991a, 1991b; Nisbett & Miyamoto,
166 2005;). Westerners like Americans who come from a predominantly individualistic culture
167 with independent self-construals have a tendency to process visual information in an
168 analytical manner (Nisbett & Miyamoto, 2005). This tendency would make them process a

visual target with minimal influence from its context. Easterners like Japanese who come from predominantly collectivistic cultures with inter-dependent self-construals have a tendency to process visual information in a holistic manner (Nisbett and Miyamoto, 2005). This would make them process a visual target in relation to its context. This logic is what makes some researchers to attribute East-West differences in contextual effects to culture-specific cognitive styles, namely holistic and analytical thinking (e.g., Kitayama et al., 2003; Masuda et al., 2008). Surprisingly, the culture specificity in this attribution is made by *assuming* that cognitive styles originate from socio-cultural constructs like self-construals, which is controversial. Sometimes, even when Westerners and Easterners exhibit analytical and holistic cognitive styles, respectively, individuals from these two cultures do not necessarily differ in their self-construals (e.g., McKone et al., 2010). Therefore, measuring social constructs is crucial, and in fact should be mandated, to attribute culture-specific cognitive styles to socio-cultural constructs.

Proposed study

Since cognitive styles are believed to be different between cultures, we reasoned that individuals coming from an Eastern culture (e.g., Malaysian-Chinese), likely with a holistic cognitive style and interdependent self-construals, would rely more on the scene context when perceiving an object in it. In contrast, individuals coming from a Western culture (e.g., White-British living in the United Kingdom), likely with an analytical cognitive style and independent self-construals, would rely less on the scene context. More specifically, we expected that Easterners would use predictions arising from (the activation of prior knowledge related to) a scene's gist to guide perception of objects to a greater extent than Westerners, resulting in stronger semantic consistency effects in Easterners. Our expectation

is also backed by a comparable principle suggested for individuals with autism spectrum disorder, whose bias to perceive visual information analytically may reflect impaired top-down guidance from expectations (Happé & Frith, 2006).

We aimed to test this possibility by measuring semantic consistency effects on perceptual representations (i.e., perceived sharpness), in a sample of White-British living in the United Kingdom (UK) and in a sample of Malaysian-Chinese living in Malaysia, after ensuring that they have lived in their respective countries since birth. We found semantic consistency effects on perceived sharpness of objects (Rossel et al., 2022, 2023) particularly intriguing because they inform how cognition—in the form of predictions—penetrates perception, and how perception is therefore constrained by cognitive functions. To this end, we used a paradigm initially designed by Lupyan (2017) and later adapted by Rossel et al. (2022; 2023) to measure semantic consistency effects on perceived sharpness.

Rossel et al. (2023) presented two images side-by-side to participants. Each image had the same object (e.g., a boat) that was filtered to have an identical level of blur objectively, but the background scene was different. One had a background that was semantically consistent with the object (e.g., a harbour scene), while the other had a background semantically inconsistent with the object. Participants were asked to adjust the blur level of one of the two images until the objects in both images appeared equally blurry (or sharp) to them. Despite objects in both scenes having the same objective level of blur, participants perceived the object in the semantically consistent scene as relatively less blurry (i.e., sharper), because they increased its blur level to match the object in the inconsistent scene. It is worth noting here that this effect only occurred when the objects' blur level was high (i.e., the signal was

unreliable) and not when it was low (i.e., the signal is reliable). Accordingly, we followed Rossel et al.'s (2023) method, but selectively presented objects only at a high level of blur. Based on participants' adjustments of blur, we calculated the difference in adjusted blur levels between the object in the consistent scene and that in the inconsistent scene. This was our measure of the semantic consistency effect (SCE).

First, we compared mean SCEs between the nominally distinct cultures in our participants, at the group-level. Second, we examined whether the influence of culture on SCEs is mediated by our participants' cognitive styles and self-construals. Cognitive styles were measured using a computerised equivalent of Kitayama et al.'s (2003) frame-line task. It is a behavioural task that measures the extent to which context (a square-shaped frame) interferes with judging the perceived length of a line inside the frame (Kitayama et al., 2003). The greater the interference, the more holistic an individual is thought to be. We expected that greater tendencies to process information holistically will predict larger SCEs.

Self-construals were initially measured using the self-construal scale originally proposed by Singelis (1994). This popularly used scale conceptualizes self-construals as two orthogonal and unitary dimensions, namely interdependence and independence. However, some critics of this conceptualisation argue that it is oversimplified, and that it does not comprehensively capture the nuances of cross-cultural and inter-individual variability in self-construals (e.g., Hardin et al., 2004; Vignoles et al., 2016). Addressing these issues, more recent research has shown that self-construals are multifaceted, and that it is more informative to model self-construals as multiple dimensions that are bipolar in nature (Vignoles et al., 2016). While one end of each dimension would characterise independence, the other end would characterise

interdependence. Accordingly, we also measured self-construals using some relevant dimensions extracted from Vignoles et al.'s (2016) multi-dimensional model (and its later developments) of self-construals. For all dimensions measured, we expected that degree of polarity towards the interdependent pole of a dimension will predict stronger SCEs.

Methods

Participants

An a-priori estimation of sample size was carried out using G*Power 3.9.1.7 (Faul et al., 2007). We intended to perform two types of analyses to assess cultural differences. One was an independent-samples *t*-test comparing contextual effects between the two distinct groups of participants (Asians and Caucasians), for which G*Power estimated a sample size of 128 participants (64 per group), to obtain a medium effect size of $d' = 0.5$, with 80% statistical power and a significance criterion of $\alpha = 0.05$. The second analysis intended to assess whether the effect of culture on semantic consistency effects is mediated by individual differences in cognitive styles and/or self-construals. To test these, we ran a series of mediation analyses, following the regression-based method of Baron and Kenny (1986). Each analysis had one independent variable (i.e., participants' group), one dependent variable (i.e., SCE) and one mediator (i.e., cognitive styles and each of the self-construal dimensions). For this, we identified that a sample size of 118 participants would be sufficient to obtain a medium effect size for both direct and indirect effects in the mediation model. This was identified from Fritz and MacKinnon's (2007; Table 3) simulation-based recommendations of sample size to obtain medium-sized effects for both direct and indirect effects in the mediation model.

To ensure sufficient statistical power for both analyses, we recruited 130 participants with normal or corrected-to-normal vision, and with no known diagnosis of Amblyopia. Sixty-five of those were White-British participants (age: mean (M) = 21, standard deviation (SD) = 3; Sex: 15 males) recruited at Bournemouth University, United Kingdom (UK). They had lived in the UK since birth. The remaining 65 participants were Malaysian-Chinese participants (age: M = 21, SD = 3; Sex: 27 males), recruited at University of Nottingham Malaysia, and they had lived in Malaysia since birth. White-British participants were compensated with course-credits. Malaysian-Chinese participants were compensated with course-credits and/or 10 Malaysian Ringgits.

Materials

Object perception task. This task was designed to measure perceived sharpness (or blur) of objects that are presented in consistent and inconsistent scenes. To identify appropriate stimuli for this task, we extracted some images under the creative commons license from Pixabay and Google. First, 40 images of objects were extracted. Second, for each object (e.g., jeep), we extracted one scene that appears to have high semantic consistency (e.g., road), and another one that appears to have low semantic consistency (e.g., kitchen). This resulted in a total of 80 scene images extracted. These initial extractions were based on the authors' subjective impressions from visual inspection.

Each scene (e.g., scene of a river) extracted was cropped to a square, ensuring that the cropped-out regions of the scene do not compromise the semantic content of it. Scenes were then resized to have a height of 500 pixels and converted to greyscale by independently weighting and summing the red, green and blue channels of the image according to the CIE

procedure ($0.299 \times R + 0.587 \times G + 0.114 \times B$). In greyscale images, pixel intensities ranged between 0 (black) and 255 (white), with 256 potential grey levels.

Each object (e.g., a boat) extracted was first cropped such that the object's outline is abutting the image borders. Next, each object was resized so that the maximum of its two dimensions (height and width) do not exceed 85 pixels, while maintaining the same aspect ratio as the originally cropped image. Objects were then converted to greyscale using the same CIE procedure as above, and pixels outside the object's outline were replaced with transparent pixels. Finally, objects were centred on a uniformly grey (pixel intensity = 128) square-shaped image with a height of 85 pixels. Matlab 2024a (Mathworks) was used for resizing images and converting them to greyscale, but for all other image processing steps mentioned above, we used Adobe Photoshop CS6.

The two groups of participants in this study came from different cultures. Their familiarity with the objects we present may have differed, and this could extend to the perceived semantic consistency between objects and scenes as well (see Weigl et al., 2024). To minimise potential influences of these factors on participants' adjustments of blur levels in the object perception task (see below), we conducted a process of stimulus validation online, using Testable (<https://testable.org>; Rezlescu et al., 2020). This validation involved 20 White-British (age: $M = 21$, $SD = 3$; sex: 5 males) and 20 Malaysian-Chinese (age: $M = 22$, $SD = 1$; sex: 8 males) participants, who did not participate in other tasks of the main study. Each participant viewed 80 experimental trials, in their own computers. At the beginning of each trial, a unique object was presented alone, at the left-hand side of the screen. Participants indicated on a 5-point Likert scale, how familiar they are with the object (1 – not at all, and 5

– absolutely). Once they completed it, a scene appeared on the right-hand side of the screen (interval 1). Participants indicated the semantic consistency between the object and the scene, on a 5-point Likert scale (1 – not at all, 5 – absolutely). The scene was then replaced with another scene (interval 2), and participants also rated that object-scene combination for semantic consistency. One interval presented a semantically consistent scene, and the other a semantically inconsistent scene; consistency in both cases were pre-determined by the authors. The order of presenting those scenes were randomised across intervals.

Participants' responses were first used to exclude objects with significant differences in familiarity between groups, determined via two-tailed Mann-Whitney U-tests. All 40 object images we validated produced non-significant differences in familiarity ratings between the two groups, and mostly had ceiling levels of familiarity (see supplemental materials, Section S1, Table S1). Next, the semantic consistency of object-scene combinations (i.e., two scenes per object) were evaluated. For any given object, object-scene combinations were identified as suitable for the object perception task if two criteria were met (see supplemental materials, Section S1, Tables S2-S6): (1) consistent combinations have significantly higher mean semantic consistency ratings than inconsistent combinations in both participant groups (two-tailed Wilcoxon signed-rank tests), and (2) mean consistency ratings do not differ between groups for either combination type (two-tailed Mann-Whitney U-tests). Both criteria were met for 39 out of 40 objects validated. Among those, we selected 20 objects with high differences in semantic consistency ratings between its consistent and inconsistent combinations, and those that looked conspicuous to the authors when the least amount of blur is applied (see supplemental materials for the final list of objects). These object-scene combinations were retained for the object perception task. We used non-parametric comparisons to assess all validation criteria as the data were not normally distributed.

338

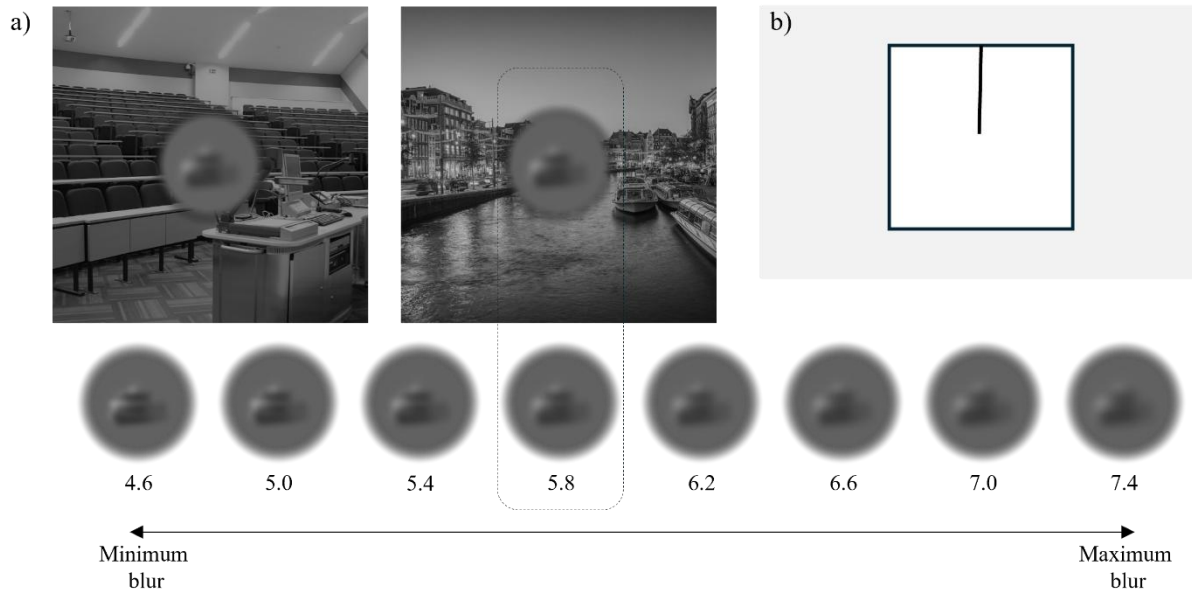
339 Following a similar (but not an identical) method as Rossel et al. (2022; 2023), we created
340 filtered duplicates of each object image retained from the validation. This involved
341 convolving a 2-dimensional Gaussian kernel with the object images, using the *imgaussfilt*
342 function in Matlab. In the experimental task, we presented two different versions of each
343 object, one labelled as the “target”, and the other labelled as the “sample”. To create filtered
344 targets, we always used a standard deviation (SD) of 5.8 for the Gaussian kernel. The SD here
345 is different to what was used by Rossel et al. (2023) to demonstrate an advantage of semantic
346 consistency for highly blurred objects. This is because the SD has been adjusted to suit the
347 size of our objects, as our objects were smaller than what they used. For each filtered sample
348 that was presented to participants, the initial SD was selected randomly from a set of values
349 ranging from 4.6 to 7.4 with a spacing of 0.4. As described in the procedure section below,
350 the SD of the sample can change based on the participant’s response. Both filtered targets and
351 samples were only superimposed on scenes after inserting them in the centre of a uniformly
352 grey square-shaped background with a height of 177 pixels. The resulting image was
353 multiplied with a circularly symmetric tapered-cosine window of 60 pixels radius, to ensure
354 that objects are visible through a soft-edged circular aperture. With these stimuli, we created
355 two unique combinations of objects superimposed on scenes, characterised by the version of
356 the object (target versus sample) and the semantic congruency of the object to the scene
357 (congruent versus incongruent): sample-consistent (SC) paired with target-inconsistent (TI),
358 and sample-inconsistent (SI) paired with target-consistent (TC). An example of one of these
359 object-scene pairings is provided in Figure 1a.

360

Moreover, our main measure with regards to these visual stimuli was how subjective perception of blur in objects differed depending on whether they were placed in semantically consistent or inconsistent scenes. Accordingly, to ensure that there are minimal objective differences in low-level visibility between various scenes and between various objects, we used Matlab and equated all the scenes and all the objects (separately) to have the same low-level visibility. To do that, we first calculated each scene's (or each object's) visibility using a method proposed by Ismail et al. (2019). In this method, visibility is calculated as the dot product between an image's power spectrum and a window of visibility that models contrast sensitivity as a function of spatial frequencies. The visibility index here gives a measure of the effective contrast of an image, after taking into account the human visual system's non-uniformities in contrast sensitivity to various spatial frequencies. After calculating the visibility of all images in a set (scenes or objects), we found the average visibility across all (unfiltered) images within a set, and defined half of this value as the "reference visibility". The amplitude spectra of all the images in a set (except the image that holds the reference visibility) were uniformly adjusted to reach the visibility of the reference image, thereby equating all images for low-level visibility. Amplitude spectra of all filtered target objects were also adjusted to reach the reference visibility of unfiltered objects, and variants of the sample were created by simply manipulating the level of Gaussian blurring of these adjusted targets.

Figure 1

Examples of stimuli used in a) the object perception task, and b) the frame-line task.



Notes: In panel “a”, the target object of a passenger ship is superimposed on a semantically inconsistent scene (lecture room; left-hand side) and the sample object of the same passenger ship is superimposed on a semantically consistent scene (river; right-hand side). In this example, the target and sample objects have both been set to have an identical level of blur ($SD = 5.8$), and the sequence of objects in the lower end of the panel illustrates what the sample object would look like if the participant was to adjust its blur level with keypresses (from a minimum achievable blur of $SD = 4.6$ and a maximum achievable blur of $SD = 7.4$, in steps of $SD = 0.4$).

Frame line task. This task, originally developed by Kitayama et al. (2009), was used as a behavioural measure of cognitive styles. Specifically, this task evaluated participants’ abilities to integrate (in the “relative” version) or ignore (in the “absolute” version) contextual information. Details about both versions are provided in the Procedure section, but the basic

stimulus for both tasks is a square frame and a vertical line that originates from the top edge of the frame and extend into the frame's interior space (see Figure 1b).

Questionnaires measuring self-construals. Two self-report questionnaires were used to measure self-construals in our participants. The self-construal scale (Singelis, 1994) is a 24-item questionnaire that measure self-construals along the two orthogonal and separate dimensions, interdependence and independence (Supplemental Materials Section S2). The scale has 12 items per dimension, and for each item, participants' responses were obtained on a 7-point Likert scale (1 = strongly disagree and 7 = strongly agree). The second multi-dimensional scale included items taken from a selected number of dimensions in Vignoles et al.'s (2016) 7-dimensional model of self-construals (Supplemental materials Section S3). We first included dimensions for which Vignoles et al. (2016) demonstrated statistically significant differences between Westerners (which included people from the United Kingdom) and Southern/Eastern Asians (which included Malaysians). These dimensions were *difference vs similarity* (6 items measuring how self is distinguished from others), *self-direction vs receptiveness to influence* (6 items measuring how decisions as made with reference to self and others), *consistency vs variability* (6 items measuring how an individual operates in difference contexts), *self-expression vs harmony* (6 items measuring how individuals communicate with others). For these dimensions, the reliability across nations (United Kingdom and Malaysia) ranges between acceptable (*Cronbach's* $\alpha = 0.68$) to good ($\alpha = 0.85$) (see Krys et al., 2021). We also included an additional dimension from a latest development of Vignoles et al.'s model, which is *de-contextualised vs contextualised self* (6 items measuring the extent to which an individual is embedded in their context), as we believed this dimension is closely related to our measure of perception in context. People from United Kingdom and those from Eastern nations like China can be contrasted in this

dimension too (Yang, 2018; Study 3), but we could not find any contrasts with Malaysians for the same dimension. Nonetheless, this dimension has shown a reliability of $\alpha = 0.74$ when tested against Malaysians, and $\alpha = 0.82$ for individuals in the United Kingdom (Krys et al., 2021). For all items used in the multi-dimensional scale, we used a 7-point Likert scale to obtain responses (1 = completely disagree and 7 = completely agree).

Apparatus

All the questionnaires were delivered online using the survey platform Qualtrics (Provo, UT). To present the behavioural (object-perception and frame-line) tasks to the Malaysian-Chinese participants, we used a gamma-linearized Samsung 27-inch W-LED screen with a spatial resolution of 1920 (width) $\times$ 1080 (height) pixels, and a refresh rate of 60 Hz. To present the same tasks to the White-British participants, we used a gamma-linearized BenQ 24-inch LCD screen with a spatial resolution of 1920 (width) $\times$ 1080 (height) pixels, and a refresh rate of 60 Hz. To maintain identical visible sizes of stimuli (in degrees of visual angle), Malaysian-Chinese were seated 63 centimetres (cm) from the screen, and White-British participants were seated 55 cm from the screen, both using the support of a chin rest. Participants completed the two behavioural tasks in a dark room. For the frame-line task, stimuli on the screen were viewed through a circular aperture placed on the screen to occlude the vertical and horizontal borders of the screen, while displaying as much of the screen as possible. For both the object perception and frame-line tasks, stimulus presentation and response collection were controlled via Matlab, using the Psychophysics toolbox v3.0.19 (Brainard, 1997; Pelli, 1997)

Procedure

Experimental procedures were approved by the Faculty of Science and Engineering Faculty Research Integrity and Research Ethics Committee, at University of Nottingham Malaysia (approval code: AMHI090425), and the Ethics Committee of Bournemouth University (approval code: 61979). All participants began participation by providing informed consent, followed by completing the two self-construal questionnaires on Qualtrics. Next, after a two-to three-minutes break, participants completed the object perception task. In this task, participants were presented with 80 experimental trials. Twenty of those trials presented the SC_TI pair, with the sample stimulus containing an object semantically consistent with its background scene and the target stimulus containing an object semantically inconsistent with the background. Another 20 trials presented the SI_TC pair, where the sample's object was semantically inconsistent with its background and the target's object was semantically consistent with its background. The third set of 20 trials presented an SC_TC pair, while the fourth set presented an SI_TI pair. Within each set of 20 trials, a unique combination of object and scene were only used once. Trials from the four unique pairings (SC_TC, SI_TI, SC_TI, and SI_TC) were fully randomised and intermixed within a single sequence of the 80 experimental trials. Trials with SC_TI and SI_TC pairings allowed us to assess semantic consistency effects, while SC_TC and SI_TI pairings allowed us to estimate the presence of response biases, if any.

Each of the 80 experimental trials commenced with a white fixation cross shown in the centre of a mid-grey screen for 1000 milliseconds (ms). This was followed by the sample-target combination presented on the screen, until a confirmation response (see below) was received from the participants. The sample and target were presented on either side of the vertical

meridian of a mid-grey screen, but along the same horizontal meridian. Above each image, a label indicated to the participant which stimulus is the sample and which one is the target. The sample (and the target) could appear on the left- or right-hand side of the screen randomly. Participants were instructed to adjust the blur level of the sample's object to match the blur level of the target's object, by pressing the "up" or "down" arrow keys in the keyboard, to decrease (make it sharper) or increase (make it blurrier) the blur level, respectively. Each keypress adjusted the blur level at an SD of ± 0.4 . The adjustments were not allowed to go outside the permitted range of adjustments (i.e., below a blur level of SD = 4.6 or above SD = 7.4). Once participants were satisfied with their adjustment, they pressed the space bar as a confirmation response to terminate the trial. Participants were required to make a minimum of 2 adjustments before proceeding to the next trial (e.g., increasing and then decreasing the blur level, or vice versa, even if they perceive equal blur levels in the sample and target from the very beginning of the trial).

487

Lastly, after another 2-3 minutes break, participants completed the frame-line task. This task consisted of two versions, which participants completed in a counter-balanced order. Regardless of the version, each trial started with the presentation of a white fixation cross in the centre of a mid-grey screen for 500 ms. Next, a single frame-line stimulus (figure 1b) was presented for 1000 ms. This ("reference" stimulus) was followed by the presentation of a noise mask for 1000 ms. The mask was a uniform distribution of grey values between black and white, freshly generated for each trial. The purpose of the mask was to interfere with any afterimages the frame-line stimulus would leave, as afterimages can serve as a reference for the task that the participants are going to do. After the mask, participants were shown a second frame-line stimulus (the "test" stimulus), which remained on the screen until a response was confirmed. In the absolute version of the task, participants adjusted the line of

the test to have the same exact absolute line length as the reference. In the relative version of the task, participants also adjusted the length of the line, but to ensure that the proportion of the line to the frame's height in the test was identical to the same proportion in the reference. The length was adjusted by using the up, down, left and right arrows. Pressing up or right arrows increased the length by ± 10 or 1 pixel(s), respectively. Pressing the down or left arrows decreased the length by ± 10 or 1 pixel(s), respectively. Once participants were satisfied with the adjustment, they pressed the space bar to continue to the next trial. Participants were required to make a minimum of 2 adjustments before proceeding to the next trial (e.g., increasing and then decreasing the line length, or vice versa, even if they perceive equal line lengths/proportions between the test and the reference stimuli from the very beginning of the trial).

As shown in Table 1, we used five unique combinations of reference and test stimuli, and the same combinations were used in both versions of the task. To obtain these combinations, we increased the line lengths and frame heights used by Kitayama et al. (2003) by two folds and converted those values from millimetres to pixels. Within a given version of the task, each combination was presented 5 times, which led to a total of 25 trials per version (5 combinations $\times$ 5 repeats). The 25 trials were presented in a fully randomised sequence. As can be seen in table 1, the starting line length of the test stimulus was either 10 pixels (in half the trials) or 10 pixels less than the test frame's height (in half the trials).

Table 1

Combinations of reference and test stimuli used in the frame-line task

Reference stimulus		Test stimulus	
Height of the frame	Length of the line	Height of the frame	(starting) length of the line
178	124	358	10 or 348
204	58	306	10 or 296
254	106	254	10 or 244
306	174	204	10 or 194
358	62	178	10 or 168

Note: All values are in pixels.

Data Analyses (based on stage 1 registration)

Cross-cultural comparison of semantic consistency effects. Our primary aim was to compare semantic consistency effects between the two nominally distinct cultures. Prior to that, we examined whether there were any response biases in assigning blur levels to objects in semantically consistent or inconsistent scenes. To do these, we first obtained the difference between the blur level assigned to the sample and the blur level of the target, for each participant separately. This difference was obtained across all trials within pairings of SC_TC (consistent sample_consistent target), SI_TI (inconsistent sample_inconsistent target), SC_TI (consistent sample_inconsistent target) and SI_TC (inconsistent sample with consistent target), separately. To calculate, we subtracted the blur level of the target from the blur level assigned to the sample. At this point, we intended to minimise the inclusion of data coming

from participants passively responding to stimuli without active engagement and effort. Accordingly, if a response in any given trial (starting from the third trial onward) marked the third consecutive instance where a participant had utilised extreme bounds of the adjustment range (SD of 4.6 or 7.4), those trials were excluded from the subtraction procedure. However, an exception was made to this criterion if the trial involved the same exact stimulus being presented for the third consecutive time. If this exclusion criteria resulted in excluding data from more than 20% of trials in any experimental condition, the participant's data was meant to be removed from further analyses, and a replacement participant was meant to be recruited to meet the intended sample size.

If there were no response biases, the calculated difference scores should not deviate from zero for SC_TC and SI_TI pairings. However, participants from one or both cultures may have had response biases, such as a tendency to overestimate blur when objects and scenes are inconsistent, and/or underestimate when they are consistent. We confirmed the presence of such biases in each culture separately, by conducting two-tailed one-sample *t*-tests on the difference scores derived from SC_TC and SI_TI pairings. Response biases aside, we expected the calculated difference score to be positive for SC_TI pairs and negative for SI_TC pairs, as objects in semantically consistent scenes are perceived to be sharper. Based on these two difference scores, we calculated an overall semantic consistency effect (SCE) for each participant, using equation 1 below. We then ran a one-tailed independent-samples *t*-test to assess whether the mean SCE score in White-British participants is significantly lower than that of Malaysian-Chinese.

$$SCE = \frac{Difference_{SC_TI} + (-1 \times Difference_{SI_TC})}{2}$$

(Eqn. 1)

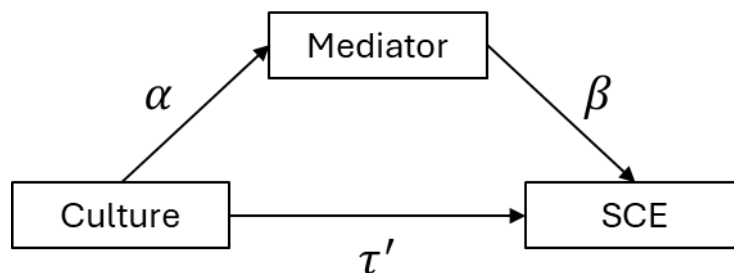
564

565 **Mediations from cognitive styles and self-construals.** Our secondary aim of this study was
566 to assess whether the influence of culture on semantic consistency effects were mediated by
567 cognitive styles (holistic vs analytical thinking) and/or self-construals. For this assessment,
568 we ran a series of mediation analyses following Baron and Kenny's (1986) regression-based
569 method. As shown in figure 2, the mediation model in each analysis had nominal culture
570 (British Caucasians and Chinese Malaysians) as the independent variable, and SCEs as the
571 dependent variable. The mediator was unique to each analysis. To confirm whether a
572 complete mediation is present, we followed four criteria: 1) differences in culture has a direct
573 effect on SCEs (path τ' in figure 2), 2) differences in culture significantly account for
574 variations in the mediator (α path in figure 3), variations in the mediator significantly account
575 for variations in SCEs after controlling for culture (β path in figure 2), and 4) there is no
576 direct effect of culture on SCEs after controlling for effects in paths α and β . If the 4th
577 criterion was not met, but the direct effect under criteria 4 was smaller than the direct effect
578 under criteria 1, we considered that a partial mediation.

579

580 **Figure 2**

581 *Path diagram for the mediation model.*



582

583

First, for cognitive styles, we obtained measures of tendencies to process information holistically and analytically, from the frame-line task. For each participant, we obtained the mean error (in pixels) in reproducing the absolute line lengths for the absolute version of the frame-line task (“absolute error”). Errors here are due to a tendency to incorporate the context frame when judging absolute lengths (Kitayama et al., 2003), and therefore, it represents a bias towards holistic processing. For each participant, we also calculated the error in reproducing the proportion in the relative task (“relative error”). Errors occur here because of a tendency to ignore to context when estimating the proportion (Kitayama et al. 2003), and therefore, it represents a bias to process information analytically. For both tasks, when obtaining mean errors, we utilised the same data exclusion criteria that we used for the object perception test. If a response in any given trial (starting from the third trial onward) marked the third consecutive instance where a participant had utilised extreme bounds of the adjustment range (10 pixels or 10 pixels less than test frame’s height), that was meant to be excluded from estimating mean errors. However, an exception was made to this criterion if the trial involves the same exact stimuli being presented for the third consecutive time. If this exclusion criteria results in excluding data from more than 20% of trials in any of the two tasks, the participant’s data was meant to be removed from further analyses, and a replacement participant was meant to be recruited to meet the intended sample size. Once we have assessed this, we ran a mediation analysis for errors from each of the two tasks.

Second, we obtained individual measures of self-construals. For the Singelis’ (1994) self-construal scale, we obtained the difference in scores between the interdependence and independence dimensions. We labelled this difference score as Singelis’ Interdependent Score (SIS), with positive SIS scores indicating greater interdependence (Kitayama et al., 2009; Park and Kitayama, 2014). The SIS scores were included as a mediator in the next mediation

analysis. Lastly, for the dimensions extracted from the multi-component model of self-construals, we obtained a score for each of the five dimensions (e.g., *difference vs similarity*). Positive scores for each dimension represent greater independence. To test the role of each of the five dimensions, we ran a mediation analysis with the respective dimension scores.

Results

Semantic consistency effects. First, when calculating semantic consistency effects (SCEs) for individual participants, we identified trials to exclude based on recurrent use of extreme bounds in adjusting blur levels as outlined above. Trials were excluded from 15 Malaysian-Chinese participants (an average of 5.92% of trials, with an SD of 5.76%, range = 1.25 to 17.25%) and none of them had more than 20% of their trials excluded within a given experimental condition. So, no participants were removed from analyses. Trials were also excluded from 15 White-British participants (an average of 5.92% of trials, with an SD of 5.76%, range = 1.25 to 17.5%), and two of those participants had more than 20% of trials (i.e., 25%) excluded from within an experimental condition. However, since these were SC_TC and SI_TI conditions that do not affect calculations of the SCE, we retained those two participants in the analyses.

Table 2

Statistics associated with biases in estimating blur

Group	Condition	M	SE	<i>t</i> -statistic	<i>p</i> -value	Cohen's <i>d</i>	<i>BF</i> ₁₀
Malaysian-	SC_TC	-0.060	0.020	-2.966	0.004	-0.368	7.191
Chinese	SI_TI	-0.042	0.018	-2.311	0.024	-0.287	1.604
White-	SC_TC	-0.071	0.023	-3.136	0.003	-0.389	11.13
British	SI_TI	-0.094	0.024	-3.860	<.001	-0.479	86.55

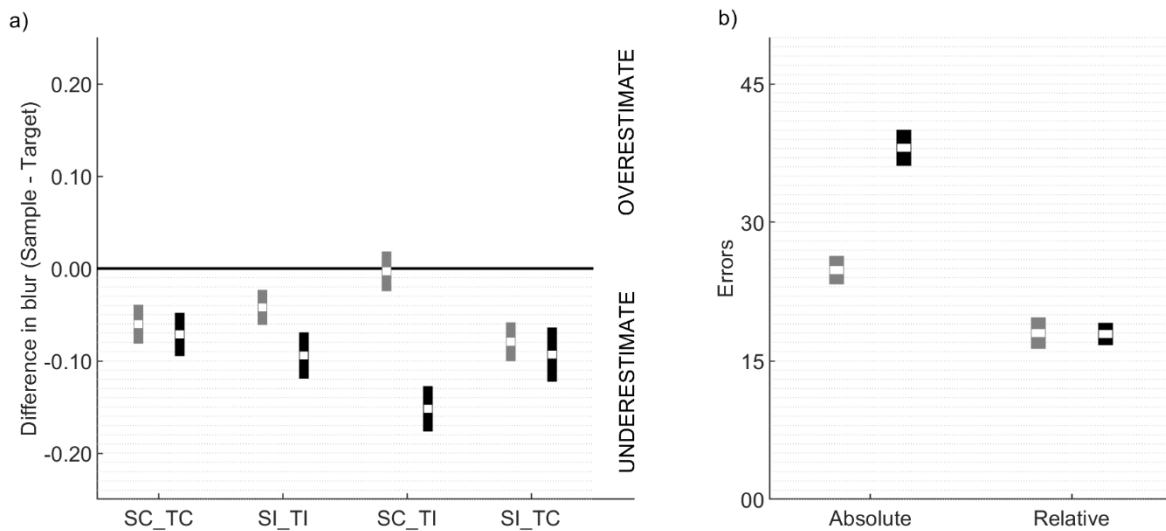
Using only included trials, for every participant, we calculated mean difference scores for the SC_TC (both sample and target in consistent scenes) and SI_TI (both sample and target in inconsistent scenes) conditions. In both the Chinese and White samples (Figure 3a), participants consistently **underestimated** the blur level of the sample, in both conditions (see Table 2). **Participants were generally biased in perceiving the sample as blurrier when it objectively had an equal blur level as the target. An independent-samples *t*-test comparing the difference in blur for the SC_TC condition between Malaysian-Chinese and White-British participants indicated that their mean difference scores were comparable, $t(128) = 0.363$, $p = 0.717$. A similar analysis for the SI_TI condition also produced a statistically non-significant difference between the two groups, $t(128) = 1.711$, $p = 0.090$. Accordingly, the two groups seemed to be comparable in response biases.**

Next, we calculated difference scores for SC_TI (sample consistent and target inconsistent) and SI_TC (sample inconsistent and target consistent) using included trials for each participant. In the Malaysian-Chinese sample (Figure 3a), SC_TI produced higher difference

scores (mean (M) = -0.003 , standard error of the mean (SE) = 0.021) than SI_TC (M = -0.079 , SE = 0.020) (Figure 3a). In White-British participants, SC_TI produced lower difference scores than (M = -0.152 , SE = 0.024) than SI_TC (M = -0.093 , SE = 0.029) (Figure 3a). We then calculated SCE scores using included trials for every participant by following Eqn. 1. The SCE scores were higher in the Chinese sample (M = 0.038 , SE = 0.013) than in the White sample (M = -0.029 , SE = 0.015), as revealed by a one-tailed independent-samples t -test, $t(123) = 3.449$, $p < 0.001$, *Cohen's d* (d) = 0.605 , with the Bayes factor showing very strong evidence in favor of the alternative hypothesis ($BF_{10} = 35.98$). In addition to what we proposed in stage 1 registration, we went one step further to confirm whether SCEs were in fact present in each culture, by comparing SCE scores against zero using Bonferroni-corrected two-tailed one-sample t -tests (adjusted significance criterion $\alpha = \frac{0.05}{2} = 0.025$), and supplementing it with the Bayesian counterpart of the same analysis. SCE scores were significantly positive in Chinese participants, $t(64) = 2.96$, $p = 0.004$, $d = 0.367$), with the Bayes factor indicating moderate evidence towards SCE being higher than zero ($BF_{10} = 7.07$). However, SCE scores did not differ from zero for White participants, $t(59) = -1.990$, $p = 0.051$, $d = 0.278$), with the Bayes factor favouring neither the null nor the alternative hypothesis ($BF_{10} = 0.863$).

Figure 3.

Summary of data obtained for the two cultural groups in a) the four experimental conditions of the object perception task, and b) the two versions of the frame-line task.



Notes: In panel “a”, grey bars denote difference in blur scores across all Malaysian-Chinese participants, and black bars denote the same for White-British participants. Difference scores above 0 indicate overestimation of the sample’s blur and those below 0 indicate underestimation. In panel “b”, grey bars denote errors (in pixels) for Malaysian-Chinese participants, whereas black bars denote the same for White-British participants. In both panels, the white band at the vertical midpoint of any bar represent the mean of a score/error; and the upper and lower edges of the bar represent associated ± 1 standard error of the mean.

Mediations from cognitive styles and self-construals. Although the stage 1 registration only intended to run mediation analyses, we also examined whether cognitive styles and self-construals indeed differed between the two cultural groups, as would be expected based on literature. Before comparing errors in the two versions of the frame-line task between the two

groups, we checked whether any of the participants' responses met the exclusion criteria, and none did for both groups. The mean errors calculated for the two versions of the task from each participant were then subjected to a 2×2 mixed-subjects analysis of variance, with participant group (Chinese and White) as the between-subjects factor and task version (absolute and relative) as the within-subjects factor. We found: 1) a significant main effect of task version, indicating higher errors in the absolute task, $F(1,128) = 80.85, p < 0.001, \eta_p^2 = 0.385$, 2) a significant main of group, $F(1,128) = 17.01, p < 0.001, \eta_p^2 = 0.117$, and 3) a significant interaction between task and group, $F(1,128) = 19.63, p < 0.001, \eta_p^2 = 0.133$. When we further explored the interaction effect using Bonferroni-corrected paired-samples t -tests adjusted significance criterion $\alpha = \frac{0.05}{2} = 0.025$, we found that errors for the absolute task were higher than those for the relative task in both Chinese ($t(64) = 3.686, p < 0.001, d = 0.457, BF_{10} = 51.44$) and White participants ($t(64) = 8.472, p < 0.001, d = 1.051, BF_{10} = 1.837 \times 10^9$). An additional independent-samples t -test showed that the difference in errors between the two tasks (absolute – relative) was smaller for the Chinese ($M = 6.82, SE = 1.85$) than the White participants ($M = 21.05, SE = 2.50$), $t(128) = -4.431, p < 0.001, d = -0.777$, with the Bayes factor strongly favouring the alternative ($BF_{10} = 917.1$). We also compared the scores calculated from the various dimensions of self-construals measured by the two questionnaires between the two groups (see Table 3), using a series of independent-samples t -tests. We found no differences between the groups for all dimensions, except for *difference vs similarity* where Chinese participants reported a greater tendency to be different from others.

Table 3

Summary of scores and statistical analyses outcomes for measured dimensions of self-construals.

Self-construal dimension	<i>M (SE)</i>		<i>t</i> -statistic	<i>p</i> -value	<i>d</i> (<i>BF</i> ₁₀)
	Malaysian- Chinese	White-British			
Interdependent [#]	56.22 (0.95)	57.72 (0.74)	-1.250	0.214	-0.219 (0.381)
Independent [#]	57.00 (0.91)	56.02 (0.94)	0.754	0.452	0.132 (0.243)
Difference vs similarity [†]	36.45 (0.88)	32.88 (0.90)	2.845	0.005	0.499 (6.938)
Self-direction vs receptiveness to influence [†]	32.80 (1.05)	34.17 (0.90)	-0.989	0.325	-0.173 (0.292)
Consistency vs variability [†]	26.48 (1.09)	26.51 (1.22)	-0.019	0.985	-0.003 (0.188)
Self-expression vs harmony [†]	31.08 (1.15)	29.52 (1.02)	1.014	0.313	0.178 (0.299)
De-contextualized vs contextualized self [†]	37.89 (1.06)	40.45 (0.87)	-1.868	0.064	-0.328 (0.906)

Note: [#] Dimensions from Singelis' (1994) self-construal scale. [†] Dimensions from the multi-dimensional scale.

Next, we report the outcomes of the series of mediation analyses that were conducted with bias-corrected bootstrapping (1000 iterations). Across all our mediation models (Table 4), the direct effect of culture on SCE scores reached statistical significance ($p < 0.001$), with a path

coefficient $\tau' = 0.034$. The simple path coefficient α (i.e., linear relationship between predictor and mediator) reached statistical significance only for the errors in the absolute version of the frame line task ($p < 0.001$) and the self-construal dimension *difference vs similarity* ($p = 0.005$). The simple path coefficient β did not reach statistical significance in any of the mediation models. None of the models satisfied the criteria for a partial or complete mediation.

Table 4

Statistics from the mediation analyses.

Model	Mediator	Path coefficients (95% confidence intervals)			
		τ'	α	β	τ
1	Absolute task errors	-0.034*	6.628*	0.0004	-0.037*
		(-0.053, -0.014)	(4.277, 9.028)	(-0.001, 0.002)	(-0.063, -0.012)
2	Relative task errors	-0.034*	-0.052	0.0007	-0.034*
		(-0.054, -0.015)	(-2.116, 1.602)	(-0.001, 0.003)	(-0.053, -0.015)
3	SIS	-0.034*	-1.246	0.001	-0.032*
		(-0.054, -0.015)	(-3.067, 0.606)	(-0.001, 0.003)	(-0.052, -0.015)
4	Difference vs Similarity	-0.034*	-1.785*	-0.0005	-0.035*
		(-0.054, -0.016)	(-3.078, -0.601)	(-0.003, 0.003)	(-0.054, -0.015)
5	Self-direction vs receptiveness to influence	-0.034*	0.685	0.0001	-0.034*
		(-0.051, -0.012)	(-0.639, 1.948)	(-0.003, 0.003)	(-0.052, -0.012)
6	Consistency vs variability	-0.034*	0.015	0.0005	-0.034*
		(-0.052, -0.014)	(-1.695, 1.491)	(-0.002, 0.003)	(-0.052, -0.014)

7	Self-expression vs harmony	-0.034*	-0.777	0.0005	-0.033*
		(-0.051,-0.014)	(-2.408,0.562)	(-0.002,0.003)	(-0.051,-0.014)
8	Decontextualised vs	-0.034*	1.277	0.0009	-0.035*
	contextualised self	(-0.053,-0.014)	(-0.086,2.670)	(-0.002,0.004)	(-0.054,-0.015)

*Note: In all models, the predictor is the cultural group, and the outcome variable is the semantic consistency effect (SCE). * indicates significance at the level of $p \leq 0.005$.*

Discussion

The predictability of an object in a scene is known to influence how that object is perceived. When an object's signal is unreliable or noisy, its perceptual representation is enhanced if it occurs in a predictable rather than an unexpected scene. However, when the same object is conspicuous, its perceptual representation is enhanced if it occurs in an unexpected scene. This interaction between object-scene semantic consistency and signal reliability is generally supported by both behavioural and neuroimaging evidence (Brandman & Peelen, 2017; Leticevscaia et al., 2024; Rossel et al., 2022, 2023). However, it is unclear if this particular semantic consistency effect manifests similarly across different cultures. Our study was the first to question this, as it would shed some insights on the extent to which different cultures rely on expectations when perceiving objects. We compared semantic consistency effects on perceived sharpness of noisy objects between adult participants from two distinct cultures: White-British and Malaysian-Chinese. We expected Malaysian-Chinese participants to experience stronger semantic consistency effects on perceived sharpness than White-British participants, and our findings confirmed this expectation.

749

750 In our study—which closely followed the experimental design of Rossel et al. (2023)—we
751 presented participants with two object-scene composites side by side. Participants adjusted
752 the blur level of an object (the “sample”) in one composite to match the blur level of the same
753 object in the second composite (the “target”). When both composites contained identical
754 scenes, participants from both cultures overestimated the blur level of the sample, regardless
755 of the semantic consistency of the scene to the object. This was clearly a response bias.
756 However, in Malaysian-Chinese participants, we also observed a systematic shift in matched
757 blur levels between the two composite pairs that were designed to capture the semantic
758 consistency effect. Specifically, when the sample’s scene was semantically consistent with the
759 object and the target’s scene was semantically inconsistent, participants overestimated the
760 blur level of the sample. In contrast, when the sample’s scene was semantically inconsistent
761 and the target’s scene was consistent, participants underestimated the blur. This shift indicates
762 that Malaysian-Chinese participants perceived the object in the consistent scene as sharper
763 when the objects in both composites had identical blur levels. Although we anticipated
764 weaker semantic consistency effects in the White-British participants, our results did not
765 support this hypothesis. While we managed to replicate the semantic consistency effect in an
766 Asian sample, we could not replicate it in a White-British sample.

767

768 When we examined the role of culture, we considered semantic consistency effects to be a
769 result of a top-down effect, where expectations/knowledge about semantic consistencies
770 guide object perception. Several past studies demonstrate that East Asians (compared to
771 Westerners) have a greater tendency to perceive a visual target in relation to its context. The
772 logic here is a cascadic effect. An independent self-construal is thought to encourage

individuals to engage in an analytical approach of processing visual information, while an interdependent self-construal encourages holistic processing (Kitayama & Salvador, 2024; Markus & Kitayama, 1991a, 1991b). The analytical tendency would make an observer less sensitive to contextual information (Ito et al., 2013; Ji et al., 2000; Kitayama et al., 2003; Masuda et al., 2008). On this basis, we hypothesised (and found) that Easterners demonstrate a stronger semantic consistency effect, insofar as scene-related expectations provide a context that guides perceiving an object (Bar, 2004). However, when East-West dichotomies in context sensitive perception are reported in past studies, the sources of this cultural difference (i.e., social constructs and/or associated cognitive styles) are in many cases *assumed* rather than explicitly measured. Therefore, we first quantified social constructs and cognitive styles and then we assessed whether these two mediate the relationship between culture and semantic consistency effects (if any).

Contrary to what is expected from cultural differences reported in the literature, we generally found self-construals to be comparable between our White-British and Malaysian-Chinese participants. The most popularly used measure of self-construals, Singelis' (1994) scale, found no difference in independence or interdependence between the two cultures. The comprehensiveness of this scale is unlikely to account for the absence of a cultural difference, because we also employed a multidimensional scale that captures nuanced variations in self-construals, with demonstrated cross-cultural reliability (Krys et al., 2021; Vignoles et al., 2016). Even across several distinct dimensions we found no clear group-level cultural differences in self-construals, with one surprising exception. Malaysian-Chinese participants reported a higher tendency to consider themselves as unique or different from others in the society (i.e., independent self-construal). Therefore, we cannot make a binary claim that the White-British participants are independent and Malaysian-Chinese are interdependent. Next,

we compared our participants' performance in the two versions of the frame-line task that was aimed at distinguishing holistic versus analytical tendencies in visual processing. Both groups of participants made more errors in the absolute version than the relative version, suggesting that both groups presented holistic tendencies. Interestingly, holistic processing tendencies were higher in White-British participants.

Our findings on self-construals contradict some large-scale research reporting clear east-west differences (e.g., Vignoles et al., 2016), but also direct comparisons between smaller convenience samples of British and (mainland) Chinese participants (Yang & Vignoles, 2020). This may not be surprising because the literature also highlights inconsistencies in demonstrating conventionally expected cultural differences in self-construals (e.g., Kitayama et al., 2009; see Cross et al., 2011 for a review). Several factors can lead to the mixed findings. First, although culture plays a dominant role in shaping one's self-construal, other factors like socioeconomic development, religious heritage, and exposure to other cultures can influence too (Lim & Ma-Kellams, 2025; Vignoles et al., 2016). Second, implicit measures of self-construals (e.g., implicit social orientation questionnaire; Kitayama & Park, 2007) may sometimes capture cross-cultural differences that are otherwise missed by explicit measures like scales (Kitayama et al., 2009; Lim & Ma-Kellams, 2025). Third, collectivism, which seems to encourage interdependent self-construals, might be reducing in Asian cultures, which would inevitably dilute East-West cultural differences in self-construals (Kim, 2024). Lastly, a convenience sample (as in our case) may not always be representative of a culture.

Findings from the frame-line task were rather counterintuitive, contradicting the classic east-west dichotomy in cognitive styles (Chua et al., 2022; Ji et al., 2000; Kitayama et al., 2003; McKone et al., 2010; Nisbett & Miyamoto, 2005). However, the frame-line task faces a replication crisis too. There are many instances where comparable holistic tendencies are reported between Easterners and Westerners, for instance between Americans and Chinese (Hakim et al., 2017), or between Germans and Japanese (Kitayama et al., 2009). These mixed findings could be due to the variability in the convenience samples selected or could in fact be due to the malleability of culture-distinguishing factors like holistic tendencies (Hakim et al., 2017). More importantly, despite self-construals being largely comparable between the two cultures that we tested, the British demonstrated stronger holistic tendencies, contradicting the conventional mapping between self-construals and cognition. We found the qualitative parallel between our British participants' greater holistic tendencies and their lower scores for one self-construal dimension (*difference versus similarity*) worthy of further exploration, as it suggested a positive relationship between holistic tendencies and interdependent self-construals. However, when we correlated the two we did not find any associations (Pearson's $r = -0.110$, $p = 0.211$). Overall, we believe it is safe to conclude that we observed no clear relationships between self-construals and holistic tendencies in visual processing (see Kitayama et al., 2009; McKone et al., 2010 for similar findings).

Of more pertinence to our study is the finding that neither self-construals nor the behavioural measures of cognitive styles mediated the relationship between culture and semantic consistency effects. Sometimes, assuming a mean difference between cultures can be problematic, as we have done in our mediation analyses, because it implies dichotomy rather than a continuous relationship. For instance, in Chua et al.'s (2022) study, Asian and European participants were indistinguishable in self-construals based on group-level

comparisons, and yet, levels of independent self-construals predicted holistic tendencies in visual processing. However, when we explored path coefficients in the mediation models, the relationships between self-construals/holistic tendencies and semantic consistency effects (after controlling for culture) were negligible. We emphasise on two things here: 1) cultural differences in semantic consistency effects that we observe cannot be reduced to individual differences in self-construals or holistic tendencies in cognition, and 2) it is highly problematic to make blanket assumptions and attribute any observed cultural difference in visual perception to self-construals and/or cognitive styles, unless otherwise the said characteristics are also measured.

In our case, nominal culture clearly exerted dissociable effects on context sensitivity in the frame-line and object perception tasks. White-British participants who were most sensitive to context in the frame-line task (demonstrating greater holistic tendencies) did not experience semantic consistency effects guided by contextual knowledge. Malaysian-Chinese participants who were less sensitive to the context in the frame-line task experienced stronger semantic consistency effects. Moreover, context sensitivity in the frame-line task (measured as errors in the absolute version) did not mediate the relationship between culture and semantic consistency effects either.

Often, the holistic tendency in the frame-line task is explained by invoking attentional control mechanisms, with those demonstrating greater context sensitivity (e.g., Eastern cultures) having a greater propensity to *attend* to contextual information outside of focal targets (Hakim et al., 2017; Hedden et al., 2008; Kitayama et al., 2003; Nisbett & Miyamoto, 2005). Whether such culture-specific attentional biases translate well into object processing within

complex scenes is questionable. Some studies (e.g., Chua et al., 2005) indicate that Eastern observers distribute overt attention more evenly between focal objects and their surrounding scenes, rendering perception and memory for focal objects more dependent on background context. In contrast, Western observers tend to concentrate overt attention on focal objects, resulting in perception and memory that are relatively independent of contextual. However, such cultural distinctions are not always replicated (Evans et al., 2009). Of specific interest to us is a recent study by Leticevscaia et al. (2024) which involved decoding object categories from magnetoencephalography responses in temporal lobe regions like the visual cortex. They found that neural representations of objects are enhanced from expectations arising from the scene context - a typical semantic consistency effect. Interestingly, this effect occurred even when participants did not pay attention to the scene, suggesting that they are purely driven by automatic influences from scene-related expectations rather than attention. This makes dissociable context sensitivities in the two tasks (frame-line and object perception) less surprising in our participants, as they can be mediated by distinct top-down mechanisms.

So, what could be driving stronger semantic consistency effects in our Malaysian-Chinese sample than the White-British sample? One possibility is that statistical object-scene regularities learnt by the two cultures may have been different. However, we minimised interference from this by conducting the stimulus validation process. All objects we used in the main study were equally familiar (and at the ceiling level) to participants from both cultures, and both cultures provided comparable semantic consistency ratings for all our object-scene pairs (consistent as well as inconsistent). However, comparable semantic consistency ratings do not strictly mean that prior expectations about object-scene regularities are identical between two groups of participants. An individual who has associated multiple

objects to a scene may give similar semantic consistency ratings to a given object (e.g., A) as another individual who has associated only that given object (A) to the scene. This leads us to the second possibility—which is largely speculative at this point—that fundamental differences in aspects related to predictive coding (e.g., nature of prior expectations, sensitivity to prediction errors) may account for the observed cultural differences.

It must be noted that the second possibility is grounded on the assumption that the semantic consistency effect is predominantly a result of statistical regularities informing object processing. This is plausible because: 1) scene selective regions (e.g., those part of the Parahippocampal cortex) do not just encode scenes, but they are also implicated in learning about object-scene associations, and 2) these scene selective regions have feedback pathways to object processing regions that are distinct from feedback pathways involved in other forms of top-down guidance, like attentional control which is engaged during the frame-line task (Aminoff et al., 2013; Bar, 2004; Hedden et al., 2008; Leticevscaia et al., 2024). The engagement of distinct feedback mechanisms might also explain why overarching interpretive constructs like self-construals and/or holistic cognition did not influence semantic consistency effects in our case.

Explaining cultural differences with regards to predictive coding is difficult because of the lack of findings and the mixed nature of existing findings. In some cases, individual differences in visual perception are attributed to differences in the nature of prior expectations. For instance, cultural differences in susceptibility to visual illusions arising from spatial context has been attributed to differences in prior expectations that mimic the statistical regularities in their respective (distinct) environments (Bremner et al., 2016). In

addition, reduced susceptibility to visual illusions in autistic individuals are attributed to broader prior expectations (Pellicano and Burr, 2020), although autistic traits cannot always explain individual differences in visual perception involving prior expectations (e.g., Tulver et al., 2019). As far as sensitivity to prediction errors are concerned, we only found one study that explored cross-cultural differences, and they reported no differences (Wright et al., 2018). Nonetheless, whether fundamental differences in predictive coding can explain cultural differences in semantic consistency effects remains an open question that requires further inquiry.

Registered protocol and data availability statement

We have made a copy of the stage 1 registration protocol (which had received a principle acceptance on the 27th of February 2025), anonymized raw participant data, data summaries, and Matlab codes to run behavioral experiments, publicly available in an Open Science Framework's archive: <https://doi.org/10.17605/OSF.IO/J74DE>

941

942

943

944

945

946

947

References

- 948 Aminoff, E. M., Kveraga, K., & Bar, M. (2013). The role of the parahippocampal cortex in
949 cognition. *Trends in cognitive sciences*, 17(8), 379-390.
- 950 Bar, M. (2004). Visual objects in context. *Nature Reviews Neuroscience*, 5(8), 617-629.
- 951 Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social
952 psychological research: Conceptual, strategic, and statistical considerations. *Journal*
953 *of personality and social psychology*, 51(6), 1173.
- 954 Biederman, I., Mezzanotte, R. J., & Rabinowitz, J. C. (1982). Scene perception: Detecting
955 and judging objects undergoing relational violations. *Cognitive psychology*, 14(2),
956 143-177.
- 957 Brainard, D. H. (1997). The psychophysics toolbox. *Spatial vision*, 10(4), 433-436.
- 958 Brandman, T., & Peelen, M. V. (2017). Interaction between scene and object processing
959 revealed by human fMRI and MEG decoding. *Journal of Neuroscience*, 37(32), 7700-
960 7710.
- 961 Castelano, M. S., & Henderson, J. M. (2007). Initial scene representations facilitate eye
962 movement guidance in visual search. *Journal of Experimental Psychology: Human*
963 *Perception and Performance*, 33(4), 753.
- 964 Chua, H. F., Boland, J. E., & Nisbett, R. E. (2005). Cultural variation in eye movements
965 during scene perception. *Proceedings of the national academy of sciences*, 102(35),
966 12629-12633.
- 967 Chua, S. Y., Rentzelas, P., Frangou, P., Kourtzi, Z., Lintern, M., & Mavritsaki, E. (2022).
968 Cultural differences in visual perceptual learning. *International Journal of*
969 *Psychology*, 57(3), 377-386.
- 970 Cornelissen, T. H., & Võ, M. L. H. (2017). Stuck on semantics: Processing of irrelevant
971 object-scene inconsistencies modulates ongoing gaze behavior. *Attention, Perception,*
972 *& Psychophysics*, 79, 154-168.
- 973 Cross, S. E., Hardin, E. E., & Gercek-Swing, B. (2011). The what, how, why, and where of
974 self-construal. *Personality and Social Psychology Review*, 15(2), 142-179.
- 975 Davenport, J. L., & Potter, M. C. (2004). Scene consistency in object and background
976 perception. *Psychological science*, 15(8), 559-564.
- 977 Evans, K., Rotello, C. M., Li, X., & Rayner, K. (2009). Scene perception and memory

- revealed by eye movements and receiver-operating characteristic analyses: Does a cultural difference truly exist?. *The Quarterly Journal of Experimental Psychology*, 62(2), 276-285.
- Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior research methods*, 39(2), 175-191.
- Fize, D., Fabre-Thorpe, M., Richard, G., Doyon, B., & Thorpe, S. J. (2005). Rapid categorization of foveal and extrafoveal natural images: Associated ERPs and effects of lateralization. *Brain and cognition*, 59(2), 145-158.
- Fritz, M. S., & MacKinnon, D. P. (2007). Required sample size to detect the mediated effect. *Psychological science*, 18(3), 233-239.
- Furmanski, C. S., & Engel, S. A. (2000). An oblique effect in human primary visual cortex. *Nature neuroscience*, 3(6), 535-536.
- Girshick, A. R., Landy, M. S., & Simoncelli, E. P. (2011). Cardinal rules: visual orientation perception reflects knowledge of environmental statistics. *Nature neuroscience*, 14(7), 926-932.
- Hakim, N., Simons, D. J., Zhao, H., & Wan, X. (2017). Do easterners and westerners differ in visual cognition? A preregistered examination of three visual cognition tasks. *Social Psychological and Personality Science*, 8(2), 142-152.
- Hardin, E. E., Leong, F. T., & Bhagwat, A. A. (2004). Factor structure of the self-construal scale revisited: Implications for the multidimensionality of self-construal. *Journal of Cross-Cultural Psychology*, 35(3), 327-345.
- Happé, F., & Frith, U. (2006). The weak coherence account: detail-focused cognitive style in autism spectrum disorders. *Journal of autism and developmental disorders*, 36, 5-25.
- Hedden, T., Ketay, S., Aron, A., Markus, H. R., & Gabrieli, J. D. (2008). Cultural influences on neural substrates of attentional control. *Psychological science*, 19(1), 12-17.
- Hong, Y. Y., Morris, M. W., Chiu, C. Y., & Benet-Martinez, V. (2000). Multicultural minds: A dynamic constructivist approach to culture and cognition. *American psychologist*, 55(7), 709.
- Hussain Ismail, A. M., Solomon, J. A., Hansard, M., & Mareschal, I. (2019). A perceptual bias for man-made objects in humans. *Proceedings of the Royal Society B*, 286(1914), 20191492.
- Ito, K., Masuda, T., & Man Wai Li, L. (2013). Agency and facial emotion judgment in context. *Personality and Social Psychology Bulletin*, 39(6), 763-776.
- Ji, L. J., Peng, K., & Nisbett, R. E. (2000). Culture, control, and perception of relationships in the environment. *Journal of personality and social psychology*, 78(5), 943.
- Ji, L. J., & Yap, S. (2016). Culture and cognition. *Current opinion in Psychology*, 8, 105-111.
- Kim, S. Y. (2024). Examining 35 years of individualism-collectivism research in Asia: A meta-analysis. *International Journal of Intercultural Relations*, 100, 101988.
- Kitayama, S., Duffy, S., Kawamura, T., & Larsen, J. T. (2003). Perceiving an object and its context in different cultures: A cultural look at new look. *Psychological science*, 14(3), 201-206.
- Kitayama, S., & Park, H. (2007). Cultural shaping of self, emotion, and well-being: How does it work?. *Social and Personality Psychology Compass*, 1(1), 202-222.
- Kitayama, S., Park, H., Sevincer, A. T., Karasawa, M., & Uskul, A. K. (2009). A cultural task analysis of implicit independence: comparing North America, Western Europe, and East Asia. *Journal of personality and social psychology*, 97(2), 236.
- Krys, K., Park, J., Kocimska-Zych, A., Kosiarczyk, A., Selim, H. A., Wojtczuk-Turek, A., ...

- & Adamovic, M. (2021). Personal life satisfaction as a measure of societal happiness is an individualistic presumption: Evidence from fifty countries. *Journal of Happiness Studies*, 22, 2197-2214.
- LaPointe, M. R., Lupianez, J., & Milliken, B. (2013). Context congruency effects in change detection: Opposing effects on detection and identification. *Visual Cognition*, 21(1), 99-122.
- Leticevscaia, O., Brandman, T., & Peelen, M. V. (2024). Scene context and attention independently facilitate MEG decoding of object category. *Vision Research*, 224, 108484.
- Lim, J., & Ma-Kellams, C. (2025). Comparing self-construal scales: Cultural implications for the United States, East Asia, Southeast Asia, Latin America, and the Middle East. *Methods in Psychology*, 12, 100189.
- Lupyan, G. (2017). Objective effects of knowledge on visual perception. *Journal of experimental psychology: human perception and performance*, 43(4), 794.
- Malcolm, G. L., & Henderson, J. M. (2010). Combining top-down processes to guide eye movements during real-world scene search. *Journal of Vision*, 10(2), 4-4.
- Markus, H. R. & Kitayama, S. (1991a). Cultural variation in the self-concept. *The Self: Interdisciplinary approaches/Springer*.
- Markus, H.R., & Kitayama, S. (1991b). Culture and the self: Implications for cognition, emotion and motivation. *Psychological Review*, 98(2), 224-253
- Masuda, T., Ellsworth, P. C., Mesquita, B., Leu, J., Tanida, S., & Van de Veerdonk, E. (2008). Placing the face in context: cultural differences in the perception of facial emotion. *Journal of personality and social psychology*, 94(3), 365.
- Matsumoto, D. (1999). Culture and self: An empirical assessment of Markus and Kitayama's theory of independent and interdependent self-construals. *Asian Journal of Social Psychology*, 2(3), 289-310.
- McKone, E., Davies, A. A., Fernando, D., Aalders, R., Leung, H., Wickramariyaratne, T., & Platow, M. J. (2010). Asia has the global advantage: Race and visual attention. *Vision research*, 50(16), 1540-1549.
- Nisbett, R. E., & Miyamoto, Y. (2005). The influence of culture: holistic versus analytic perception. *Trends in cognitive sciences*, 9(10), 467-473.
- Ortiz-Tudela, J., Jiménez, L., & Lupiáñez, J. (2023). Scene-object semantic incongruity across stages of processing: From detection to identification and episodic encoding. *Frontiers in Cognition*, 2, 1125145.
- Palmer, T. E. (1975). The effects of contextual scenes on the identification of objects. *Memory & cognition*, 3(5), 519-526.
- Park, J., & Kitayama, S. (2014). Interdependent selves show face-induced facilitation of error processing: Cultural neuroscience of self-threat. *Social cognitive and affective neuroscience*, 9(2), 201-208.
- Peelen, M. V., Berlot, E., & de Lange, F. P. (2024). Predictive processing of scenes and objects. *Nature Reviews Psychology*, 3(1), 13-26.
- Pelham, B., Hardin, C., Murray, D., Shimizu, M., & Vandello, J. (2022). A truly global, non WEIRD examination of collectivism: The global collectivism index (GCI). *Current Research in Ecological and Social Psychology*, 3, 100030.
- Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial vision*, 10, 437-442.
- Rao, R. P., & Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature neuroscience*, 2(1), 79-87.

- Rezlescu, C., Danaila, I., Miron, A., & Amariei, C. (2020). More time for science: Using Testable to create and share behavioral experiments faster, recruit better participants, and engage students in hands-on research. *Progress in brain research*, 253, 243-262
- Rossel, P., Peyrin, C., Roux-Sibilon, A., & Kauffmann, L. (2022). It makes sense, so I see it better! Contextual information about the visual environment increases its perceived sharpness. *Journal of experimental psychology: human perception and performance*, 48(4), 331.
- Rossel, P., Peyrin, C., & Kauffmann, L. (2023). Subjective perception of objects depends on the interaction between the validity of context-based expectations and signal reliability. *Vision Research*, 206, 108191.
- Spaak, E., Peelen, M. V., & de Lange, F. P. (2022). Scene context impairs perception of semantically congruent objects. *Psychological Science*, 33(2), 299-313.
- Singelis, T. M. (1994). The measurement of independent and interdependent self construals. *Personality and social psychology bulletin*, 20(5), 580-591.
- Tse, P. U., & Cavanagh, P. (2000). Chinese and Americans see opposite apparent motions in a Chinese character. *Cognition*, 74(3), B27-B32.
- Tulver, K., Aru, J., Rutiku, R., & Bachmann, T. (2019). Individual differences in the effects of priors on perception: A multi-paradigm approach. *Cognition*, 187, 167-177.
- Vignoles, V. L., Owe, E., Becker, M., Smith, P. B., Easterbrook, M. J., Brown, R., ... & Bond, M. H. (2016). Beyond the 'east-west' dichotomy: Global variation in cultural models of selfhood. *Journal of Experimental Psychology: General*, 145(8), 966.
- Weigl, M., Pietsch, J., Kapsali, E., Shao, Q., Zheng, Z., Li, J., Kray, J., & Mecklinger, A. (2024). ORCA: a picture database of object-scene arrangements for cross-cultural and aging research. *Behavior Research Methods*, 56(2), 513-528.
- Wiebel, C. B., Valsecchi, M., & Gegenfurtner, K. R. (2013). The speed and accuracy of material recognition in natural images. *Attention, Perception, & Psychophysics*, 75, 954-966.
- Willey, C. R., & Liu, Z. (2022). Re-assessing the role of culture on the visual orientation perception of the rod and frame test. *PlosOne*, 17(10), e0276393.
- Wright, N. D., Grohn, J., Song, C., Rees, G., & Lawson, R. P. (2018). Cultural effects on computational metrics of spatial and temporal context. *Scientific Reports*, 8(1), 2027.
- Wu, Chia-Chien, Farahnaz Ahmed Wick, and Marc Pomplun. "Guidance of visual attention by semantic information in real-world scenes." *Frontiers in psychology* 5 (2014): 54.
- Yang, S. (2018). *Multidimensional self-construals: Testing the model and refining measurement* (Doctoral dissertation, University of Sussex).
- Yang, S., & Vignoles, V. L. (2020). Self-construal priming reconsidered: Comparing effects of two commonly used primes in the UK and China. *Open Science Journal*, 5(3).

1126
1127
1128
1129
1130
1131
1132
1133
1134
1135
1136
1137
1138
1139
1140
1141
1142
1143
1144