Reproducibility of visual-field asymmetries

Brederoo, Sanne G.; Nieuwenstein, Mark R.; Cornelissen, Frans W.; Lorist, Monicque M.

Published in:
Cortex

DOI:
10.1016/j.cortex.2018.10.021

Accepted Manuscript

Reproducibility of visual-field asymmetries: Nine replication studies investigating lateralization of visual information processing

Sanne G. Brederoo, Mark R. Nieuwenstein, Frans W. Cornelissen, Monicque M. Lorist

PII: S0010-9452(18)30356-3

DOI: https://doi.org/10.1016/j.cortex.2018.10.021

Reference: CORTEX 2455

To appear in: *Cortex*

Received Date: 13 June 2018

Accepted Date: 17 October 2018

Reproducibility of visual-field asymmetries: Nine replication studies investigating lateralization of visual information processing

Sanne G. Brederoo[a,b,c]*, Mark R. Nieuwenstein[a,b], Frans W. Cornelissen[b,c,d], & Monicque M. Lorist[a,b,c]

[a] Experimental Psychology, Faculty of Behavioural and Social Sciences, University of Groningen, Groningen, The Netherlands

[b] Research School Behavioural and Cognitive Neurosciences, University of Groningen, Groningen, The Netherlands

[c] Neuroimaging Center Groningen, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands

[d] Laboratory of Experimental Ophthalmology, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands

**Corresponding author:*

Postal address: Dept. of Experimental Psychology, Grote Kruisstraat 2/1, 9712 TS, Groningen, The Netherlands

Office phone: +31 50 3636879

Email addresses:

sannebrederoo@gmail.com (S.G. Brederoo)

m.r.nieuwenstein@rug.nl (M.R. Nieuwenstein)

f.w.cornelissen@umcg.nl (F.W. Cornelissen)

m.m.lorist@rug.nl (M.M. Lorist)

Abstract

Numerous behavioral studies suggest that the processing of various types of visual stimuli and features may be more efficient in either the left or the right visual field. However, not all of these visual-field asymmetries (VFAs) have been observed consistently. Moreover, it is typically unclear whether a failure to observe a particular VFA can be ascribed to certain characteristics of the participants and stimuli, to a lack of statistical power, or to the actual absence of an effect. To increase our understanding of lateralization of visual information processing, we have taken a rigorous methodological and statistical approach to examine the reproducibility of various previously reported VFAs. We did so by performing (near-)exact replications of nine representative previous studies, aiming for sufficient power to detect the effects of interest, and taking into consideration all relevant dependent variables (reaction times and error rates). Following Bayesian analyses –on our data alone as well as on the combined evidence from the original and replication studies– we find precise and reliable evidence that support VFAs in the processing of faces, emotional expressions, global and local information, words, and in the distribution of spatial attention. In contrast, we find less convincing evidence for VFAs in processing of high and low spatial frequencies. Finally, we find no evidence for VFAs in categorical perception of color and shape oddballs, and in the judgments of categorical and coordinate spatial relations. We discuss our results in the light of their implications for theories of visual lateralization.

Keywords: visual-field asymmetries; replication; lateralization; Bayes factor; behavioral tasks

1. Introduction

Depending on the nature of visual information, presenting it in either the left (LVF) or right (RVF) visual field can influence the efficiency with which observers process it. Behavioral experiments in which visual stimuli are presented to the LVF and RVF have, for example, shown that the majority of observers show LVF-advantages for face information, while they show RVF-advantages for words. The visual-field asymmetries (VFAs) resulting from such visual half-field or free-viewing tasks have been suggested to reflect differential hemispheric specialization, or lateralization, of the processing of different types of visual information (Beaumont, 1982; Bourne, 2006; Voyer, Voyer, & Tramonte, 2012).

Over the past decades, behavioral experiments have demonstrated VFAs for a variety of stimulus types, and these phenomena have in turn formed the basis for a number of theories regarding lateralization of visual information processing (for overviews, see Hellige, 1995; Dien, 2008; Hellige, Laeng, & Michimata, 2010; Karim & Kojima, 2010). Importantly, however, there is reason for concern about the reliability of some of these findings. Specifically, a number of VFAs extracted in such studies tend to have a relatively low test-retest and split-half reliability, when compared to behavioral asymmetries in the auditory domain (Voyer, 1998), and the results of different studies on the same types of visual information often lack consistency in their outcomes. As a case in point, consider the results of studies investigating the lateralization of global and local information processing of hierarchical stimuli. While the general assumption is that there is an RVF-advantage when processing of the local elements is task-relevant, and an LVF-advantage when processing the global form is task-relevant (Van Kleeck, 1989), most studies using visual half-field tasks with hierarchical stimuli have found evidence for only one of these two VFAs (for a recent review, see Brederoo, Nieuwenstein, Lorist, & Cornelissen, 2017). Concomitantly, the interpretation of such failures to demonstrate a particular VFA is often difficult because it is unclear whether a null result can be taken as evidence for the null hypothesis or as evidence that the study did not have sufficient power to detect the effect of interest.

The inconsistent findings have promoted the approach of using convergent evidence from, for example, patient and neuroimaging studies, to arrive at insights about the extent to which the left (LH) and right (RH) hemispheres might be specialized for processing certain types of visual input. While this approach provides insight into whether lateralization occurs at the implementational, neural level, the investigation of which aspects of lateralization also produce reliable behavioral effects is an important enterprise in its own right, for several reasons. To start, the availability of reliable behavioral manifestations of lateralization can be of practical importance in distinguishing between clinical populations (Luh & Gooding, 1999) and in studying the effects of aging (Lux, Marshall, Thimm, & Fink, 2008). Secondly, behavioral studies are usually cheaper and easier to implement than patient or neuroimaging studies, and they therefore provide a highly useful means to examine how various factors influence the lateralized processing of visual information. Lastly, insight into the behavioral manifestations of lateralization is also of importance for practical reasons when it comes to designing applications aimed at maximizing the efficiency of visual information processing. For these reasons, verifying the reliability of behavioral indices of lateralization of visual information processing is valuable for the field.

In the current study, we investigated the reliability of several behavioral manifestations of lateralized visual information processing by determining whether we could replicate the earlier-found VFAs. The importance of replication research has received growing emphasis by the scientific community in recent years. Researchers (e.g., Pashler & Wagenmakers, 2012; Schmidt, 2009) and journal editors (Wagenmakers & Forstmann, 2014) have been encouraged to improve reproducibility of scientific findings by engaging in replication research, of which the large-scale replication project of the Open Science Framework is an example (Open Science Collaboration, 2015). This project raised awareness of the importance of studying reproducibility of effects in psychological science, and stressed that "Replication can increase certainty when findings are reproduced and promote innovation when they are not." (Open Science Collaboration, 2015, p. 7). With this goal in mind, we attempted to replicate nine studies that yielded evidence for lateralization of visual information processing in behavioral outcomes, with each targeting a different type of visual information.

In selecting our targets for the replication studies, we aimed to arrive at a representative set of tasks that have previously been found to yield VFAs for various types of visual features and stimuli. Specifically, our selection included several phenomena that have dominated the field of visual lateralization research over the past 50 years (i.e., VFAs for neutral and emotional faces, global and local visual information, high and low spatial frequencies, categorical and coordinate spatial relations, the distribution of spatial attention, and visually presented words), as well as some that have resulted from more recent studies (i.e., VFAs showing categorical effects in the perception of colors and shapes). Importantly, this selection of phenomena also entailed the inclusion of studies employing different presentation conditions (e.g., free-viewing and visual-half field paradigms) and exposure durations (from 30 ms to 10 s) for a wide diversity of tasks and outcome measures (i.e., target detection, target identification, S1-S2 matching, choice bias), thereby yielding a broad range of phenomena that can be said to be representative of previous studies examining the behavioral manifestations of lateralized visual information processing. Accordingly, our study not only allowed for an examination of the reproducibility of a large number of VFAs found in previous studies, but it also enabled us to examine how reproducibility varied across VFAs for different types of visual information and tasks.

In designing our replication studies, we strove to replicate the original experiments as exactly as possible –either by copying the original methods or by using the original experiment programs when possible– and we conducted a priori power analyses to ensure that our sample sizes would be large enough to have sufficient power to observe the effects of interest. In addition, we examined both error rates (ERs) and reaction times (RTs), so as to allow us to exclude the occurrence of a speed-accuracy trade-off as an alternative account of any observed lateralization effect. Furthermore, in addition to a more conventional analysis using null hypothesis significance testing (NHST), we used Bayesian analyses, as these enable an assessment of the extent to which a non-significant outcome provides evidence in favor of the null hypothesis (Dienes & Mclatchie, 2017). Lastly, we also calculated a meta-analytical Bayes factor (Rouder & Morey, 2011), which is a novel Bayesian analysis method that combines results of several studies in order to arrive at a more robust estimate of the presence or absence of a particular effect.

2. General Methods

2.1 Tasks

Each of the to-be replicated tasks had been described in more than one earlier published study. For our replication studies, we selected those studies that were pioneering, or were an updated version of pioneering tasks, based on more recent findings. The tasks used were the Face Similarity Task (FST) (C. Gilbert & Bakan, 1973), Face Emotionality Task (FET) (Levy, Heller, Banich, & Burton, 1983), Hierarchical Letter Task (HLT) (Yovel et al., 2001), Picture Matching Task (PMT) (Peyrin, Mermillod, et al., 2006), Color Oddball Task (COT) (A.L. Gilbert et al., 2006), Shape Oddball Task (SOT) (A.L. Gilbert et al., 2008), Cross-dot Matching Task (CMT) (Van der Ham & Borst, 2011, 2016), Landmark Task (LT) (Linnell et al., 2014), and Lexical Decision Task (LDT) (Willemin et al., 2016).

2.2 Participants

Participants were recruited from the student population of the University of Groningen. All participants were right-handed as assessed by self-report (LT), measured using the Edinburgh Handedness Inventory (Oldfield, 1971) (LDT), or measured using the Flanders handedness questionnaire (Nicholls, Thomas, Loetscher, & Grimshaw, 2013) (all other tasks). All participants had normal or corrected-to-normal vision, which was measured using a Snellen test (PMT), or based on participants' self-report (all other tasks). Participants received course credits or a monetary compensation in exchange for their participation. The ethical committee of the Psychology Department of the University of Groningen approved all experiments, and participants always gave written informed consent before the start of an experiment.

To determine the minimum number of participants needed to find the smallest effect of interest in the original study with 80% power (at α = .05, one-sided), we conducted power analyses using the G*Power 3.1.9.2 software (Faul, Erdfelder, Lang, & Buchner, 2007), based on the original study's effect sizes (Cohen's d_z). The achieved power for each of the effects of interest is reported below, in the subsections where we report the results of each study.

2.3 Procedure

The experiments took place in a dimly lit and sound-attenuating cabin. Stimuli were presented on a 22" (1280 x 1024, 100 Hz, Iiyama Vision Master Pro 513) or 19" (1024 x 768, 100 Hz, Iiyama Vision Master Pro 454) CRT-monitor. In each experiment the distance to the monitor was fixed using a chin rest. The experiments were implemented in DMDX (Forster & Forster, 2003) (LDT), or E-Prime 2.0 (Psychology Software Tools, Pittsburgh, PA) (all other tasks), running on a Windows 7 operating system. Responses were collected using a QWERTY-keyboard (LT; LDT) or an in-house manufactured button box (all other tasks).

2.4 Statistical analyses

In all analyses, we subtracted performance on RVF-trials from performance on LVF-trials, and therefore any negative test statistic indicates an LVF-advantage whereas any positive test statistic indicates an RVF-advantage. For studies that examined VFAs across different task conditions (HLT; PMT; COT; SOT; CMT), we conducted planned comparisons for the visual-field contrasts even when the repeated measures ANOVA did not show a significant interaction with task condition. The ANOVA tables describing the results of the full models can be found in Appendix A.

In line with the original studies' analyses, we report the outcomes of one-sided dependent samples *t*-tests contrasting LVF- and RVF-performance, or one-sample *t*-tests comparing a VFA to a mean of zero. However, to decide on the success or failure of a replication, rather than using frequentists *t*-tests and focusing on the *p*-

value that can be derived from such a test, we used Bayesian *t*-tests (using the BayesFactor package for R). The reason for this is that the frequentist statistical method allows the researcher to reject the null hypothesis, but not to accept it, and as such does not allow the conclusion that a replication attempt has failed. The Bayes factors that we derived from the Bayesian *t*-tests reflect the amount of evidence in favor or against the alternative and null hypotheses, thus allowing us to decide on the success or failure of our replication. To interpret the resulting Bayes factors (BF_{10}) we adopted the classifications proposed by Jeffreys (1961) (i.e., a BF_{10} > 3.16, > 10, > 31.6, or > 100 respectively entails substantial, strong, very strong, or decisive evidence for the alternative hypothesis, while a BF_{10} < .316, < .1, < .0316, or < .01 respectively entails substantial, strong, very strong, or decisive evidence for the null hypothesis)[1]. In our analyses, we concluded a VFA was successfully replicated when the BF_{10} exceeded 3.16, and we concluded that the replication had failed when the BF_{10} was below .316. When the BF_{10} was within the .316 – 3.16 interval, we concluded that there was not sufficient evidence to decide on the success or failure of the replication.

2.4.1 Effects of interest. The nine studies that we attempted to replicate produced a variety of outcome measures. Specifically, three of the experiments produced a measure of bias towards one of the visual fields (FST; FET; LT), while the effects for the other six experiments were expressed in terms of differences in ERs and/or RTs. Four experiments compared conditions for which opposing VFAs were expected (HLT; COT; SOT; LDT), and two experiments additionally measured the effect of a modulating task factor (PMT) or participant factor (CMT) upon the

---

[1] Alternative classifications have been proposed (e.g., Dienes, 2014), but these would lead to a more liberal approach in deciding a replication has failed, rendering them less suitable for the current studies.

found VFAs. The effects of interest in the replication studies were restricted to those outcomes that yielded a significant effect (i.e., had a *p*-value smaller than .05) in the original study.

2.4.2 Additional analyses. Aside from examining the replicability of the effects that were found to be significant in the original studies, we also conducted a number of additional analyses. To start, we examined each VFA in terms of differences in both ERs and RTs. The motivation for examining both ERs and RTs was to determine whether a speed-accuracy trade-off occurred, and whether such a trade-off could explain any discrepancy between the effects found in the original study and in our replication attempt (Hellige & Sergent, 1986). In addition, a test of both RTs and ERs appeared to be warranted by logic, as any beneficial effect of hemispheric specialization could in principle surface in both accuracy and processing time.

A second point of departure from the original analyses derived from the fact that each of the studies that tested the LVF-RVF contrasts under different task conditions (HLT; PMT; COT; SOT; CMT) failed to find some of the predicted VFAs. Since four of these studies used relatively small sample sizes ($N < 17$), these studies may have been underpowered to detect all predicted VFAs. Therefore, we additionally examined the VFAs that were predicted based on theory, but not found in the original studies.

2.4.3 Combined evidence. Finally, for each of the predicted VFAs (significant and non-significant) in the original studies, we calculated a combined Bayes factor based on the statistics of the effect in the original and replication studies. This meta-analytic Bayes factor (Rouder & Morey, 2011) allows the assessment of the total amount of evidence for the predicted VFAs under study (i.e., the effects of interest as well as those effects addressed with the additional analyses).

3. General Results

3.1 Data Exclusion

Data of participants whose accuracy did not exceed 50% were excluded from the analyses. This resulted in exclusion of 18 of the 322 (i.e., 5.6%) tested participants (HLT: 7; PMT: 2; COT: 1; SOT: 6; CMT: 2). The ensuing descriptions of the participants in each of the replication studies pertain to the remaining participants who were included in the analyses.

For all analyses of RTs, we first subjected the data to the outlier removal procedure described by Van Selst and Jolicoeur (1994). The percentage of trials removed as a result of this procedure ranged between 1.6% and 2.7% over studies.

4. Replication Studies

In the following sections, we describe the experimental set-up, methods and results for each of the nine replication studies and we provide a short discussion of the results. In cases in which we did not successfully replicate an effect, we discuss whether differences between the original and replication studies might have caused this. The presentation of the nine replication studies is ordered by the publication dates of the original studies.

4.1 Face Similarity Task (FST)

Faces have been suggested to be the most widely studied type of visual stimulus (Yovel, Wilmer, & Duchaine, 2014). The first to show an LVF-bias for face processing in a group of healthy adults were C. Gilbert and Bakan (1973). They asked participants to judge the similarity of construed symmetric face images to the original face images. Specifically, participants had to choose between a symmetric face image that was made by mirroring the left half of the original face, and a symmetric face that was made by mirroring the right half of the original face. The right-handed participants more often found the left-side symmetric composite to resemble the original face than the right-side symmetric composite. This finding was interpreted to indicate a bias towards the LVF in perceiving faces, caused by RH-dominance in face processing (C. Gilbert & Bakan, 1973). This free-viewing face paradigm and adaptations of it have been widely used since (for an overview, see Voyer, Voyer, & Tramonte, 2012). The current study is a replication attempt of the pioneering Face Similarity Task (FST) used by C. Gilbert and Bakan (1973; Experiment 4 (subsample of right-handed participants)).

4.1.1 Methods.

4.1.1.1 Participants. Thirty-four participants (17 women) performed the FST. Their mean age was 20 years (range = 18-27).

4.1.1.2 Stimuli. Fifty-three neutral face images (28 female and 25 male) photographed in straight view were selected from the Karolinska Directed Emotional Faces (KDEF) face database (Lundqvist, Flykt & Öhman, 1998). For each of the original images, we also created two mirror images in which the face was mirrored along the vertical axis. By using both the original and the mirrored images, we aimed to prevent any asymmetries in the features of the model's face to influence choice behavior. The symmetric faces were created in Adobe Photoshop, by mirroring half of a face over the midline, and softening the break line; one consisting of twice the left half of the face (left-side composite), and one consisting of twice the right half (right-side composite).

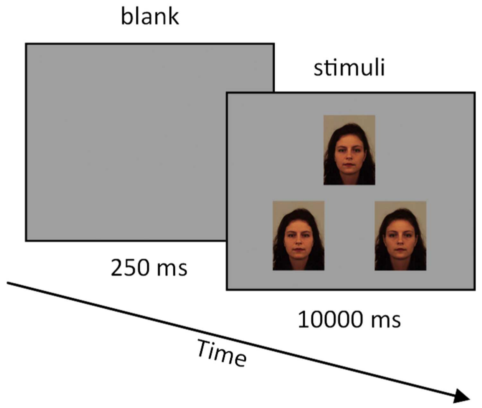

Figure 1. Timeline of a trial in the Face Similarity Task (FST).

4.1.1.3 Procedure. On each trial, a blank screen lasting 250 ms was followed by the stimulus consisting of three versions of the same face: the original (or mirrored) face at the top, and the symmetric versions at the lower left and lower right (Figure 1). The participants were instructed to indicate which of the two lower faces resembled the upper face most by pressing a corresponding button. In making this judgment, participants were asked to go with their first instinct, and to base their decision solely on the face of the person. The next trial started after the participant had made a response, or after a response period of 10 s (in 0.3% of trials no response was recorded). The pictures were shown in randomized order, and presented on a grey background. Symmetric left- and right-side composites were randomly presented at the left or right side of the screen.

Participants started the experimental session with a block of the FST, followed by the FET (see section 4.2), and another task including face stimuli that will not be described here. They concluded the test session, which lasted about 45 min in total, with a second block of the FST. Half of the participants saw the original symmetric faces in the first block and their mirror images in the second block, and vice versa for the other half of the participants.

4.1.1.4 Effects of interest. Following the original study, we computed a measure of LVF-bias by comparing the proportion of choice for the left-side composite in the block using the original face images, to the proportion of choice for the right-side composite in the block using its mirror images. Because one block used the original face images and the other used its mirror images, a choice for the left-side composite in one block and for the right-side composite in the other block is twice a choice for the same symmetric composite face. By making the comparison as we did (following C. Gilbert and Bakan, 1973), we controlled for participants' choosing a composite based on some specific feature that is present in the model's one face half. For example, a model may have a specific feature (e.g., a birthmark) on one of the sides of the face that is particularly striking to a participant and leads them to choose the composite containing it. In the block using mirrored images, this participant will then likely choose the same composite, containing the specific feature. However, if participants' choices are most strongly influenced by an LVF-bias in face perception, they will choose the composite face that reflects what they see on the left side of the face more frequently. Hence, the hypothesis was that the proportion choice for the left-side composite would be higher in the block using original faces images than the proportion choice for the right-side composite in the block using mirrored face images, indicating an LVF-bias.

Based on the original study's finding of an effect size of d_z = -.943 we had more than 99% power to detect this VFA with our sample size. No additional analyses were planned.

4.1.1.5 Differences with original study. Our version of the FST is a partial replication of C. Gilbert and Bakan's Experiment 4 from their 1973 paper, with differences pertaining to the stimulus set and testing procedure. The original study used 14 face pairs, of which printouts were presented to the participants. No details were provided about how participants were required to make their response, and how much time was allowed for this. Our replication attempt used 53 face pairs, which were digitally presented, with a maximum viewing time of 10 s. We used different face images than those used in the original study, but their symmetric versions were constructed in the same manner. In the original study, participants received the block using mirrored (original) images immediately following the block using the original (mirrored) images, while in the replication attempt these blocks were separated by two other tasks involving face stimuli.

The original study compared performance in left- and right-handed participants, finding a diminished LVF-bias for left-handed participants (C. Gilbert & Bakan, 1973). We tested only right-handed participants, and we thus relate our results to the right-handed group of the original study.

4.1.2 Results. We replicated the LVF-bias in the FST (BF_{10} = 5,858, $t[33] = -5.34$, $p < .001$, $d_z = -.916$). Participants more often judged the left-side composite face to resemble the original most in the block using the original face images (59%), than that they judged the right-side composite face to resemble the (mirrored) original most in the block using mirrored face images (47%) (mean choice for left-side composite over blocks = 56%, SD = 6.7%). Combining the original and replication studies' results, we found decisive evidence for the presence of an LVF-bias (BF_{10} = 189,722,311).

4.1.3 Discussion. Our replication attempt for the finding of a behavioral manifestation of lateralized face processing in the FST was successful. Specifically, we replicated the original study's LVF-bias, as participants more often chose the composite face that was constructed from the left half of the original face. When combining the original study's results and the results of our replication study in a meta-analytic Bayes factor, the evidence is decisive in demonstrating an LVF-bias in the FST. Furthermore, the effects in the original and replication studies were comparable in terms of direction and size, while the studies used different face images. This suggests that the likelihood of observing an LVF-bias for face processing in the FST is robust to different face images.

4.2 Face Emotionality Task (FET)

In 1983, Levy et al. devised a free-viewing face task using chimeric faces with half the face showing an emotional expression and the other half showing a neutral expression. This Face Emotionality Task (FET) is a widely used task to study lateralization of processing emotional expressions (e.g., Coronel & Federmeier, 2014; Innes, Burt, Birch, & Hausmann, 2016). We attempted to replicate Levy et al.'s 1983 study.

4.2.1 Methods.

4.2.1.1 Participants. The same thirty-four participants that completed the FST also performed the FET.

4.2.1.2 Stimuli. Images from the KDEF (Lundqvist et al., 1998) were adapted to form a set of 39 emotional chimeric faces; one half of the face showed an emotional expression, while the other half showed a neutral expression (T. Beking, personal communication, 2014). For each image, we created a version with the emotion showing in the left half of the face and a version with the emotion showing in the right half of the face (its mirror image). Twenty images showed the emotion happiness (10 female and 10 male models), and 19 images showed the emotion anger (10 female and 9 male models) in one half of the face.

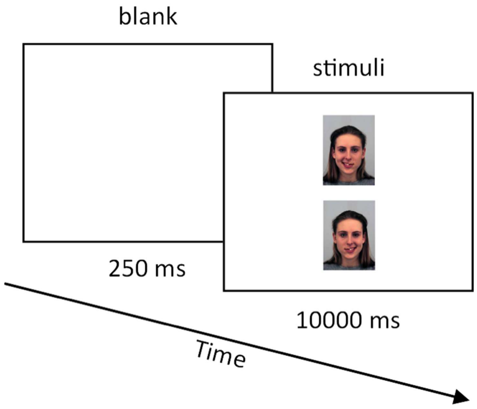

Figure 2. Timeline of a trial in the Face Emotionality Task (FET).

4.2.1.3 Procedure. On each trial, following a blank screen of 250 ms, the participant was shown an emotional chimeric face and its mirror image, one above the other (Figure 2). The participant was asked to indicate which of the two faces showed the strongest emotional expression, by pressing one of two buttons. The next trial started after the participants' response, or after 10 s (in 0.6% of the trials no response was recorded). The 39 stimuli were presented in randomized order, on a white background. The location of the face with the emotional expression on the left side was randomized over trials.

4.2.1.4 Effects of interest. The effect of interest was whether participants more often judged the face with the emotion on the left side as more emotional than the face with the emotion on the right side (i.e., LVF-bias). Based on the original study's effect size of d_z = -.689 for right-handed participants, we had 99% power to detect this VFA with our sample size. No additional analyses were planned.

4.2.1.5 Differences with original study. Our version of the FET is a partial replication of the study by Levy et al. (Levy et al., 1983), with differences pertaining to the stimuli and procedure. The original study used 36 pairs of 9 male actors showing the emotion 'happy', and the images were presented to the participants on slides. No details were provided with regard to response procedure, or how much time was allowed to make a response. The replication study used 39 pairs of male (19 items) and female (20 items) actors, showing the emotions 'happy' (20 items) or 'angry' (19 items), which were presented digitally. In the replication attempt we used different face images than those used in the original study. Furthermore, the original study allowed the response 'can't decide', while in the replication study we asked participants to make a choice on each trial. If a participant had not responded within 10 s, it was considered a miss and these trials were not included in our analysis.

The original study compared left- and right-handed participants, and found the left-handed participants to show a weaker LVF-bias (Levy et al., 1983). We tested only right-handed participants, and we accordingly relate our results to those of the right-handed participants in the slide presentation group of the original study.

4.2.2 Results. We replicated the LVF-bias in the FET (BF_{10} = 2,824, $t[33] = -5.07$, $p < .001$, $d_z = -.870$). Participants more often judged faces to have a stronger emotional expression when the left side expressed the emotion (bias = 65%, SD = 18%). When combining the effects found in the original and replication studies, there is decisive evidence for the presence of an LVF-bias (BF_{10} = 2.88647E+12).

4.2.3 Discussion. The results of this replication attempt were successful in replicating the original study's LVF-bias for emotional face processing. As was the case for the FST, the meta-analytic Bayes factor indicates that the evidence combined across the original and replication studies is decisive in demonstrating an LVF-bias in the FET. While the original study used only male faces with 'angry' expressions, we found highly similar results using male and female faces with angry and happy expressions. Accordingly, we can conclude that the LVF-bias observed in the FET is robust to different emotional expressions and the actors' sex.

4.3 Hierarchical Letter Task (HLT)

In 1979, Martin studied VFAs in processing the global and local elements present in so-called Navon letters. Using a Hierarchical Letter Task (HLT), she found an RVF-advantage for processing of local information, which was complemented by an LVF-advantage for processing of global information in a later study (Sergent, 1982). While these asymmetries have been replicated, there has also been a substantial number of studies that did not show a significant LVF-advantage for global processing and/or RVF-advantage for local processing (e.g., Boles, 1984; Boles & Karner, 1996; Van Kleeck, 1989). Discrepancies between these studies were argued to be due to differences in stimulus- and task-characteristics (Yovel et al., 2001). Yovel et al. addressed the influence of a number of stimulus and task factors on the surfacing of VFAs in ERs and RTs using an HLT. Their results showed that requiring participants to divide attention over equally salient local and global stimulus levels produced more robust VFAs than other versions of the HLT. Accordingly, we selected this improved paradigm (Yovel et al., 2001, Experiment 1C) for our replication attempt.

4.3.1 Methods.

4.3.1.1 Participants. Twenty-one participants (9 women) with a mean age of 20 years (range = 18-23) performed the HLT. The presented data are a subset of a larger data set (Brederoo et al., 2017).

4.3.1.2 Stimuli. Stimulus letters were T and H (targets), and Y and N (distractors). All stimuli were incongruent, that is, the identity of the letters presented at the global level always differed from that of the letters shown at the local level. The global stimulus was comprised of local stimuli placed within a 5 x 5 grid, with a global/local ratio of 0.14. The hierarchical letters were presented in black on a white background, and they subtended 3.5° of visual angle, with their inner edges positioned at 0.5° from the central fixation point. The mask consisted of a 5 x 5 grid of hash tags. During unilateral presentation blocks, one hierarchical letter was presented, in either the LVF or RVF. During bilateral presentation blocks, one hierarchical letter appeared in the LVF and another in the RVF, but only one of them contained the target.

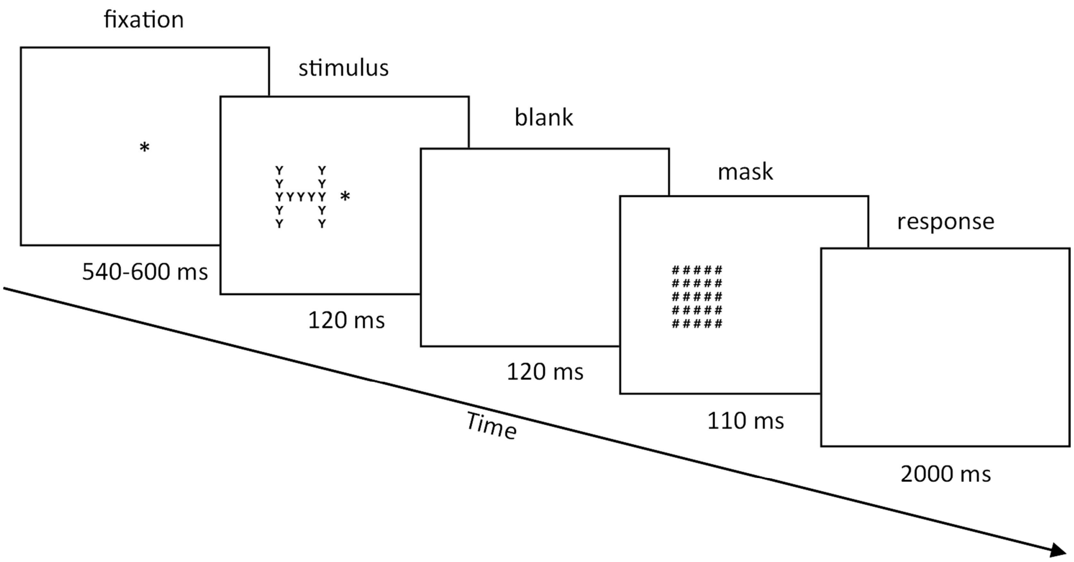

Figure 3. Timeline of a trial (unilateral presentation, global target) in the Hierarchical Letter Task (HLT).

4.3.1.3 Procedure. A trial started with the presentation of a central fixation asterisk that was present for a duration jittered between 540-600 ms (Figure 3). Next, a single stimulus was presented in the LVF or RVF, during unilateral presentation blocks, or two stimuli were presented, one in each visual field, during bilateral presentation blocks, for 120 ms. This display was followed by a blank screen of 120 ms during unilateral blocks and of 220 ms during bilateral blocks. After the blank, one or two masks were presented in place of the stimuli, for 110 ms. Participants were required to identify the target letter as quickly as possible, regardless of the level at which it appeared, or on which side it appeared. They did so by pressing one of two buttons using their index or middle finger. As in the original study, finger-response mapping and response hand were counterbalanced over participants. The next trial started after the participant had made a response, or after the response period of 2 s was over.

Participants completed four blocks of 80 trials, amounting to 320 experimental trials in total. They were allowed to take self-paced breaks between the blocks. Throughout the experiment, target letters appeared either at the global or the local level, of only one stimulus. In the first two blocks, unilateral stimuli were presented, while in the last two blocks bilateral stimuli were presented. Within blocks, the target appeared in the LVF and RVF equally often, and on the global and local level equally often, in a randomized manner. Before the start of the unilateral as well as the bilateral blocks, participants were given sixteen practice trials. Twelve of the participants completed 706 trials in a similar task using hierarchical figures, before starting this task. The results are no different for these participants than for the nine participants who only completed the HLT[2].

4.3.1.4 Effects of interest. The effects of interest were the RVF-local advantage in ERs (based on the original study's effect size of d_z = .716, we had 94% power to detect the effect with our sample size), and in RTs (original d_z = .557, 80% power), and the LVF-advantage for global processing in RTs (original d_z = -.835, 98% power).

[2] We checked whether the length of the task session affected the VFAs in an ANOVA. There showed to be no indication of this (Session Length x Level x Visual Field: $F[1,19] = .721$, $p = .406$ in ERs; and $F[1,19] = .147$, $p = .706$ in RTs).

4.3.1.5 Additional analyses. The only additional effect we examined was the LVF-advantage for global processing in ERs which was not found to be significant in the original study.

4.3.1.6 Differences with original study. Our version of the HLT is a partial replication of the original study (Yovel et al. 2001; experiment 1C), with slight changes regarding the stimuli and trial procedure. Specifically, we chose to replace the E and F of the original study by a T and H, because these are symmetric around the midline, thus preventing an asymmetric stimulus from causing different effects depending on the visual field of presentation. In the original experiment, level saliency of the stimuli was modulated by varying the global/local ratio (288 trials in total). As the equally salient stimuli were shown to produce more robust effects in the original study, we only used equally salient stimuli in our replication attempt (352 trials in total). In place of the manipulation of level saliency, we introduced two blocks using bilateral stimulus presentation, in addition to the unilateral stimulus presentation that the original study employed. We chose to include these blocks with bilateral stimuli because previous research (e.g., Boles, 1987) suggests that VFAs should be expected to be larger when both visual fields are stimulated. Thus, to increase our chance of producing VFAs with the HLT, we additionally included bilateral trials. Furthermore, the original study reported to have placed the local elements in a 3 x 5 grid, but we chose a 5 x 5 grid, because the N and Y could not be produced in a 3 x 5 grid. The original study used a 9 x 8 grid of small letters as a mask, but since no information was provided about the identity of the letters used for the mask, we used a 5 x 5 grid of hash tags. In the original study, the stimulus duration was 100 ms, and the duration of the mask was 1000 ms. Based on a pilot study we changed the durations of the stimuli and masks (see section 4.3.1.3).

4.3.2 Results. As predicted, the VFAs were present during both unilateral and bilateral presentation blocks, but they were larger during bilateral presentation than during unilateral presentation (see Appendix A). To assess our success of replication, in the following analyses we take into account all trials, as this gives us the greatest degree of power to detect the VFAs.

4.3.2.1 Effects of interest. We replicated the RVF-advantage in local processing in ERs (BF_{10} = 26.8, $t[20] = 3.36$, $p = .002$, $d_z = .733$) (LVF: 27%, SD = 17%; RVF: 21%, SD = 18%), and in RTs (BF_{10} = 8.39, $t[20] = 2.75$, $p = .006$, d_z = .600) (LVF: 845 ms, SD = 134 ms; RVF: 805 ms, SD = 137 ms). We replicated the LVF-advantage in global processing in RTs (BF_{10} = 409, $t[20] = -4.69$, $p < .001$, d_z = -1.023) (LVF: 741 ms, SD = 115 ms; RVF: 809 ms, SD = 150 ms).

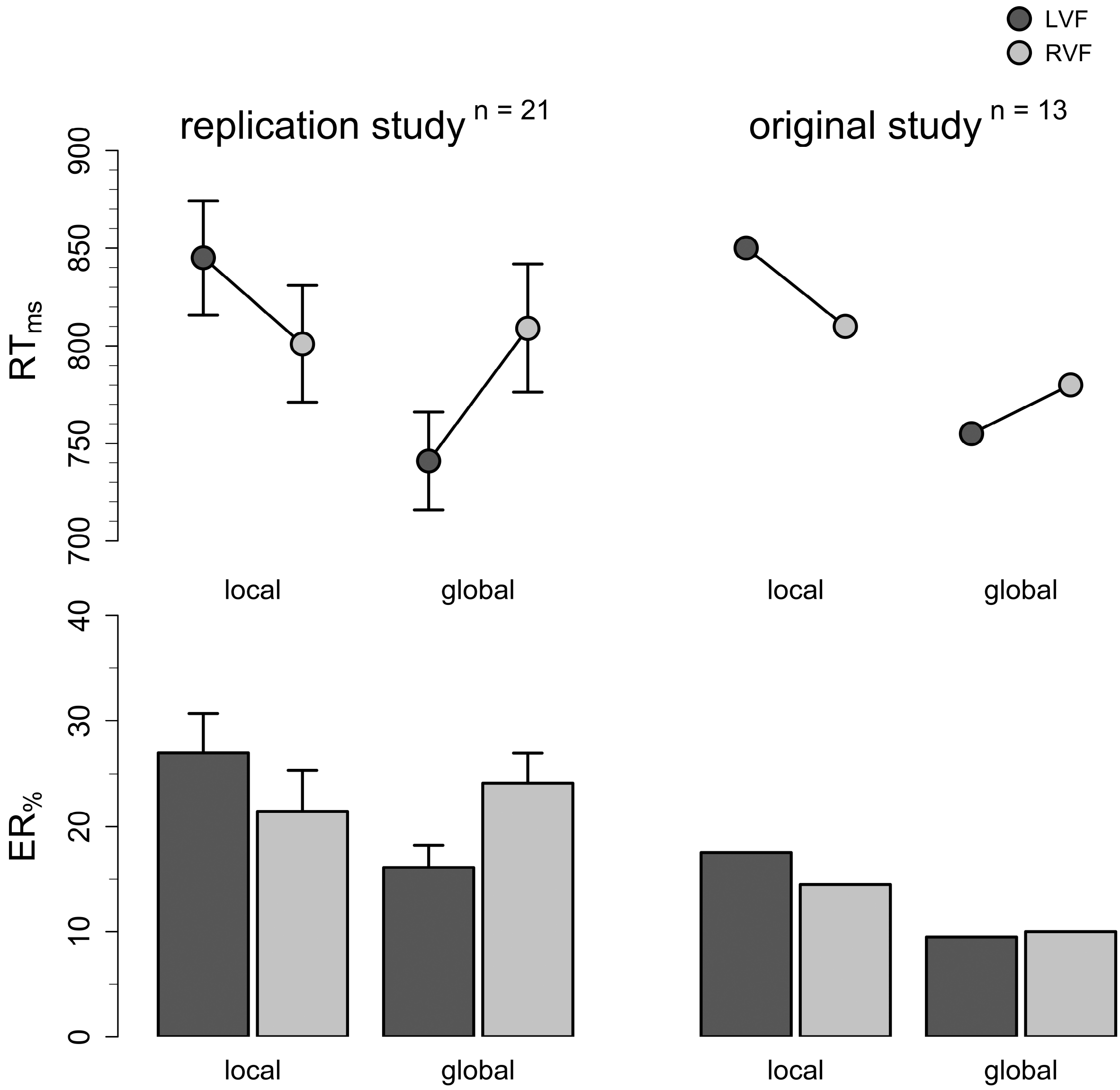

Figure 4. Error rates (lower panels) and reaction times (upper panels) of the replication (left panels) and original (right panels) studies' Hierarchical Letter Task. The means of the original study are estimated from the bottom-left panel of Figure 5 in Yovel et al. (2001, p. 1375). Error bars represent standard errors of the means

4.3.2.2 Additional analyses. In ERs, we found substantial evidence for an LVF-advantage in global processing (BF_{10} = 237, t[20] = -4.43, p < .001, d_z = -.967) (LVF: 16%, SD = 9.7%; RVF: 24%, SD = 13%).

4.3.2.3 Combined evidence. When combining the results of the original and replication studies, there is decisive evidence for the presence of an RVF-advantage for local processing in ERs (BF_{10} = 329) and for the presence of an LVF-advantage for global processing in RTs (BF_{10} = 10,124). There is very strong evidence with regard to the RVF-advantage for local processing in RTs (BF_{10} = 40.7).

4.3.3 Discussion. The outcome of this replication attempt of the HLT was successful as it yielded the expected behavioral manifestations of lateralized processing of global and local information. Specifically, our results were similar to those of the original study, in showing an RVF-advantage for local processing in both ERs and RTs, and in showing an LVF-advantage for global processing surfacing in RTs, and additionally in ERs. Accordingly, the meta-analytic Bayes factor also yielded strong support the presence of an RVF-advantage for local processing and an LVF-advantage for global processing, as measured with the HLT. It is of further interest that, in line with predictions (Boles, 1987; Hunter & Brysbaert, 2008), the VFAs were larger during the bilateral than the unilateral presentation blocks.

4.4 Picture Matching Task (PMT)

The idea that the two hemispheres differentially process high spatial frequencies (HSF) and low spatial frequencies (LSF) was first put forward by Sergent (1982), who used the results in an HLT (see section 4.3.1 for task description) to arrive at these conclusions. In 1992, Kitterle, Hellige, and Christman more directly tested the role of spatial frequencies by assessing VFAs in response to gratings, and reported that HSF gratings were more easily classified when presented in the RVF, whereas LSF gratings were more easily classified when presented in the LVF. As pointed out by Peyrin et al. (2003), much of the theory regarding lateralization of spatial frequency processing was based on studies using hierarchical stimuli, rather than on studies that explicitly demonstrated differing VFAs by manipulating the spatial frequency content of stimuli. One exception is the study by Kitterle et al. (1992), which used gratings to show an LVF-advantage for LSF processing and an RVF-advantage for HSF processing. However, these VFAs were found in only one of four task conditions, and the study used a sample of only 5 participants. Peyrin et al. (Peyrin, Chauvin, Chokron, & Marendaz, 2003) introduced a Picture Matching Task (PMT) in which more complex stimuli were used than the gratings used by Kitterle et al. (1992). Using unfiltered and filtered images of natural scenes, Peyrin and colleagues successfully produced LVF-advantages for LSF processing and RVF-advantages for HSF processing (Peyrin et al., 2006, 2003). In addition, Peyrin, Mermillod, et al. (2006) showed that the time allowed for processing of the filtered stimuli affected the surfacing of the VFAs. Acknowledging the importance of processing time as a potential modulator of VFAs in spatial frequency processing, we attempted to replicate the 2006 study of Peyrin, Mermillod, and colleagues.

4.4.1 Methods.

4.4.1.1 Participants. Thirty-one participants (15 women) performed the PMT. Their mean age was 21 years (range = 18-25).

4.4.1.2 Stimuli. The stimulus set comprised four black-and-white images of natural scenes (a city, a highway, a beach, and a mountain), two filtered versions of each of these images, and a backward mask. The HSF filtered images were created using a high-pass filter with a cut-off of 24 cycles per filter. The LSF filtered images were created using a low-pass filter with cut-off of 16 cycles per image. The size of the images was 4.8° x 4.8° of visual angle, and they were presented on a grey background at either the center of the screen, in the LVF, or RVF. When presented in the LVF or RVF, the inner edge of the image was positioned at a distance of 2° from the center. The mask contained a mean frequency typical of that of the set of natural scene images from which the stimuli had been selected (see Peyrin, Mermillod, et al., 2006).

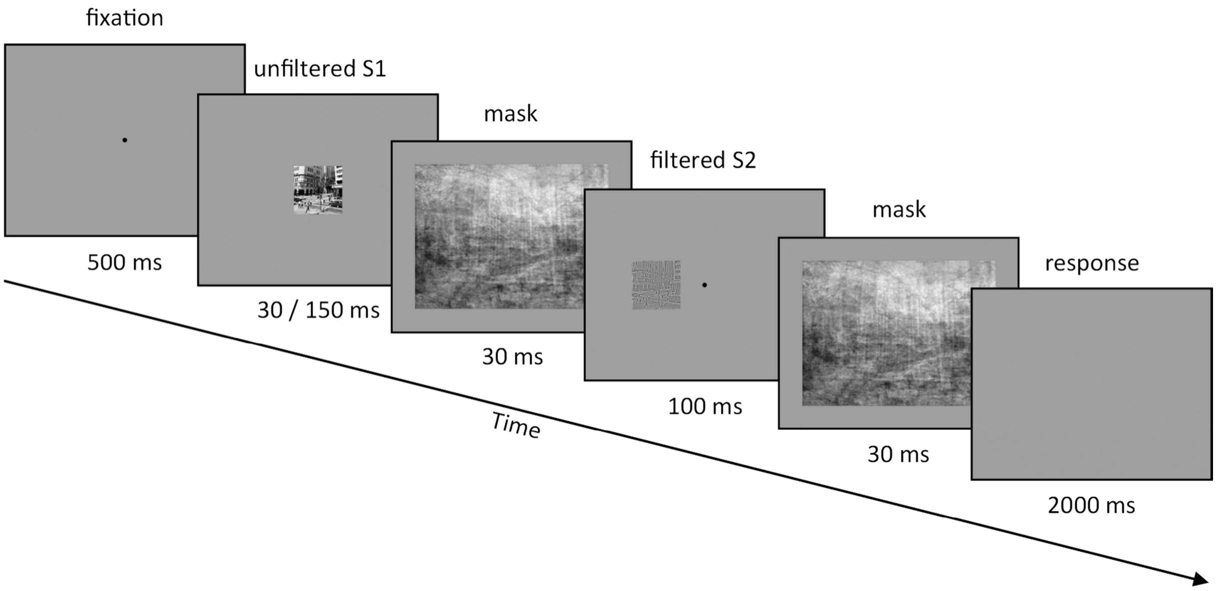

Figure 5. Timeline of a trial (high spatial frequency S2) in the Picture Matching Task (PMT).

4.4.1.3 Procedure. Each trial began with a centrally presented fixation point for 500 ms (Figure 5). Subsequently, one of the four unfiltered images (S1) was presented centrally. The S1 was presented for 30 ms or 150 ms, after which it was replaced by the mask, which remained on the screen for 30 ms. Immediately following the mask, a second image (S2) was presented for 100 ms. The S2 could be either an HSF or LSF filtered image of the S1, or of one of the other images, and was presented in the LVF or RVF. After 100 ms, the mask replaced the S2 and it was again shown for 30 ms. From the offset, participants had 2 s to indicate whether the S2 depicted the same natural scene as the S1. They did so by pressing two buttons simultaneously with their index fingers each time when they detected a match, as quickly as possible. They were instructed not to press any buttons on no-match trials. After the 2 s response interval the next trial started automatically.

The participants started the task with three practice blocks. First, they performed 32 trials in which the S2, like the S1, was an unfiltered image and presented centrally. Next, they performed 64 trials in which the S2 appeared either in the LVF or RVF, but was still an unfiltered image. The final practice block consisted of 64 trials during which the S2 again always appeared in the center of the screen, but was either an HSF or LSF filtered image[3]. After the practice blocks, participants completed four experimental blocks of 64 trials in each of the S1 duration conditions, with self-paced breaks between blocks. Within each block, HSF and LSF trials, and match and non-match trials, occurred equally often, and both types of trials were randomized. Half of the participants started with the 30 ms condition, followed by the 150 ms condition, and vice versa for the other half.

4.4.1.4 Effects of interest. The four effects of interest all pertained to RTs. Specifically, the original study showed a LVF-advantage for LSF trials regardless of S1 duration (based on the original study's effect size of d_z = -1.06, we had more than 99% power to detect the effect with our sample size), and this effect was also found to be significant for each duration condition (short duration: d_z = -1.20, more than 99% power; long duration: d_z = -.647, 97% power). In addition, the original study found a significant RVF-advantage for HSF trials in the short duration condition only (original d_z = .615, 96% power).

4.4.1.5 Additional analyses. In addition to examining the above-mentioned effects of interest, we also analyzed the RVF-advantage for HSF trials in the long duration condition, and we also tested the significance of this VFA averaged across

[3] The original article states that the total practice procedure consisted of eight trials with unfiltered images (Peyrin, Mermillod, et al., 2006). The practice procedure as adopted for this replication, however, is copied from the original experiment E-Prime file, as shared with us by the main author of the study, who confirmed that this in fact was the practice procedure used in the experiment described in the 2006 publication.

the two duration conditions, in RTs. Furthermore, we analyzed each of the six effects' counterparts in ERs.

4.4.1.6 Differences with original study. The PMT is a full replication of the original study (Peyrin, Mermillod, et al., 2006), as the first author of the original study shared the experiment E-Prime file and stimulus image files, which we adjusted for Dutch participants (the original included French instructions). The only difference between the original study and our replication study concerned the number of trials. In the original study, participants completed 256 trials in total. Because of the use of a go/no-go procedure, this amounted to 16 trials per condition for analysis. In our replication experiment, we chose to double the number of trials (Brysbaert & Stevens, 2018).

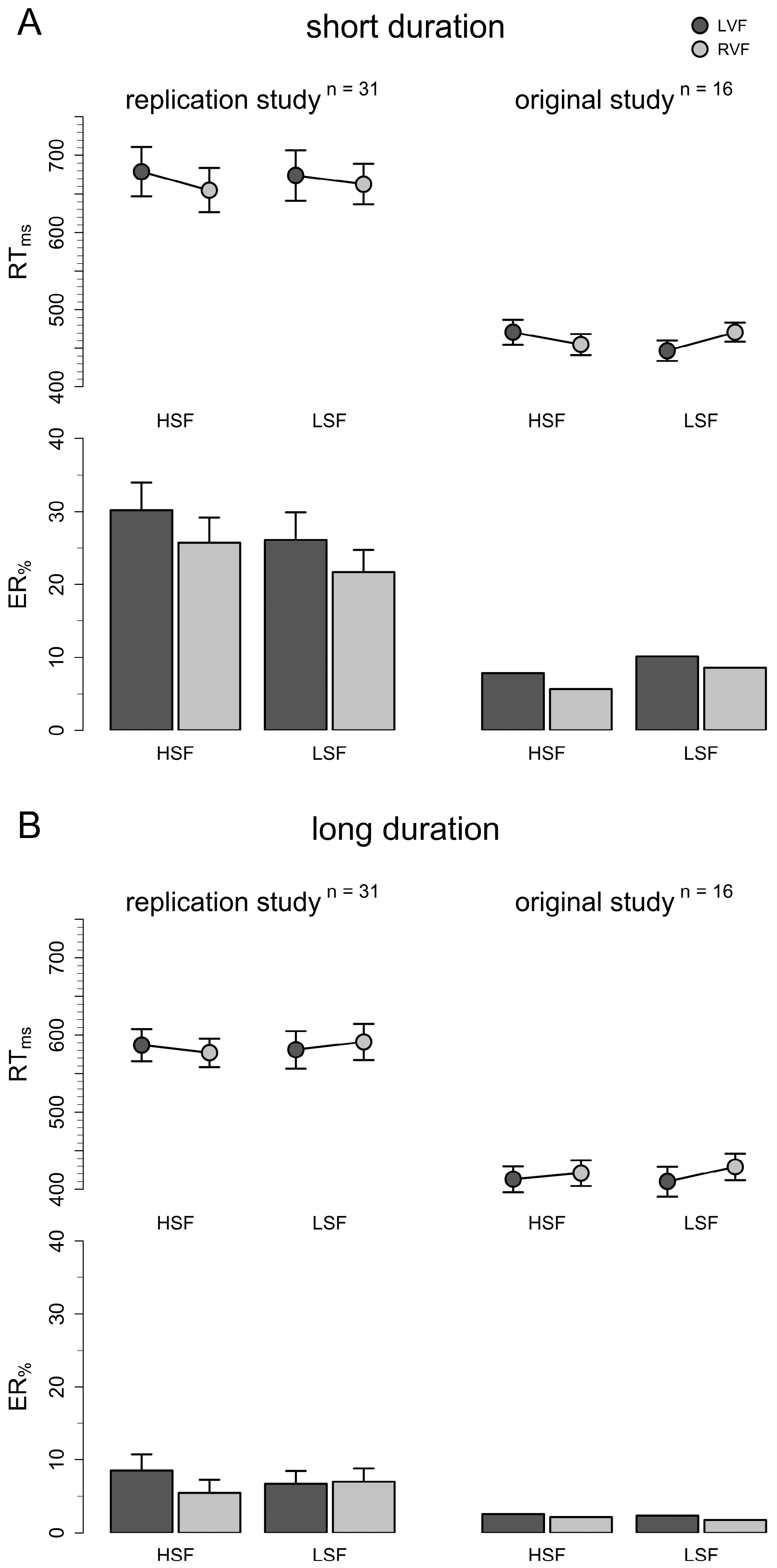
A
short duration
LVF
RVF
replication study n = 31
original study n = 16
RTms
700
600
500
400
HSF
LSF
HSF
LSF
ER%
40
30
20
10
0
HSF
LSF
HSF
LSF
B
long duration
replication study n = 31
original study n = 16
RTms
700
600
500
400
HSF
LSF
HSF
LSF
ER%
40
30
20
10
0
HSF
LSF
HSF
LSF

Figure 6. Error rates (lower panels) and reaction times (upper panels) of the replication (left panels) and original (right panels) studies' Picture Matching Task, of the results in the short S1 duration (30 ms) condition (A), and the results in the long S1 duration (150 ms) condition (B). The means of the original study are copied from Table 1 of Peyrin, Mermillod, et al. (2006, p. 218). Error bars represent standard errors of the means.

4.4.2 Results.

4.4.2.1 Effects of interest. We failed to replicate the LVF-advantage for LSF images in the short duration condition in RTs, indicated by substantial evidence against its presence in our data (BF_{10} = .116, *t*[30] = .78, *p* = .779, d_z = .140) (LVF: 674 ms, SD = 182 ms; RVF: 663 ms, SD = 146 ms). For the long duration condition, our results were inconclusive with regard to the presence of this VFA (BF_{10} = .594, *t*[30] = -1.14, *p* = .132, d_z = -.205) (LVF: 581 ms, SD = 135 ms; RVF: 591 ms, SD = 131 ms), and the average across duration conditions likewise failed to produce convincing evidence for this VFA (BF_{10} = .789, *t*[30] = -1.36, *p* = .093, d_z = -.243) (LVF: 613 ms, SD = 120 ms; RVF: 622 ms, SD = 122 ms).

There was also indecisive evidence with regard to the RVF-advantage for HSF image processing in the short duration condition in RTs (BF_{10} = 1.30, *t*[30] = 1.70, *p* = .050, d_z = .305) (LVF: 679 ms, SD = 178 ms; RVF: 655 ms, SD = 159 ms).

4.4.2.2 Additional analyses. We did not find conclusive support for the presence of an RVF-advantage for HSF images in RTs, when combining the short and the long S1 conditions (BF_{10} = 3.03, *t*[30] = 2.20, *p* = .018, d_z = .395) (LVF: 623 ms, SD = 121 ms; RVF: 607 ms, SD = 110 ms), or when considering the long duration condition only (BF_{10} = .744, *t*[30] = 1.31, *p* = .100, d_z = .236) (LVF: 587 ms, SD = 117 ms; RVF: 577 ms, SD = 102 ms).

In the ER data of the replication study we found substantial evidence for an RVF-advantage for HSF image processing when combining the short and the long S1 duration conditions (BF_{10} = 5.97, t[30] = 2.56, p = .008, d_z = .460) (LVF: 19%, SD = 14%; RVF: 16%, SD = 12%), as well as in the long duration condition only (BF_{10} = 4.83, t[30] = 2.45, p = .010, d_z = .440) (LVF: 8.5%, SD = 13%; RVF: 5.5%, SD = 9.9%). In the short duration condition alone, the evidence for this VFA was inconclusive (BF_{10} = 1.29, t[30] = 1.69, p = .050, d_z = .304) (LVF: 30%, SD = 21%; RVF: 26%, SD = 19%). With regard to the LVF-advantages for LSF image processing in ERs, we found substantial evidence against the presence of this VFA when combining the short and the long S1 duration conditions (BF_{10} = .087, t[30] = 1.40, p = .915, d_z = .251) (LVF: 16%, SD = 12%; RVF: 14%, SD = 11%), in the short duration condition only (BF_{10} = .081, t[30] = 1.58, p = .938, d_z = .284) (LVF: 26%, SD = 21%; RVF: 22%, SD = 17%), and in the long duration condition only (BF_{10} = .230, t[30] = -.23, p = .411, d_z = -.041) (LVF: 6.7%, SD = 9.7%; RVF: 7.0%, SD = 10%).

4.4.2.3 Combined evidence. When combining the original and replication results, there is substantial evidence for an RVF-advantage for HSF processing in RTs in the short duration condition (BF_{10} = 9.04), but substantial evidence against the presence of this VFA in the long duration condition (BF_{10} = .230).

Combining the original and replication results further shows there to be strong evidence for the presence of an LVF-advantage for LSF processing in RTs (BF_{10} = 19.3), substantial evidence for this VFA in the long duration condition alone (BF_{10} = 3.52), and inconclusive evidence for this VFA in the short duration condition alone (BF_{10} = .592).

4.4.3 Discussion. We were not successful in replicating the expected VFAs for processing of high and low spatial frequencies using the task that was introduced by Peyrin et al. (2006). However, two LVF-advantages for LSF processing and one RVF-advantage for HSF processing were in the expected direction, and combining the evidence for these VFAs in meta-analytical Bayes factors (Rouder & Morey, 2011) resulted in at least substantial evidence for their presence. We additionally found evidence for an RVF-advantage for HSF processing that was not predicted based on the original study's results (Peyrin, Mermillod, et al., 2006), but could be expected based on the theory regarding lateralization of spatial frequency information.

Given the large difference between the original study's and replication study's effect sizes, and the larger error margin on the former than the latter, it seems likely that the effects in the original studies were an overestimation of the true effect sizes, which is not an uncommon problem in replication research (Anderson & Maxwell, 2015). Consequently, while the effects may in fact have been present, our study may not have had enough power to detect them. Furthermore, the notion that the LH is specialized in processing HSF information while the RH is specialized in processing LSF information is supported by neuroimaging data (for a review, see Kauffmann, Ramanoël, & Peyrin, 2014), which suggests that behavioral methods may be less sensitive to measure lateralized processing of this type of visual information, especially with a limited sample size.

4.5 Color Oddball Task (COT)

Using an oddball task, A.L. Gilbert et al. (2006) showed that participants were faster to detect colored targets when these had different color names than the distractors, supporting the notion of categorical perception for colors. Importantly, they found that this effect was only present for targets presented in the RVF. In contrast, participants were faster to detect colored targets that had the same name as the distractors, when these were presented in the LVF compared to the RVF. The authors concluded that language affects visual processing of colors in the RVF, but not in the LVF, and called this the 'lateralized Whorf effect'. Since the appearance of this paper, many more publications have followed, supporting and extending this finding (e.g., Daoutis, Pilling, & Davies, 2006; Drivonikou et al., 2007; Siok et al., 2009; but see Brown, Lindsey, & Guckes, 2011; Witzel & Gegenfurtner, 2011), but often using different tasks. We attempted to replicate the Color Oddball Task (COT) described in the original study of A.L. Gilbert et al. (2006; Experiment 2 (no-interference block)).

4.5.1 Methods.

4.5.1.1 Participants. Thirty-two participants (17 women) performed the COT. All participants had normal color vision, and their native language was either Dutch or German. Mean age was 20 years (range = 18-25).

4.5.1.2 Stimuli. The stimulus colors were chosen to resemble those used by A.L. Gilbert et al. (2006). We used two shades of green (G1 and G2), and two shades of blue (B1 and B2). The interstimulus distances in CIEL*a*b* space were ΔE = 4.6 for the G1-G2 pair, ΔE = 3.6 for the G2-B1 pair, and ΔE = 5 for the B1-B2 pair. A stimulus array consisted of a ring with a diameter of 8.5° of visual angle, of twelve 1° colored circles, presented on a grey background. Eleven of these circles had the same color, and one circle, the oddball, was colored differently. The oddball could appear in one of eight positions; four on the left and four on the right side of the ring. The two uppermost and two lowermost circles were never oddballs. The color of the oddball was either from the same category as the distractors (i.e., G1-G2, or B1-B2), or from a different category (i.e., G1-B1, G2-B1, or G2-B2).

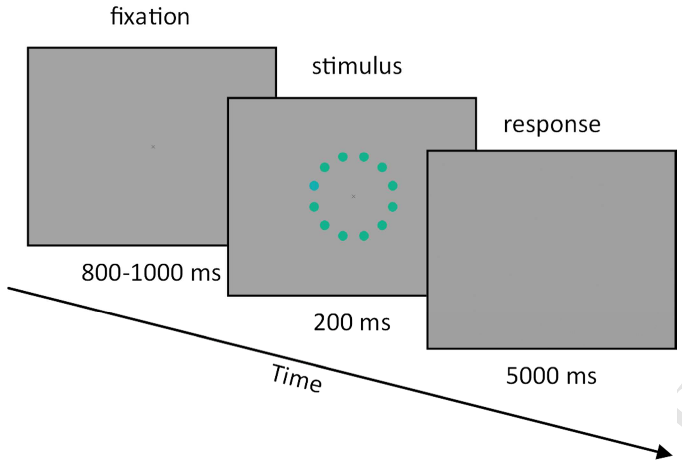

Figure 7. Timeline of a trial (between-category) in the Color Oddball Task.

4.5.1.3 Procedure. Each trial started with the presentation of a fixation cross, with presentation duration jittered between 800-1000 ms (Figure 7). With the fixation cross remaining on screen, the stimulus ring was presented for 200 ms. Next, a blank screen was presented during which participants could make their response; a left index finger button press if the oddball had appeared on the left side of the ring, and a right index finger button press if it had appeared on the right side of the ring. Participants were asked to respond as fast and accurately as possible. The next trial started after the participants' response, or after 5 s if no response was made.

Each of the oddball-distractor combinations and oddball-positions occurred equally often. Participants completed four blocks of 80 trials, and were allowed to take self-paced breaks between blocks. The experimental session started with a naming task to establish participants' green-blue lexical boundary, on which inclusion of their data in the analyses was based. In this task, one circle was presented centrally on a grey background, for 200 ms. Each of the four possible colors (G1, G2, B1 and B2) was presented ten times, in a randomized order. Participants were asked on each trial to indicate whether the colored circle had been green or blue, by pressing the G-key or B-key on a QWERTY-keyboard. They were not required to respond as fast as possible, but were encouraged to go with their first intuition. The lexical green-blue boundary was defined as the estimated value where blue would be reported half of the time. After the naming task, the participants were given 32 practice trials in the COT before the experimental trials started. Participants received all instructions in their native language.

Sixteen of the participants completed the SOT (described in section 4.6), before starting the COT, and vice versa for the other 16.

4.5.1.4 Effects of interest. The effects of interest were the RVF-advantage for between-category discrimination in RTs (based on the original study's (no-interference blocks) effect size of d_z = .742, we had 99% power to detect the effect with our sample size[4]), and the LVF-advantage for within-category discrimination in RTs (original d_z = -.684, power 97%).

4.5.1.5 Additional analyses. Additionally, we analyzed the two effects' counterparts in ERs.

4.5.1.6 Differences with original study. The COT is a partial replication of A.L. Gilbert et al.'s (2006) Experiment 2 (no-interference block). The replication experiment differs from the original study on a number of aspects. Firstly, the appearance of the stimuli in the replication study was not identical to that in the original study. Because A.L. Gilbert et al. did not report the specific color values in a way that makes them reproducible, the specific colors of the stimuli used in the replication experiment were likely different from the original color values. Furthermore, in the original study, the stimulus ring consisted of colored squares. However, since using squares leads to differences in the distance from the center to

[4] With the exclusion of the four participants who failed to put the naming boundary between G2 and B1 (see section 4.5.2).

the inner edge of the stimulus depending on its position in the ring, we chose to use colored circles instead. Another possible difference with regard to the appearance of the stimuli is the size of the stimulus ring. Because A.L. Gilbert et al. do not report on its size in their 2006 paper, we chose to use the ring size that they report in their 2008 paper (A.L. Gilbert et al., 2008) on a variation of the oddball task using shapes (see section 4.6 on the SOT)[5].

Secondly, in the original study participants completed on average 500 trials in an oddball task (250 of which in the no-interference block), and in the replication study participants completed on average 560 trials in an oddball task (320 of which in the COT). The exact number of trials depended on the order of the three task conditions in the original study, and on the ordering of the COT and SOT tasks[6] in the replication study.

Thirdly, we excluded the two uppermost and two lowermost positions in the ring as potential oddball positions. As it has been suggested that a strip of 1-3° along the vertical meridian of the visual field is projected bilaterally, rather than in a lateralized fashion (Bunt & Minckler, 1977; Jordan & Paterson, 2009; but see Ellis & Brysbaert, 2010), the stimuli in these positions are likely projected to both hemispheres. The COT is used with the assumption that the oddballs are projected to the hemispheres contralateral to the visual fields, rendering the uppermost and lowermost positions unsuitable oddball locations.

In the original study, the authors showed that a verbal interference task could disrupt the surfacing of VFAs. As the focus of these replication studies lies in the

[5] Pilot studies showed that varying the size of the stimulus ring did not affect the presence of the VFAs.
[6] As a first analysis step we checked whether in our experiment task order (COT or SOT first) affected the VFAs in an ANOVA. There showed to be no indication of this (Task Order x Color Pair x Visual Field: $F[1,26] = .32$, $p = .576$ in ERs; and $F[1,26] = .44$, $p = .513$ in RTs).

reproducibility of VFAs, we did not use any interference tasks, and relate our results only to the no-interference block of the original study.

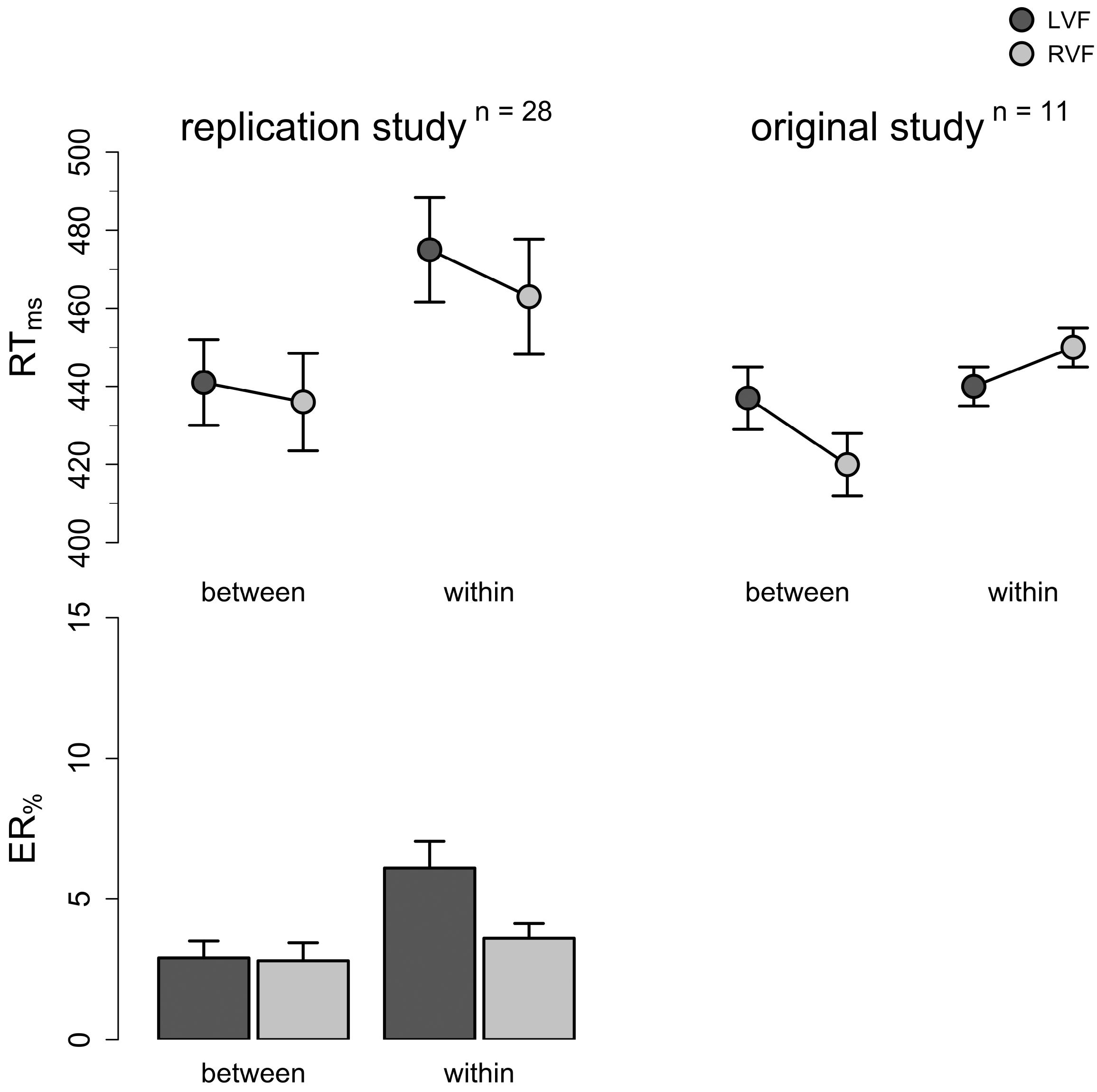

Figure 8. *Error rates (lower panel) and reaction times (upper panels) of the replication (left panels) and original (right panel) studies' Color Oddball Task. The mean reaction times of the original study are estimated from Figure 2B in A.L. Gilbert et al. (2006, p. 491) (error rates were not reported for each of the conditions). Error bars represent standard errors of the means.*

4.5.2 Results. Three of the participants put the blue-green boundary between G1 and G2, and one put it between B1 and B2, and these participants' data were excluded from analyses. The remaining 28 participants (15 women) put the blue-green boundary between G2 and B1.

4.5.2.1 Effects of interest. We failed to replicate the LVF-advantage for within-category discrimination, as the evidence against the presence of this VFA was substantial (BF_{10} = .077, *t*[27] = 1.89, *p* = .965, d_z = .357) (LVF: 475 ms, SD = 71 ms; RVF: 463 ms, SD = 78 ms). There was inconclusive evidence with regard to the RVF-advantage for between-category discrimination in RTs (BF_{10} = .490, *t*[27] = .95, *p* = .176, d_z = .179) (LVF: 441 ms, SD = 58 ms; RVF: 436 ms, SD = 66 ms).

4.5.2.2 Additional analyses. In ERs, we found substantial evidence against the presence of an RVF-advantage for between-category discrimination (BF_{10} = .224, *t*[27] = .14, *p* = .445, d_z = .027) (LVF: 2.9%, SD = 3.2%; RVF: 2.8%, SD = 3.4%), and we found strong evidence against the presence of an LVF-advantage for within-category discrimination (BF_{10} = .054, *t*[27] = 3.41, *p* = .999, d_z = .644) (LVF: 6.1%, SD = 5.0%; RVF: 3.6%, SD = 2.8%). The latter effect indicates that participants in fact seemed to perform better on within-category discrimination when the oddball was presented in the RVF than when it was presented in the LVF.

4.5.2.3 Combined evidence. When combining the results of the original and replication studies, there is substantial evidence against the presence of an LVF-advantage for within-category discrimination (BF_{10} = .125), and inconclusive evidence regarding the presence of an RVF-advantage for between-category discrimination (BF_{10} = 1.87), in RTs.

4.5.3 Discussion. We did not succeed in replicating the lateralized Whorf effect for color perception. The combined evidence of the original and replication studies indicates evidence against the presence of an LVF-advantage for within-category discrimination, and the combined evidence is inconclusive with regard to the

RVF-advantage for between-category discrimination. We will now consider whether the differences between the original and replication studies could account for the differences in results.

Although the original (A.L. Gilbert et al., 2006) and replication studies differed in the specific color values used, we do not consider this difference detrimental to the comparability of the results. First, while the exact color values may differ, we used the same selection criteria as the original study did, resulting in a color set with similar characteristics. Specifically, like in the original study, the colors formed two within-category pairs and one between-category pair, as confirmed by the naming task. Additionally, like in the original study, the colors' interstimulus distances in CIEL*a*b* space were larger for the within-category pairs than for the between-category pair, and participants' performance was worse for within-category discrimination than for between-category discrimination (see Appendix A). Second, the lateralized Whorf effect has previously been found with different color sets (Drivonikou et al., 2007; Roberson, Pak, & Hanley, 2008), and even with stimuli outside the color domain (A.L. Gilbert et al., 2008, see also section 4.6 on the SOT). In sum, we do not consider differences in the exact color values to be a potential cause for the differences in results between the original and replication studies. While the shape of the stimuli may have affected their processing, we would expect this to have been the same for stimuli in the LVF and RVF, and as such not to have affected the VFAs.

The original study used on average 12% more trials than the replication study (COT and SOT combined). However, we showed that in the replication study, the order of the tasks, and hence, the number of completed trials in an oddball task, did not affect the VFAs for categorical color perception. A.L. Gilbert et al. (2006) also do not report that the order of tasks affected the VFAs in their experiment. We, therefore, do not regard this difference in the number of trials as a potential explanation for the differing results between the original and replication studies.

In conclusion, rather than lateralized categorical color perception, our replication study supports a general RVF-advantage for color discrimination (see also Appendix A). This RVF-advantage is, contrary to predictions, larger for within-category discrimination than for between-category discrimination, in ERs. This is in direct contrast to the results of A.L. Gilbert et al. (2006), who report an LVF-advantage for within-category discrimination, and has certain implications for the likeliness that there is lateralization of the influence of color categories on color discrimination. We return to this point in section 5.3.

4.6 Shape Oddball Task (SOT)

In 2008, A.L. Gilbert et al. showed that the lateralized Whorf effect generalized to stimuli other than colors. Specifically, they showed that categorical perception of cat and dog shapes produced an RVF-advantage for between-category discrimination, and an LVF-advantage for within-category discrimination. This study has been less influential than A.L. Gilbert et al.'s 2006 study, but given the important theoretical implications of a lateralized Whorf effect beyond the realm of colors, we chose to also attempt to replicate the Shape Oddball Task (SOT) (A.L. Gilbert et al., 2008; Experiment 1).

4.6.1 Methods.

4.6.1.1 Participants. Twenty-seven participants (14 women) who performed the COT also completed the SOT. Their mean age was 20 years (range = 18-25).

4.6.1.2 Stimuli. The two black shapes of dog figures and two black shapes of cat figures as used in the original study of A.L. Gilbert et al. (2008) were copied from their article. The radius of the stimulus ring was 8.5° of visual angle.

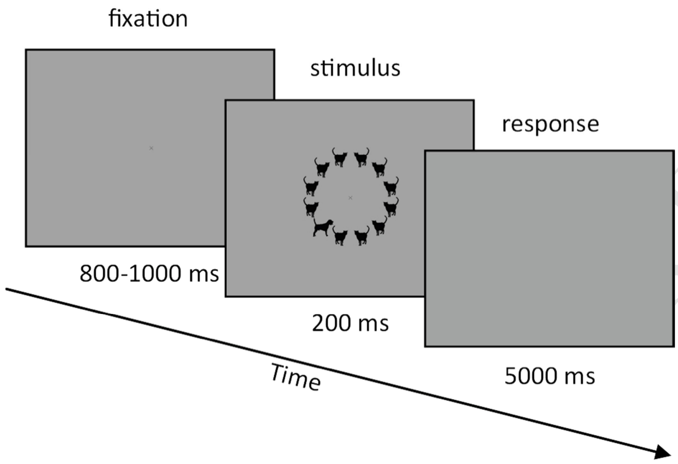

Figure 9. Timeline of a trial (between-category) in the Shape Oddball Task.

4.6.1.3 Procedure. The procedure of the SOT is identical to that of the COT (see section 4.5) (Figure 9), with two exceptions. Firstly, the SOT used animal shapes instead of colored circles. All the stimulus shapes faced the center of the screen. There were four between-category pairs (combining a dog shape with a cat shape) and two within-category pairs (combining two dog shapes or two cat shapes). Secondly, participants completed 384 trials, which were divided over 4 blocks of 96 trials, and preceded by 32 practice trials. Participants were allowed to take self-paced breaks between blocks.

Thirteen of the participants completed the COT (see section 4.5) before starting the SOT, and vice versa for the other 14 participants.

4.6.1.4 Effects of interest. The effects of interest were the RVF-advantage for between-category discrimination in RTs (based on the original study's effect size of

d_z = .525, we had 83% power to detect the effect with our sample size[7]), and the LVF-advantage for within-category discrimination in RTs (original d_z = -.6, 91% power).

4.6.1.5 Additional analyses. We additionally investigated the effects of interests' counterparts in ERs.

4.6.1.6 Differences with original study. In the original study, the next trial would only start after the participant had made a response, but in the replication study we limited response time to 5 s (in 0.6% of the trials no response was recorded). Like in the COT, we did not use the two uppermost and lowermost positions in the ring as potential oddball locations. Finally, the original study used 864 trials, while participants in the replication study performed 585 trials on average in an oddball task (depending on whether they had started with the COT or the SOT[8], see section 4.5 on the COT).

[7] With the exclusion of the one participant who failed to put the naming boundary between cats and dogs (see section 4.6.2).

[8] Like for the COT, as a first analysis step we checked whether task order (COT or SOT first) affected the VFAs in an ANOVA. There showed to be no indication for this (Task Order x Color Pair x Visual Field: F[1,24] = .02, p = .878 in ERs; and F[1,24] = 1.84, p = .188 in RTs).

Figure10. Error rates (lower panel) and reaction times (upper panels) of the replication (left panels) and original (right panel) studies' Shape Oddball Task. The mean reaction times of the original study are estimated from the bottom-left panel of Figure 4 in A.L. Gilbert et al. (2008, p. 93) (error rates were not reported for each of the conditions). Error bars represent standard errors of the means.

4.6.2 Results. One participant regarded one cat shape as a dog, and this participant's data were excluded from the analyses. The remaining 26 participants (13 women) correctly categorized the cat and dog shapes.

4.6.2.1 Effects of interest. In RTs, we failed to replicate the LVF-advantage for within-category discrimination, finding substantial evidence against the presence of this VFA (BF_{10} = .126, *t*[25] = .79, *p* = .782, d_z = .155) (LVF: 758 ms, SD = 207 ms; RVF: 748 ms, SD = 206 ms). Our data do not allow a conclusion to be drawn regarding the RVF-advantage for between-category discrimination (BF_{10} = .908, *t*[25] = 1.41, *p* = .085, d_z = .277) (LVF: 712 ms, SD = 174 ms; RVF: 698 ms, SD = 180 ms).

4.6.2.2 Additional analyses. In ERs, we found strong evidence against the presence of an LVF-advantage for within-category discrimination (BF_{10} = .087, *t*[25] = 1.65, *p* = .944, d_z = .324) (LVF: 38%, SD = 12%; RVF: 33%, SD = 10%). The evidence was inconclusive with regard to an RVF-advantage for between-category discrimination in ERs (BF_{10} =. 774, *t*[25] = 1.29, *p* = .104, d_z = .253) (LVF: 29%, SD = 14%; RVF: 26%, SD = 11%).

4.6.2.3 Combined results. When taking the original and replication studies' results together, there is substantial evidence against the presence of an LVF-advantage for within-category discrimination (BF_{10} = .243), and inconclusive evidence regarding an RVF-advantage for between-category discrimination (BF_{10} = 2.75), in RTs.

4.6.3 Discussion. We were not able to replicate the lateralized Whorf effects using animal shapes as stimuli. The combined evidence of the original and replication studies indicates evidence against the presence of an LVF-advantage for within-category discrimination, and the combined evidence is inconclusive with regard to the RVF-advantage for between-category discrimination. We will now consider whether the differences between the original and replication studies could account for the differences in results.

In the original study, participants had unlimited time to make their response, while in the replication study we chose to shorten this interval to 5 s. As a result, in the former a response was always recorded, while in the latter 'misses' occurred, in 0.6% of the trials. Although participants in the replication study almost never failed to respond, they did produce a higher number of errors (mean 30%) than participants in the original study (mean 5.6%). While participants in the original study did receive instructions to respond as quickly as possible, the fact that they had unlimited time to respond may have resulted in their putting more weight on accuracy than on speed. In comparison, participants in the replication study may have put more weight on a speedy response, to the detriment of accuracy. Support for this can be seen in the RTs, which are on average lower in the replication study than in the original study (Figure 9). However, given that ERs are the most informative measure when they are relatively high (Hellige & Sergent, 1986), the effects of interest could have been expected to surface in ERs, which was not the case.

Because we chose to combine the COT and SOT in one test session, we limited the number of trials in the SOT to 384 trials. This, however, is only about half the number of trials participants performed in the original study. As a result, we may have had less power to detect the RVF-advantage for between-category discrimination than we calculated based on the original study's effect size and our number of participants. In combination with the Bayes factors indicating that there was inconclusive evidence with regard to the presence of this RVF-advantage in RTs and ERs, we cannot rule out the possibility that the RVF-advantage for between-category discrimination would have been found, had our replication study used more trials. In contrast, as the Bayes factors indicated a sufficient amount of evidence against the presence of an LVF-advantage for within-category discrimination in ERs and RTs, we can be certain that the failure of replicating this effect is not due to the diminished number of trials in the replication study.

In conclusion, the results we found in our replication of the SOT are similar to those of our replication of the COT. We were not able to replicate the RVF-advantage for between-category discrimination, but in the case of the SOT, this may have been a consequence of low power. In addition, we failed to replicate the LVF-advantage for within-category discrimination. We will reflect further on the likeliness that there is a lateralized influence of shape categories on shape discrimination in section 5.3.

4.7 Cross-dot Matching Task (CMT)

Kosslyn proposed that the two hemispheres are lateralized with regard to two different types of spatial relation processing (Kosslyn, 1987). For example, a categorical spatial relation judgment (e.g., “the ball is to the right of the table”) is more easily made when the stimulus is presented in the RVF, and a coordinate spatial relation judgment (e.g., “the ball is one meter away from the table”) is more easily made when the stimulus is presented in the LVF (for reviews, see Jager & Postma, 2003; Laeng, Chabris, & Kosslyn, 2003). The task predominantly used to study lateralization of spatial relation processing is the bar-dot task (Hellige & Michimata, 1989; Kosslyn et al., 1989). However, Van der Ham, van Wezel, Oleksiak and Postma (2007) identified and attempted to overcome two drawbacks of this widely used bar-dot task.

Firstly, in bar-dot tasks, participants show a training effect over trials, resulting in categorization of the coordinate task condition into different categories of nearness. Secondly, the categorical condition of the bar-dot task seemed to be less difficult than the coordinate task condition. Consequently, Van der Ham et al. (2007) reasoned, the possibility that the VFAs had been caused by differences in difficulty between the conditions could not be ruled out. To overcome these problems, Van der Ham et al. (2007) introduced an alternative task to study lateralization of categorical and coordinate spatial relation processing, making use of cross-dot configurations. We chose to replicate the Cross-dot Matching Task (CMT) that includes a self-rating of spatial strategy (Van der Ham & Borst, 2011), as with this study the authors showed that individual differences in spatial strategy can affect the found VFAs.

In the course of this replication attempt, Van der Ham and Borst published a corrigendum to their original research article (2016). This corrigendum stated that a coding error had occurred in the analyses that were reported in the original article. The originally reported VFAs largely disappeared when these errors were corrected for. Nevertheless, we decided to report the outcomes of the replication study here, and we relate our results to the corrected results as reported in the 2016 corrigendum.

4.7.1 Methods.

4.7.1.1 Participants. Thirty-four participants (17 women) performed the CMT. Their mean age was 21 years (range = 18-28).

4.7.1.2 Stimuli. The first stimulus (S1) consisted of a centrally presented plus sign (the ‘cross’ of the cross-dot stimulus) of 0.35° degrees of visual angle, and a dot of 0.15°. The dot could appear at one of forty fixed positions, in relation to the cross. The second stimulus (S2) could either be a match or a non-match to the S1. In the categorical task, a match was defined as the dot appearing in the same quadrant (upper left, upper right, lower left and lower right, with regard to the cross) as the dot in the S1. In the coordinate task, a match was defined as the dot appearing in the same radius (inner ring, first ring, second ring, or outer ring, with regard to the cross) as the dot in the S1. The stimuli and fixation cross were presented in black on a white background, and the inner edge of the S2 was 2.5° from the center.

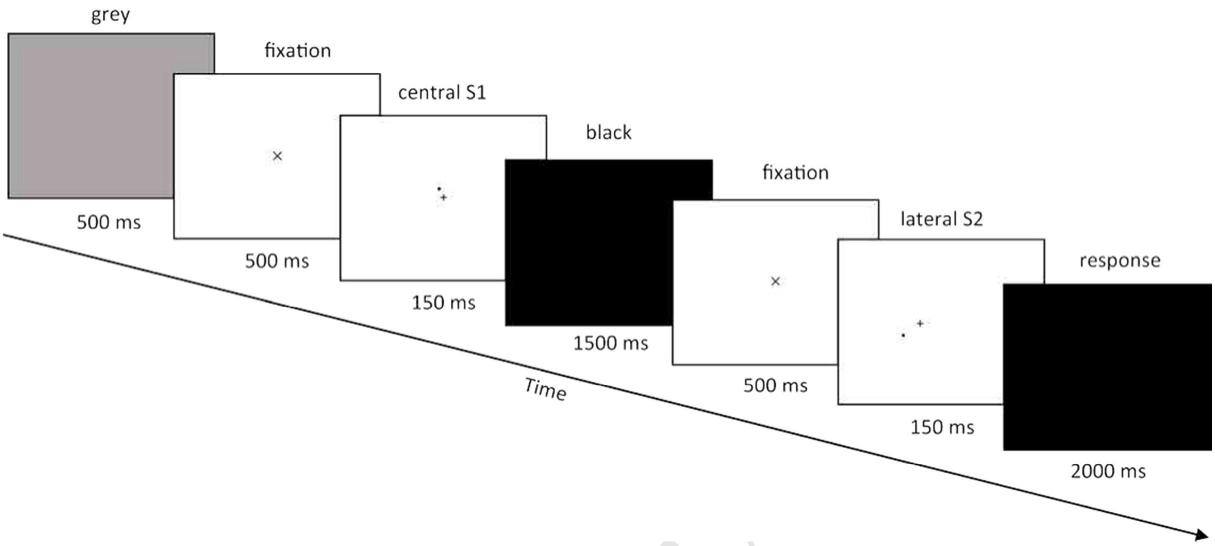

Figure 11. Timeline of a trial in the Cross-dot Matching Task.

4.7.1.3 Procedure. A grey screen lasting 500 ms signaled the start of the new trial, after which a fixation cross was presented for 500 ms (Figure 11). Next, the S1 was presented for 150 ms, followed by a black screen for 1500 ms. Following another fixation cross for 500 ms, the S2 was presented in the LVF or RVF for 150 ms. After this, a black screen appeared and participants had 2 s to make their response. Participants were asked to indicate whether the S1 and S2 were a match or non-match, by pressing one of two buttons with the index or middle finger of their right hand, as fast and accurately as possible. Finger-response mappings were counterbalanced over participants. During the instructions, it was stressed that participants should take into account the position of the dot relative to the cross, and not take into account the positioning of the cross-dot configuration on the screen.

Participants completed four blocks of 40 trials in both the categorical and the coordinate task, and each task was preceded by eight practice trials during which participants received feedback on their performance. Match and non-match trials, and LVF and RVF presentations occurred equally often, in a randomized manner. The ordering of the categorical and coordinate tasks was counterbalanced over participants.

At the end of the experiment, participants filled out a questionnaire about the strategy they had used during CMT performance. Participants were asked to rate the degree to which they had used a spatial strategy in the categorical and coordinate tasks, on a seven-point Likert scale.

4.7.1.4 Effects of interest. The original VFAs that we initially considered to be the effects of interest did not all survive the corrected analyses as reported in Van der Ham and Borst's corrigendum (2016). We relate our results to the effects reported in the corrigendum. Consequently, we had less than 80% power to detect the adjusted effects of interest.

Based on the surviving effects as reported in the corrigendum (2016), the effects of interest became the RVF-advantage for categorical processing in the high spatial strategy group in ERs (based on the original study's effect size of d_z = .421, we had 51% power detect the effect with our sample size[9]), the LVF-advantage for categorical processing in the low spatial strategy group in ERs (original d_z = -.528, 70% power), the LVF-advantage for coordinate processing in the high spatial strategy group in ERs (original d_z = -.417, 50% power) and in the group as a whole ERs (original d_z = -.304, 54% power), and the RVF-advantage for coordinate processing in the group as a whole in RTs (original d_z = .251, 42% power).

For the analyses, the participants were divided into two groups based on their median scores on the spatial strategy questionnaire, as in the original study (Van der Ham & Borst, 2011),

[9] Because we divide the participants into high and low spatial strategy groups based on a median split, there would be 17 participants in each group.

4.7.1.5 Additional analyses. In addition, we analyzed the remaining contrasts using the expected directions of the VFAs based on the theory regarding lateralized processing of categorical and coordinate spatial processing (i.e., not based on the unexpected reversed asymmetries found by Van der Ham and Borst (2016)). This resulted in the analyses of the RVF-advantage for categorical processing in the group as a whole, in ERs and RTs, and in the high and low spatial strategy groups in RTs. With regard to coordinate processing, we analyzed the LVF-advantage in the low spatial strategy group in ERs, and this VFA in both high and low spatial strategy groups in RTs.

4.7.1.6 Differences with original study. The CMT is a full replication of the original study (Van der Ham & Borst, 2011, 2016), with the exception of the number of trials. The original study used a selection of 80 of the 160 possible cross-dot configurations, each participant receiving the same fixed-order selection. In the replication attempt, we doubled the number of trials, so that each participant received all possible cross-dot configurations. For each participant, a new randomization of trial order was used. The first author of the original study shared the experiment E-Prime file and stimulus image files.

Figure 12. Error rates (lower panels) and reaction times (upper panels) of the replication (left panels) and original (right panels) studies' Cross-dot Matching Task, with the results of the participants in the high spatial strategy group (A), and the results of the participants in the low spatial strategy group (B). The means of the original study are copied from Table 2 of Van der Ham & Borst (2016, p. 41). Error bars represent standard errors of the means.

4.7.2 Results.

4.7.2.1 Effects of interest. We failed to replicate the RVF-advantage for categorical processing in the high spatial strategy group in ERs (BF_{10} = .238, *t*[16] = -.06, *p* = .523, d_z = -.014) (LVF: 21%, SD = 17%; RVF: 21%, SD = 18%). In addition, our data were inconclusive regarding the replication of the LVF-advantage for categorical processing in the low spatial strategy group in ERs (BF_{10} = .460, *t*[16] = -.71, *p* = .245, d_z = -.171) (LVF: 14%, SD = 10%; RVF: 15%, SD = 11%).

We failed to replicate the LVF-advantage for coordinate processing in ERs in the high spatial strategy group (BF_{10} = .326, *t*[16] = -.34, *p* = .370, d_z = -.082) (LVF: 29%, SD = 8.6%; RVF: 30%, SD = 10%), and in the group as a whole (BF_{10} = .221, *t*[33] = -.23, *p* = .409, d_z = -.040) (LVF: 30%, SD = 8.4%; RVF: 30%, SD = 9.4%). We failed to replicate the RVF-advantage for coordinate processing in the group as a whole in RTs (BF_{10} = .102, *t*[33] = -.95, *p* = .826, d_z = -.163) (LVF: 674 ms; SD = 134 ms; RVF: 681 ms, SD = 145 ms).

4.7.2.2 Additional analyses. We found substantial evidence against the presence of RVF-advantages for categorical processing in ERs in the group as a whole (BF_{10} = .123, *t*[33] = -.60, *p* = .722, d_z = -.102) (LVF: 18%, SD = 14%; RVF: 18%, SD = 15%), in RTs in the group as a whole (BF_{10} = .145, *t*[33] = -.33, *p* = .629, d_z = -.057) (LVF: 700 ms, SD = 139 ms; RVF: 702 ms, SD = 128 ms), in RTs in the high spatial strategy group (BF_{10} = .153, *t*[16] = -.79, *p* = .779, d_z = -.191) (LVF: 687

ms, SD = 149 ms; RVF: 694 ms, SD = 143 ms), and in RTs in the low spatial strategy group (BF_{10} = .202, *t*[16] = .30, *p* = .618, d_z = .074) (LVF: 713 ms, SD = 131 ms; RVF: 711 ms, SD = 114 ms). Furthermore, we found substantial evidence against the presence of an LVF-advantage for coordinate processing in ERs in the low spatial strategy group (BF_{10} = .240, *t*[16] = .05, *p* = .520, d_z = .013) (LVF: 31%, SD = 8.4%; RVF: 31%, SD = 8.9%), and inconclusive evidence for this VFA in RTs in the low spatial strategy group (BF_{10} = .448, *t*[16] = -.68, *p* = .254, d_z = -.164) (LVF: 683 ms, SD = 106 ms; RVF: 691 ms, SD = 104 ms) and the high spatial strategy group (BF_{10} = .441, *t*[16] = -.66, *p* = .259, d_z = -.161) (LVF: 664 ms, SD = 160 ms; RVF: 670 ms, SD = 180 ms).

4.7.2.3 Combined evidence. When combining the original and replication studies' results, there is strong evidence for an LVF-advantage for categorical processing in ERs in the low spatial strategy group (BF_{10} = 15.1), and inconclusive evidence for an RVF-advantage for categorical processing in the high spatial strategy group in ERs (BF_{10} = 1.87). However, there is strong evidence against the presence of an RVF-advantage for categorical processing in the group as a whole in ERs (BF_{10} = .098) and RTs (BF_{10} = .048).

There is inconclusive evidence with regard to the LVF-advantage for coordinate processing in ERs in the high spatial strategy group (BF_{10} = 3.10), and in the low spatial strategy group (BF_{10} = .351). In the group as a whole, there is inconclusive evidence with regard to the RVF-advantage in coordinate processing in RTs (BF_{10} = .412), and with regard to the LVF-advantage for coordinate processing in ERs (BF_{10} = 2.60). We must note that in the corrigendum, the reported degrees of freedom are not in line with the reported sample sizes[10], so we cannot be certain that the calculated Bayes factors are exact.

4.7.3 Discussion. We were not able to replicate any of the original study's VFAs. Instead, we found substantial evidence against the presence of eight of the VFAs, and inconclusive evidence regarding the other four VFAs. Of note is the fact that when we combined the results of the original and replication studies, we found substantial evidence for the existence of one VFA which was opposite from what we would expect based on Van der Ham and Borst's (2011) predictions. Furthermore, in our replication attempt, the spatial strategy used by the participant did not affect the results as predicted.

The fact that the replication study had low power to detect the effects of interest cannot be claimed to have caused the difference in results, as eight of the twelve Bayes factors indicated at least substantial evidence against presence of the VFAs, reflecting that the amount of data was sufficient to support these null hypotheses.

In considering the possible reasons for why our replication study did not replicate the effects found by Van der Ham et al. (2016), a first possibility might be the fact that we used double the number of trials compared to the original study. To examine whether this may have influenced the results, we repeated the analyses on only the first half of the trials (resulting in 18 participants' categorical blocks and 16 participants' coordinate blocks). The *p*-values and Bayes factors of the effects of interest remained qualitatively unchanged, with the exception two effects in the low spatial strategy group whose evidence for their absence as reflected by the Bayes factors changed from substantial to inconclusive (coordinate processing ERs all trials

[10] The degrees of freedom were neither in line with the reported sample sizes of the original research article.

included: BF_{10} = .240; first half of trials: BF_{10} = .440; categorical processing RTs all trials included: BF_{10} = .202; first half of trials: BF_{10} = .502), and for one effect in the high spatial strategy group with the evidence for its absence changing from inconclusive to substantial (coordinate processing ERs all trials included: BF_{10} = .326; first half of trials: BF_{10} = .243). These relatively unchanged results indicate that the failure of replication of the original results cannot be attributed to the fact that the replication study included more trials.

It remains possible that the differences in selection and randomization of the cross-dot configurations between the original and replication studies caused the differences in results. However, if this were the case, this would imply that the originally found effects depended solely on the sub-set of configurations, and/or their specific order.

We consider it noteworthy that in Van der Ham & Borst's 2016 corrigendum there seems to be a speed-accuracy trade-off for coordinate processing, with an LVF-advantage for this VFA in ERs, but an RVF-advantage in RTs. As such, their data do not seem to support the existence of an LVF-advantage for coordinate processing, as measured by the CMT.

In conclusion, we do not consider the study of Van der Ham & Borst (2016) a strong case for categorical and coordinate lateralized processing, and in that sense the findings of the replication study are in concordance with those reported in the original study's 2016 corrigendum. Based on these results, we conclude that if lateralization of categorical and coordinate spatial relation processing exists, the CMT does not seem to be an adequate task to study them. Alternatively, these results may be taken to suggest that there is low evidence for lateralization of categorical and coordinate spatial relation processing. We return to this point in section 5.3.

4.8 Landmark Task

Observers have a tendency to view the left side of space as being larger than the right side of space, an LVF-bias referred to as 'pseudoneglect' (Bowers & Heilman, 1980), which has been associated with RH-specialization for visuo-spatial attention (Çiçek, Deouell, & Knight, 2009; Zago et al., 2015; Ocklenburg & Güntürkun, 2018). Over the years, pseudoneglect has been studied in several modalities and with a variety of tasks. A review and meta-analysis by Jewell and McCourt (2000) reported that tasks using limited viewing time and a forced-choice response, such as the Landmark Task (LT), produce larger effect sizes than the more often used method-of-adjustment, or line bisection procedures. Manly, Dobler, Dodds and George (2005) devised a computerized version of the LT, after which Linnell et al. (2014) modeled their LT. Given that Linnell et al.'s description of the methods allowed for a full replication attempt, we performed a replication of their study (subsample of British participants).

4.8.1 Methods.

4.8.1.1 Participants. Forty-three participants (21 women) performed the LT. Their mean age was 22 years (range = 18-31).

4.8.1.2 Stimuli. A stimulus consisted of a horizontal line of 18.8° of visual angle, presented in black on a white background. The line was transected by a vertical line of 0.8°, positioned at -1.2°, -0.8°, -0.4°, 0.0°, +0.4°, +0.8°, or +1.2° from the midpoint. This resulted in seven conditions: three in which the left part was longer than the right, three in which the right part was longer than the left, and one with equally long left and right parts. Stimuli were equally often presented centrally, or jittered 1.1° to the right or left of the center. The mask consisted of a horizontal line, subtending from the far left to the far right of the screen, with 85 transecting lines across the length of the horizontal line, spaced 0.4° apart.

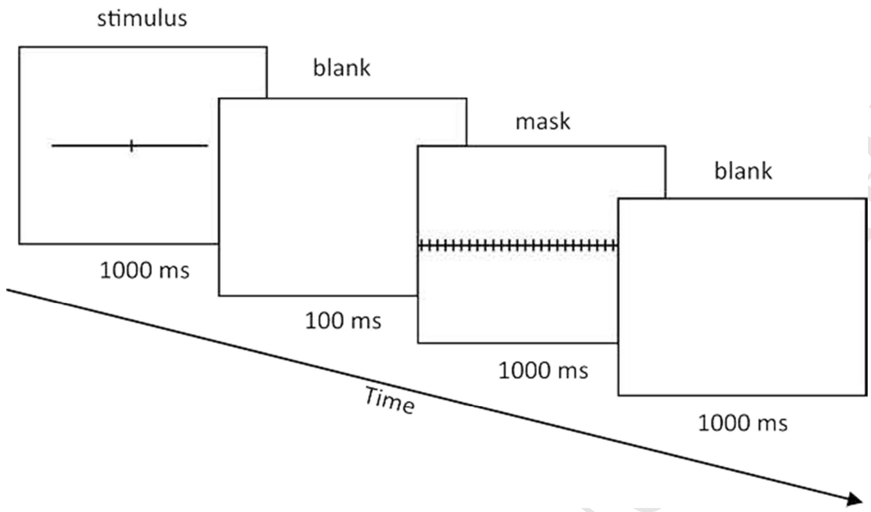

Figure 13. Timeline of a trial in the Landmark Task.

4.8.1.3 Procedure. A trial began with the presentation of the stimulus for 1000 ms (Figure 13). After a blank screen of 100 ms, the mask was presented for 1000 ms, followed by another blank screen for 1000 ms. Participants were instructed to indicate which part of the horizontal line was longer; the part left of the transection or right of the transection. The participants responded by pressing the F-key or H-key on a QWERTY-keyboard, using their left or right index finger, respectively. They could make their response from the moment of stimulus presentation to the end of the trial. They were not required to respond as fast as possible, but were asked to respond before the end of the trial.

Each transection position occurred 12 times, resulting in a total of 84 trials. Participants were not informed that in one seventh of the trials the line was transected exactly in the middle. Before the experimental task started, participants were given ten practice trials with transections at -1.4° and +1.4°, as seen from the midpoint.

4.8.1.4 Effects of interest. The effect of interest was whether participants showed an LVF-bias in that they judged the midpoint of the line to be to the left of the veridical midpoint. To assess this, each participant's point of subjective equality (PSE) was derived by finding his or her threshold for deciding that the right part of the line was longer than the left part (Linnell et al., 2014). PSEs are expressed in deviation from the veridical center, in degrees of visual angle. Based on the original study's (British participants) effect size of d_z = -.361, we had 75% power to detect this visual field bias with our sample size[11]. No additional analyses were planned.

4.8.1.5 Differences with original study. The LT was a full replication of the study by Linnell et al. (2014). The stimuli and procedure were identical to those of the original study.

The original study compared pseudoneglect between a group of British participants and a group of Himba participants, to investigate the effect of urbanization on spatial attention distribution. Our replication relates only to the British participant group of the original study.

4.8.2 Results. We replicated the original study's LVF-bias (BF_{10} = 7.66, t[42] = -2.68, p = .005, d_z = -.409) (PSE = -.08, SD = .19). When combining the original and replication studies, there is decisive evidence in favor of the presence of an LVF-bias (BF_{10} = 160).

[11] We initially performed the power analysis based on the effect size d = .497, as reported in the original research article (Linnell et al., 2014), resulting in a required sample size of 27 to reach 80% power. On later inspection, we found this effect size not to correspond to the mean and standard deviation as reported in the article, and to be an overestimation of the actual effect size. We then calculated Cohen's d_z based on the statistics reported in the article, and refer to this value in the rest of this paper.

4.8.3 Discussion. The results of this replication attempt yield behavioral support for the existence of a lateralized distribution of spatial attention. Specifically, we found substantial evidence for an LVF-bias in spatial attention in our replication attempt and the original and replication studies combined provide decisive evidence for the lateralization of spatial attention distribution.

4.9 Lexical Decision Task

LH-lateralization for language was the first described instance of hemispheric specialization of the human brain, and is now considered to be one of the most reliably lateralized processes (Hugdahl, 2000). A multitude of ways to study language lateralization has accumulated over the years (e.g., lesion studies, assessing the ear advantage in dichotic listening, neuroimaging combined with word fluency tests), but the Lexical Decision Task (LDT) is an often used way to study lateralization of written language (e.g., Cai, Paulignan, Brysbaert, Ibarrola, & Nazir, 2010; Hellige & Yamauchi, 1999). The aim of the study by Willemin et al. (2016) was to devise an LDT that could be used in five different languages (French, German, Italian, English, Dutch), and to test it in a French-speaking population. The current replication attempt investigates the reproducibility of the original VFAs in Willemin et al.'s study, in a Dutch-speaking population.

4.9.1 Methods.

4.9.1.1 Participants. Forty-eight native Dutch speakers (39 women) performed the LDT. Their mean age was 20 years (range = 17-28).

4.9.1.2 Stimuli. The international word and non-word set described in Willemin et al. (2016) was used in the LDT. This set comprises sixteen 4-, 5- or 6-letter words, which are meaningful in Dutch, German, English, French and Italian. Pseudowords, created by changing two letters from each of the words, were coupled to the words, resulting in sixteen word/non-word pairs. In addition, sixteen non-word/non-word pairs were used. The size of the letter strings was on average 3° of visual angle, the inner edge presented 2° from the center. The height of the stimuli was 0.5°. The letters were presented in Courier New (12 point), in black on a white background.

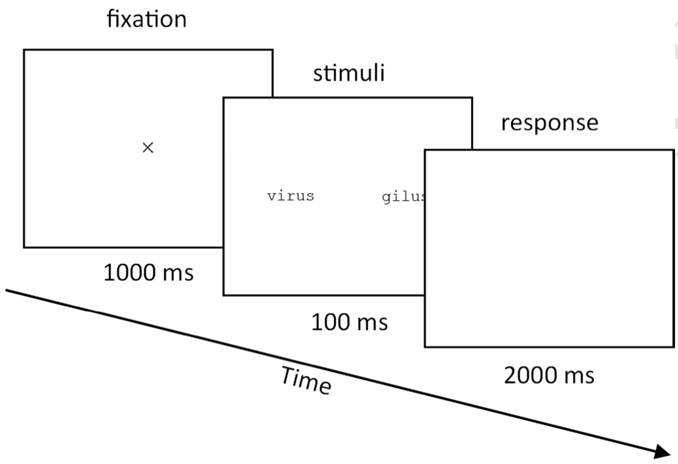

Figure 14. Timeline of a trial in the Lexical Decision Task.

4.9.1.3 Procedure. A trial started with the presentation of a fixation cross at the center of the screen, for 1000 ms (Figure 14). Next, the stimulus pair followed, with one letter string in the LVF and the other in the RVF, for 100 ms. After the stimulus pair, a blank screen followed, and participants had 2 s to respond. Participants were required to indicate whether they had seen a meaningful word on the left side, on the right side, or not at all. They did so by pressing the F-key with their left index finger, the J-key with their right index finger, or the space bar with both thumbs, respectively. Participants were asked to respond as quickly and accurately as possible.

The word/non-word pairs and non-word/non-word pairs were presented in an LVF-RVF and RVF-LVF configuration, four times each. This resulted in a total of 256 trials, which were presented in randomized order. Participants completed these trials in two blocks, with a self-paced break in between.

4.9.1.4 Effects of interest. The effects of interest were the RVF-advantage in ERs (based on the original study's effect size of d_z = .88, we had more than 99% power to detect the effect with our sample size), and the RVF-advantage in RTs (original d_z = .559, 99% power). No additional analyses were planned.

4.9.1.5 Differences with original study. The LDT was a full replication of the original study, with the exception of the native language of the participants. The original experiment DMDX-file was shared by the authors of the original study. As such, stimuli and procedure were identical to that of the original study, but with instructions in Dutch.

The original study examined influences of handedness, sex and multilingualism of the participants on the VFAs, but found no differences between the groups. For this reason, with the replication study we did not address these aspects.

Figure 15. Error rates (lower panels) and reaction times (upper panels) of the replication (left panel) and original (right panel) studies' Lexical Decision Task. The means of the original study are copied from Table 1 in Willemin et al. (2016, p. 10). Error bars represent standard errors of the means.

4.9.2 Results. We replicated the RVF-advantage for visual word processing in RTs (BF_{10} = 282, $t[47] = 4.08$, $p < .001$, $d_z = .589$) (LVF: 708 ms, SD = 98 ms; RVF: 658 ms, SD = 80 ms), and in ERs (BF_{10} = 7,260,758, $t[47] = 7.26$, $p < .001$, $d_z = 1.05$) (LVF: 36%, SD = 15%; RVF: 19%, SD = 11%). When combining the original and replication results, there is decisive evidence for the presence of an RVF-advantage in visual word processing in RTs (BF_{10} = 247,259,539) and ERs (BF_{10} = 1.113278+19).

4.9.3 Discussion. The results of this replication yield behavioral manifestations of, and thus support the existence of, lateralized processing of visually presented words. Specifically, we replicated the original RVF-advantages in ERs and RTs, in a group of native Dutch-speaking participants. Combining the results of the original and replication studies indicated that there is decisive evidence for lateralization of visual word processing.

5. General Discussion

Accurate characterization of hemispheric specialization and the resulting instances of lateralized processing of sensory information is critical to our understanding of how the human brain functions. By allowing for tightly controlled manipulations in powerful within-subject designs, behavioral studies can provide an essential contribution to our understanding of such hemispheric specialization. Importantly, however, a key requirement for such studies to be useful is that they reliably demonstrate differences in behavior or performance when visual targets of a certain type appear in either the left (LVF) or the right (RVF) visual field. In the current study, we took a rigorous empirical and statistical approach in investigating the reliability of a large number of previously found visual field asymmetries (VFAs) by means of a series of replication studies.

5.1 Summary of Results and Methodological Implications

Figure 16. Overview of all original (light) and replication (dark) studies' effect sizes (Cohen's d_z). Error bars indicate 95% confidence interval of the effect size. In the two rightmost columns the

Bayes factors of the replication study and the meta-analytical Bayes factors are presented. Asterisks to the left of the effects sizes indicate that either or both of those Bayes factors reflect there to be substantial (*), strong (**), very strong (***), or decisive (****) evidence for the presence of the expected visual field asymmetry.

An overview of our findings can be found in Figure 16. This figure shows the effect sizes and confidence intervals for a total of 41 putative VFAs that could be tested in the nine experiments we included in our replication studies (grey datapoints). In addition, Figure 16 shows for which of these effects the original studies reported the outcomes of a statistical analysis, and it illustrates the outcomes of these statistical analyses in terms of the resulting estimates of effect sizes and their confidence intervals (white datapoints). Moreover, Figure 16 also includes the Bayes factors for the effects we obtained in our replication studies and for our tests of the combined evidence from the original and replication studies, for those cases for which this computation was possible[12]. In interpreting these results, the Bayes factors can be considered to provide an index of the likelihood of the presence of a particular VFA, while the effect sizes and their confidence intervals provide insight into how strong these effects have been estimated to be, and how confident we can be about the precision of these estimates.

In evaluating the evidence depicted in Figure 16, a number of observations can be made. To start, there are five tasks that stand out in terms of producing precise, reliable evidence for moderate to strong effects of lateralization on performance, namely the Face Similarity Task (FST), the Face Emotionality Task (FET), the Hierarchical Letter Task (HLT), the Landmark Task (LT), and the Lexical

[12] We could not compute a meta-analytic Bayes factor for effects for which the original study did not report the outcome of a statistical test.

Decision Task (LDT). Specifically, the FST and the FET yielded clear evidence for an LVF-advantage in perceptual judgments of faces and their emotional expressions. The results of the HLT showed LVF- and RVF-advantages for the processing of global and local levels of hierarchical letters, respectively. Regarding the HLT, it is of note that although the outcomes of the Bayesian analyses produced strong evidence in favor of the presence of the predicted lateralization effects, the confidence intervals were relatively large, and the estimates, therefore, less precise. This indicates that future studies employing this task should use a large sample of participants so as to ensure a reliable estimate of the true effect size. The LT yielded precise and convincing evidence for an LVF-bias in the distribution of spatial attention. Finally, the LDT yielded compelling evidence for an RVF-advantage in the detection of words. Taken together, these results corroborate the existence of hemispheric specialization for the processing of faces and emotional expressions, of the global and local elements of visual stimuli, in the distribution of spatial attention, and in the processing of visually presented words. Furthermore, since the resulting effect sizes are similar to those of the original studies, we consider the current five tasks to offer highly useful, reliable tools to elicit and study the behavioral manifestations of these instances of hemispheric specialization.

A less convincing pattern of results can be seen for the Picture Matching Task (PMT) that was introduced by Peyrin et al. (2006) as a tool to study the putative lateralized processing of stimuli's high (HSF) and low (LSF) spatial frequency content. As illustrated in Figure 16, this task allows for twelve tests of lateralization of which only four were found to reach significance in the original study. Specifically, the original study by Peyrin et al. (2006) only showed evidence for lateralization effects in reaction times (RTs), but not in error rates (ERs), and the observed effects had relatively large effect sizes with relatively large confidence intervals. In comparison, our replication attempt yielded smaller effect sizes with greater precision, and these effects aligned with those of Peyrin et al. (2006) in demonstrating an RVF-advantage for processing of HSF stimuli, albeit in ERs rather than RTs. However, for LSF stimuli, the results were less clear, as the Bayes factors in our analyses were inconclusive and only the combined evidence pointed towards the presence of LVF-advantages in processing LSF stimuli. Accordingly, we conclude that further, high-power replication attempts are needed to establish the usefulness of Peyrin's PMT as a tool to elicit and study the behavioral effects of lateralized processing of high and low spatial frequencies.

Lastly, the current study also included three tasks that did not produce reliable evidence for the effects of brain lateralization on performance. To start, we obtained no evidence for RVF-advantages in detecting color or shape oddballs belonging to different categories than the distractors. For both the Color (COT) and Shape (SOT) Oddball Tasks, our results yielded more precise estimates of the effects than the original studies, and our results failed to replicate the earlier found effects. Secondly, we obtained no evidence for effects of lateralization in categorical and coordinate judgments of spatial relationships in the Cross-dot Matching Task (CMT) (Van der Ham & Borst, 2011; 2016), as we did not replicate the effects that survived in the corrigendum by Van der Ham and Borst (2016) and we also did not find evidence for a number of other effects that were predicted for this task in the original report by Van der Ham and Borst (2011). Accordingly, we conclude that the CMT, COT, and SOT do not reliably elicit behavioral manifestations of lateralized information processing.

5.2 A Generic Role for Stimulus and Task Factors?

In view of our large collection of successful and less successful replications of previous findings, an interesting question is whether there are any methodological factors that distinguish the paradigms that do and do not produce reliable behavioral effects of brain lateralization. Indeed, there are many previous studies that have attempted to demonstrate behavioral effects of lateralization and that have concluded that the observation of such effects may depend on various potential modulators (e.g., Bourne, 2006; Hellige & Sergent, 1986; Hunter & Brysbaert, 2008; Yovel et al., 2001; for reviews, see Dien, 2008; Springer & Deutsch, 1998), such as the presentation parameters used in displaying the stimuli (e.g., stimulus duration, presence of masks, bilateral vs. unilateral stimulus presentation), and the nature of the task (e.g., target detection, target discrimination, judging the similarity between two stimuli). Accordingly, we can ask the question whether the current set of results can be understood in terms of the fact that lateralization effects are more likely to surface at the behavioral level when a paradigm has a certain combination of presentation and task parameters. In considering this possibility, we note that the effects that were replicated were obtained in different types of tasks (target detection, target identification, and stimulus matching), using either a free-viewing or a visual half-field technique, and for various presentation durations. Therefore, we conclude that the likelihood of observing a behavioral effect of brain lateralization in one of the currently used paradigms is not related directly to a specific setting of parameters.

5.3 Beyond the Reliability of Specific Paradigms: Implications for Lateralization

Given that it is difficult to explain our mixed success at replicating previous findings exclusively in terms of methodological factors, an alternative account could be that the success of replication in the current study relates to whether or not a certain type of visual stimulus is indeed processed in a lateralized manner. In this view, the successfully replicated VFAs for faces, emotional expressions, global and local stimuli, spatial attention distribution, and words would be interpreted to reflect the existence of lateralization, whereas the non-replicated VFAs for the influence of categorical processing in detection of color and shape oddballs, and for judgments of spatial relations would be interpreted to reflect the non-existence of lateralized processing in these tasks. In the following sections, we discuss this possibility as we address the relationship between the current findings and those of previous studies that have examined the same instances of lateralization using different behavioral paradigms and more direct measures of brain functioning, such as studies on the effects of lateralized brain injury, and studies employing neuroimaging.

In relating the current findings to the broader context of previous studies investigating the same instances of lateralization with different methods, it becomes clear that the pattern of successful and non-successful replications across the current set of studies resonates well with the amount and consistency of the currently available evidence pertaining to the underlying instances of lateralization. To start, our finding of an LVF-advantage in the FST is consistent with a large body of findings demonstrating RH-specialization for processing faces in patients (e.g., De Renzi, Perani, Carlesimo, Silveri, & Fazio, 1994), in neuroimaging studies (e.g., Kanwisher, McDermott, & Chun, 1997), and in behavioral studies (for a recent meta-analysis, see Voyer et al., 2012). The current finding of an LVF-advantage for processing emotional expressions in the FET can likewise be considered to be “unsurprising” in view of the fact that a meta-analysis by Voyer et al. (2012) showed that many previous behavioral experiments using emotional faces have consistently demonstrated this advantage, with a large estimated pooled effect size. At the same, however, it is not yet clear whether this LVF-advantage should be interpreted as evidence for RH-dominance in processing emotional stimuli, as the results from one meta-analysis of neuroimaging studies showed no support for such generic RH-dominance in processing emotional stimuli (Wager, Luan Phan, Liberzon, & Taylor, 2003), whereas another showed that such lateralization may only pertain to the processing of faces that are difficult to perceive due to masking (Costafreda, Brammer, David, & Fu, 2008). Accordingly, an interesting question for future studies will be to examine whether the LVF-advantage for perceiving the emotional expression of faces in a chimeric face task derives from RH-specialization for processing faces, or from RH-specialization in processing emotional stimuli, under conditions with and without masks.

Likewise, our finding of convincing evidence for an LVF-bias in allocating attention converges with the results of many different types of studies showing RH-dominance in the control of spatial attention (e.g., Rafal, 1998). Lastly, our finding of an RVF-advantage in the LDT converges with a large number of studies which have shown that right-handed participants generally show LH-dominance for language in general (e.g., Springer & Deutsch, 1998; Vignau et al., 2005), and for processing linguistic visual stimuli in visual-half field studies in particular (e.g., Hunter & Brysbaert, 2008).

While our successful replication of VFAs for faces, spatial attention, and words can be considered an "unsurprising" result in view of the large and consistent body of evidence for lateralized modularity of the neural mechanisms involved in face processing, spatial attention, and language, a different opinion should apply to the lateralization for processing of global and local stimuli and of spatial frequency content. Specifically, an extensive review by Dien (2008) makes it clear that even though research on these instances of lateralization has a long history, the results of neuroimaging and patient studies have not consistently identified the presumed lateralized processing mechanisms, and the results of behavioral studies have likewise been mixed in providing evidence for the predicted VFAs. In light of these observations, the current finding that the HLT introduced by Yovel et al. (2001) produces convincing evidence for RH-global and LH-local processing biases can thus be said to be informative because it provides strong support for the existence of global-local lateralization. However, the current findings demonstrate only limited support for differential sensitivity to HSF and LSF stimuli in the PMT of Peyrin et al. (2006). Accordingly, to sustain the notion of lateralized processing of spatial frequency content would require additional and reliable observations.

Lastly, we consider the implications of the current findings for theories proposing the existence of lateralized influences of stimulus categories on making perceptual judgments. In addressing this matter, we examined the reliability of earlier findings that suggested the existence of an RVF-advantage in making a categorical judgment of the spatial relationship between two stimuli and in oddball detection when the oddball stimulus is categorically distinct from the distractors in terms of its color or shape. Importantly, our results offered little to no support for the reliability of these findings, thereby indicating that our results failed to offer support for theories proposing LH-dominance in categorical spatial judgments (e.g., Kosslyn et al., 1989) and in detecting categorically distinct visual oddballs (Gilbert et al., 2006; 2008). In considering the broader implications of these findings, it is of relevance to note that previous studies investigating the existence of LH-dominance in categorical spatial judgments have also offered only limited support for this form of lateralization (Van der Ham & Postma, 2010; Van der Ham, Raemaekers, Van Wezel, Oleksiak, & Postma, 2009; Van der Ham et al., 2007). Furthermore, an extensive review by Jager and Postma (2003) shows that behavioral tasks other than the one used in the current study have also produced mixed results, and it indicated that evidence for lateralized categorical and coordinate spatial relation processing from neuroimaging, patient, and computational modeling studies is also variable. Likewise, our failure to find support for previous findings of LH-dominance in detecting categorically distinct visual oddballs converges with the results of previous studies that also did not show evidence for VFAs using different tasks to measure lateralization of categorical color perception (Brown et al., 2011; Efron & Yund, 1996; Witzel & Gegenfurtner, 2011), and it is also consistent with the fact that there is little evidence from neuroimaging studies to support the existence of lateralization in the influence of categorical boundaries on visual search (for a review, see Witzel & Gegenfurtner, 2011). Taken together, we conclude that there is no consistent support for theories that propose a LH-dominance in categorical spatial judgments or in detecting categorically distinct targets in a visual oddball task.

5.4. Concluding Remarks and Recommendations for Future Studies

Aside from offering insight into the reliability and existence of several previously found instances of behavioral effects of brain lateralization, the current study also suggests a number of more general recommendations for future studies on lateralization. To start, our exposition of the results of previous studies (see Figure 16) shows that there has been considerable tolerance towards selective reporting when it comes to tests that fail to show predicted effects, meaning that non-significant lateralization studies often do not disclose sufficient detail to afford their use in meta-analyses. In light of the many disparate findings that have been obtained for various purported instances of lateralization, such meta-analyses are essential to assess the strength of effects, as well as the influences of modulators and publication bias. Accordingly, a first important general recommendation for future studies on lateralization is to fully disclose the results of all analyses, including those that did not yield statistically significant effects.

A second, related recommendation pertains to the degrees of freedom that researchers have when examining evidence for lateralization in behavioral and neuroimaging studies. On this point, it is noteworthy that a typical study using the visual half field paradigm has at least four opportunities to provide some evidence for lateralization, such that there might be LH- or RH-dominance on either RT or ER outcomes. In view of the degrees of freedom that these options for analysis offer, it seems crucial that researchers preregister their analysis plan so as to clarify which of these effects are predicted to occur in the light of the underlying theoretical rationale. In combination with the full disclosure of analyses and findings, such transparency will surely benefit the field by providing the evidence that is needed to identify robust instances of lateralization and to weed out any non-reliable observations and false conjectures.

Finally, we have pointed out a number of paradigms that produce reliable lateralization effects. These paradigms point towards potential underlying neural mechanisms. To establish the scope of the underlying mechanisms, we recommend that future studies should consider replication tests as well as testing variations of these paradigms.

Acknowledgements

We would like to thank Don van Ravenzwaaij for his advice on Bayesian analyses. We would like to thank Sophie Jöbkes, Lisette van Beek, Airi Yamada, Carel-Peter van Erpecum, Soma Makai-Boloni, Lukas Kowald, Louisa La Porta, and

Heike Schuler for their assistance in data collection. S.G. Brederoo was supported by a PhD scholarship as part of grant 022.004.008 awarded to the Research School for Behavioural and Cognitive Neuroscience (BCN) by the Netherlands Organisation for Scientific Research (NWO).

Author Note

The data for the Hierarchical Letter Task (section 4.3) are a subset of a larger dataset described in Brederoo et al. (2017). The current study and Brederoo et al. (2017) focus on different research questions.

Conflict of Interest: none.

References

Anderson, S. F., & Maxwell, S. E. (2015). There's more than one way to conduct a replication study: Beyond statistical significance. *Psychological Methods*, *21*, 1–12. http://doi.org/10.1037/met0000051

Beaumont, J. G. (1982). *Divided Visual Field Studies of Cerebral Organisation*. London: Academic Press Inc.

Boles, D. B. (1984). Global versus local processing: Is there a hemispheric dichotomy? *Neuropsychologia*, *22*, 445–455.

Boles, D. B. (1987). Reaction time asymmetry through bilateral versus unilateral stimulus presentation. *Brain and Cognition*, *6*, 321–333.

Boles, D. B., & Karner, T. A. (1996). Hemispheric differences in global versus local processing: Still unclear. *Brain and Cognition*, *30*, 232–43. http://doi.org/10.1006/brcg.1996.0015

Bourne, V. J. (2006). The divided visual field paradigm: Methodological considerations. *Laterality: Asymmetries of Body, Brain and Cognition*, *11*, 373–393. http://doi.org/10.1080/13576500600633982

Bowers, D., & Heilman, K. M. (1980). Pseudoneglect: Effects of hemispace on a tactile line bisection task. *Neuropsychologia*, *18*, 491–498. http://doi.org/10.1016/0028-3932(80)90151-7

Brederoo, S. G., Nieuwenstein, M. R., Lorist, M. M., & Cornelissen, F. W. (2017). Hemispheric specialization for global and local processing: A direct comparison of linguistic and non-linguistic stimuli. *Brain and Cognition*, *119*, 10–16. https://doi.org/10.1016/j.bandc.2017.09.005

Brown, A. M., Lindsey, D. T., & Guckes, K. M. (2011). Color names, color categories, and color-cued visual search: Sometimes, color perception is not categorical. *Journal of Vision*, *11*, 1–21. http://doi.org/10.1167/11.12.2

Brysbaert, M. & Stevens, M. (2018). Power analysis and effect size in mixed effects models: A tutorial. *Journal of Cognition*, *1*, 1-20.

Bunt, A. H., & Minckler, D. S. (1977). Foveal sparing: New anatomical evidence for bilateral representation of the central retina. *Archives of Ophthalmology*, *95*, 1445–1447.

Cai, Q., Paulignan, Y., Brysbaert, M., Ibarrola, D., & Nazir, T. A. (2010). The left ventral occipito-temporal response to words depends on language lateralization but not on visual familiarity. *Cerebral Cortex*, *20*, 1153–63. http://doi.org/10.1093/cercor/bhp175

Çiçek, M., Deouell, L. Y., & Knight, R. T. (2009). Brain activity during landmark and line bisection tasks. *Frontiers in Human Neuroscience*, *3*, 1–7. http://doi.org/10.3389/neuro.09.007.2009

Coronel, J. C., & Federmeier, K. D. (2014). Task demands modulate decision and eye movement responses in the chimeric face test: Examining the right hemisphere processing account. *Frontiers in Psychology*, *5*, 1–10. http://doi.org/10.3389/fpsyg.2014.00229

Costafreda, S. G., Brammer, M. J., David, A. S., & Fu, C. H. Y. (2008). Predictors of amygdala activation during the processing of emotional stimuli: A meta-analysis of 385 PET and fMRI studies. *Brain Research Reviews*, *58*, 57–70. doi:10.1016/j.brainresrev.2007.10.012

Daoutis, C. A., Pilling, M., & Davies, I. R. L. (2006). Categorical effects in visual search for colour. *Visual Cognition*, *14*, 217–240. http://doi.org/10.1080/13506280500158670

De Renzi, E., Perani, D., Carlesimo, G. A., Silveri, M. C., & Fazio, F. (1994). Prosopagnosia can be associated with damage confined to the right hemisphere–An MRI and PET study and a review of the literature. *Neuropsychologia*, *32*, 893–902. https://doi.org/10.1016/0028-3932(94)90041-8

Dien, J. (2008). Looking both ways through time: The Janus model of lateralized cognition. *Brain and Cognition*, *67*, 292–323. http://doi.org/10.1016/j.bandc.2008.02.007

Dienes, Z. (2014). Using Bayes to get the most out of non-significant results. *Frontiers in Psychology*, *5*, 1–17.

Dienes, Z., & Mclatchie, N. (2017). Four reasons to prefer Bayesian over orthodox statistical analyses. *Psychonomic Bulletin & Review*. http://doi.org/10.1017/CBO9781107415324.004

Drivonikou, G. V., Kay, P., Regier, T., Ivry, R. B., Gilbert, A. L., Franklin, A., & Davies, I. R. L. (2007). Further evidence that Whorfian effects are stronger in the right visual field than the left. *Proceedings of the National Academy of Sciences of the United States of America*, *104*, 1097–1102.

Efron, R., & Yund, E. W. (1996). Spatial nonuniformities in visual search. *Brain and Cognition*, *31*, 331–68. http://doi.org/10.1006/brcg.1996.0050

Ellis, A. W., & Brysbaert, M. (2010). Divided opinions on the split fovea. *Neuropsychologia*, *48*, 2784–5. http://doi.org/10.1016/j.neuropsychologia.2010.04.030

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*, 175–191.

Forster, K. I., & Forster, J. C. (2003). DMDX: A Windows display program with millisecond accuracy. *Behavior Research Methods*, *35*, 116–124.

Gilbert, A. L., Regier, T., Kay, P., & Ivry, R. B. (2006). Whorf hypothesis is supported in the right visual field but not the left. *Proceedings of the National Academy of Sciences of the United States of America*, *103*, 489–494.

Gilbert, A. L., Regier, T., Kay, P., & Ivry, R. B. (2008). Support for lateralization of the Whorf effect beyond the realm of color discrimination. *Brain and Language*, *105*, 91–98. http://doi.org/10.1016/j.bandl.2007.06.001

Gilbert, C., & Bakan, P. (1973). Visual asymmetry in perception of faces. *Neuropsychologia*, *11*, 355–62.

Hellige, J. B. (1995). Hemispheric asymmetry for components of visual information processing. In R. J. Davidson & K. Hugdahl (Eds.), *Brain Asymmetry*. Cambridge, MA: MIT Press.

Hellige, J. B., Laeng, B., & Michimata, C. (2010). Processing asymmetries in the visual system. In K. Hugdahl & R. Westerhausen (Eds.), *The Two Halves of the*

Brain: Information Processing in the Cerebral Hemispheres. Cambridge, MA: MIT Press.

Hellige, J. B., & Michimata, C. (1989). Categorization versus distance: Hemispheric differences for processing spatial information. *Memory & Cognition*, *17*, 770–776. http://doi.org/10.3758/BF03202638

Hellige, J. B., & Sergent, J. (1986). Role of task factors in visual field asymmetries. *Brain and Cognition*, *5*, 200–222. http://doi.org/10.1016/0278-2626(86)90055-2

Hellige, J. B., & Yamauchi, M. (1999). Quantitative and qualitative hemispheric asymmetry for processing Japanese kana. *Brain and Cognition*, *40*, 453–463. http://doi.org/10.1006/brcg.1999.1074

Hugdahl, K. (2000). Lateralization of cognitive processes in the brain. *Acta Psychologica*, *105*, 211–235. http://doi.org/10.1016/S0001-6918(00)00062-7

Innes, B. R., Burt, D. M., Birch, Y. K., & Hausmann, M. (2016). A leftward bias however you look at it: Revisiting the emotional chimeric face task as a tool for measuring emotion lateralization. *Laterality: Asymmetries of Body, Brain and Cognition*, *21*, 643–661. http://doi.org/10.1080/1357650X.2015.1117095

Jager, G., & Postma, A. (2003). On the hemispheric specialization for categorical and coordinate spatial relations: A review of the current evidence. *Neuropsychologia*, *41*, 504–15.

Jeffreys, H. (1961). *Theory of Probability*. (3rd ed.). London: Oxford University Press.

Jewell, G., & McCourt, M. E. (2000). Pseudoneglect: A review and meta-analysis of performance factors in line bisection tasks. *Neuropsychologia*, *38*, 93–110. http://doi.org/10.1016/S0028-3932(99)00045-7

Jordan, T. R., & Paterson, K. B. (2009). Re-evaluating split-fovea processing in word recognition: A critical assessment of recent research. *Neuropsychologia*, *47*, 2341–2353. http://doi.org/10.1016/j.neuropsychologia.2008.07.020

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience*, *17*, 4302–4311. https://doi.org/10.1523/JNEUROSCI.17-11-04302.1997

Karim, A. K. M. R., & Kojima, H. (2010). The what and why of perceptual asymmetries in the visual domain. *Advances in Cognitive Psychology*, *6*, 103–15. http://doi.org/10.2478/v10053-008-0080-6

Kauffmann, L., Ramanoël, S., & Peyrin, C. (2014). The neural bases of spatial frequency processing during scene perception. *Frontiers in Integrative Neuroscience*, *8*, 1–14. http://doi.org/10.3389/fnint.2014.00037

Kitterle, F. L., Hellige, J. B., & Christman, S. (1992). Visual hemispheric asymmetries depend on which spatial frequencies are task relevant. *Brain and Cognition*, *20*(2), 308–14. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/1449760

Kosslyn, S. M. (1987). Seeing and imagining in the cerebral hemispheres: A computational approach. *Psychological Review*, *94*, 148–175.

Kosslyn, S. M., Koenig, O., Barrett, A., Backer Cave, C., Tang, J., & Gabrieli, J. D. E. (1989). Evidence for two types of spatial representations: Hemispheric specialization for categorical and coordinate relations. *Journal of Experimental Psychology: Human Perception and Performance*, *15*(4), 723–735.

Laeng, B., Chabris, C. F., & Kosslyn, S. M. (2003). Asymmetries in encoding spatial relations. In K. Hugdahl & R. J. Davidson (Eds.), *The Asymmetrical Brain*. Cambridge, MA: MIT Press.

Levy, J., Heller, W., Banich, M. T., & Burton, L. A. (1983). Asymmetry of perception in free viewing of chimeric faces. *Brain and Cognition*, *2*, 404–419.

http://doi.org/10.1016/0278-2626(83)90021-0

Linnell, K. J., Caparos, S., & Davidoff, J. (2014). Urbanization increases left-bias in line-bisection: An expression of elevated levels of intrinsic alertness? *Frontiers in Psychology*, *5*, 1–7. http://doi.org/10.3389/fpsyg.2014.01127

Luh, K. E., & Gooding, D. C. (1999). Perceptual biases in psychosis-prone individuals. *Journal of Abnormal Psychology*, *108*, 283–9. http://doi.org/10.1037/0021-843X.108.2.283

Lundqvist, D., Flykt, A., & Öhman, A. (1998). The Karolinska Directed Emotion Faces – KDEF. http://www.emotionlab.se/resources/kdef

Lux, S., Marshall, J. C., Thimm, M., & Fink, G. R. (2008). Differential processing of hierarchical visual stimuli in young and older healthy adults : Implications for pathology. *Cortex*, *44*, 21–28. http://doi.org/10.1016/j.cortex.2005.08.001

Manly, T., Dobler, V. B., Dodds, C. M., & George, M. A. (2005). Rightward shift in spatial awareness with declining alertness. *Neuropsychologia*, *43*, 1721–1728. http://doi.org/10.1016/j.neuropsychologia.2005.02.009

Martin, M. (1979). Hemispheric specialization for local and global processing. *Neuropsychologia*, *17*, 33–40.

Nicholls, M. E. R., Thomas, N. A., Loetscher, T., & Grimshaw, G. M. (2013). The Flinders Handedness survey (FLANDERS): A brief measure of skilled hand preference. *Cortex*, *49*, 2914–26. http://doi.org/10.1016/j.cortex.2013.02.002

Ocklenburg, S. & Güntürkün, O. (2018). *The Lateralized Brain: The Neuroscience and Evolution of Hemispheric Asymmetries*. Cambridge, MA: Academic Press.

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, *9*, 97–113. http://doi.org/10.1016/0028-3932(71)90067-4

Open Science Collaboration, (2015). Estimating the reproducibility of psychological science. *Science*, *349*, aac4716. http://doi.org/10.1126/science.aac4716

Pashler, H., & Wagenmakers, E. (2012). Editors' introduction to the special section on replicability in psychological science. *Perspectives on Psychological Science*, *7*, 528–530. http://doi.org/10.1177/1745691612465253

Peyrin, C., Chauvin, A., Chokron, S., & Marendaz, C. (2003). Hemispheric specialization for spatial frequency processing in the analysis of natural scenes. *Brain and Cognition*, *53*(2), 278–282. http://doi.org/10.1016/S0278-2626(03)00126-X

Peyrin, C., Chokron, S., Guyader, N., Gout, O., Moret, J., & Marendaz, C. (2006). Neural correlates of spatial frequency processing: A neuropsychological approach. *Brain Research*, *1073–1074*, 1–10. http://doi.org/10.1016/j.brainres.2005.12.051

Peyrin, C., Mermillod, M., Chokron, S., & Marendaz, C. (2006). Effect of temporal constraints on hemispheric asymmetries during spatial frequency processing. *Brain and Cognition*, *62*, 214–20. http://doi.org/10.1016/j.bandc.2006.05.005

Rafal, R. (1998). Neglect. In R. Parasuraman (ed.), *The Attentive Brain*. Cambridge, MA: MIT Press.

Roberson, D., Pak, H., & Hanley, J. R. (2008). Categorical perception of colour in the left and right visual field is verbally mediated: Evidence from Korean. *Cognition*, *107*, 752–62. http://doi.org/10.1016/j.cognition.2007.09.001

Rouder, J. N., & Morey, R. D. (2011). A Bayes factor meta-analysis of Bem's ESP claim. *Psychonomic Bulletin & Review*, *18*, 682–689. http://doi.org/10.3758/s13423-011-0088-7

Schmidt, S. (2009). Shall we really do it again? The powerful concept of replication is neglected in the social sciences. *Review of General Psychology*, *13*, 90–100. http://doi.org/10.1037/a0015108

Sergent, J. (1982). The cerebral balance of power: Confrontation or cooperation? *Journal of Experimental Psychology: Human Perception and Performance*, *8*(2), 253–272. http://doi.org/10.1037/0096-1523.8.2.253

Siok, W. T., Kay, P., Wang, W. S. Y., Chan, A. H. D., Chen, L., Luke, K.-K., & Hai Tan, L. (2009). Language regions of brain are operative in color perception. *Proceedings of the National Academy of Sciences of the United States of America*, *106*, 8140–8145. Retrieved from www.pnas.org?cgi?doi?10.1073?pnas.0903627106

Springer, S. P., & Deutsch, G. (1998). *Left brain, right brain: Perspectives from cognitive neuroscience (5th ed.)* New York, NY: W. H. Freeman/Times Books/ Henry Holt & Co.

Van der Ham, I. J. M., & Borst, G. (2011). Individual differences in spatial relation processing: Effects of strategy, ability, and gender. *Brain and Cognition*, *76*, 184–190. http://doi.org/10.1016/j.bandc.2011.01.008

Van der Ham, I. J. M., & Borst, G. (2016). Corrigendum to "Individual differences in spatial relation processing: Effects of strategy, ability, and gender" [Brain Cogn. (2011) 76, 184–190], *109*, 40–42. http://dx.doi.org/10.1016/j.bandc.2016.08.006

Van der Ham, I. J. M., & Postma, A. (2010). Lateralization of spatial categories: a comparison of verbal and visuospatial categorical relations. *Memory & Cognition*, *38*(5), 582–590. http://doi.org/10.3758/MC.38.5.582

Van der Ham, I. J. M., Raemaekers, M., Van Wezel, R. J. A., Oleksiak, A., & Postma, A. (2009). Categorical and coordinate spatial relations in working memory: an fMRI study. *Brain Research*, *1297*(1987), 70–9.

http://doi.org/10.1016/j.brainres.2009.07.088

Van der Ham, I. J. M., Van Wezel, R. J. A., Oleksiak, A., & Postma, A. (2007). The time course of hemispheric differences in categorical and coordinate spatial processing. *Neuropsychologia*, *45*, 2492–8. http://doi.org/10.1016/j.neuropsychologia.2007.03.021

Van Kleeck, M. H. (1989). Hemispheric differences in global versus local processing of hierarchical visual stimuli by normal subjects: New data and a meta-analysis of previous studies. *Neuropsychologia*, *27*, 1165–1178.

Van Selst, M., & Jolicoeur, P. (1994). A solution to the effect of sample size on outlier elimination. *The Quarterly Journal of Experimental Psychology Section A*, *47*, 631–650. http://doi.org/10.1080/14640749408401131

Vigneau, M., Beaucousin, V., Hervé, P. Y., Duffau, H., Crivello, F., Houde, O., Mazoyer, B., & Tzourio-Mazoyer, N. T. (2005). Meta-analyzing left hemisphere language areas: Phonology, semantics, and sentence processing. *NeuroImage*, *30*, 1414–1432. doi:10.1016/j.neuroimage.2005.11.002

Voyer, D. (1998). On the reliability and validity of noninvasive laterality measures. *Brain and Cognition*, *36*, 209–236. http://doi.org/10.1006/brcg.1997.0953

Voyer, D., Voyer, S. D., & Tramonte, L. (2012). Free-viewing laterality tasks: A multilevel meta-analysis. *Neuropsychology*, *26*, 551–567.

Wagenmakers, E., & Forstmann, B. U. (2014). Rewarding high-power replication research. *Cortex*, *51*, 105–106. http://doi.org/10.1016/j.cortex.2013.09.010

Wager, T. D., Luan Phan, K., Liberzon, I., & Taylor, S. F. (2003). Valence, gender, and lateralization of functional brain anatomy in emotion: A meta-analysis of findings from neuroimaging. *NeuroImage*, *19*, 513–531. https://doi:10.1016/S1053-8119(03)00078-8

Willemin, J., Hausmann, M., Brysbaert, M., Dael, N., Chmetz, F., Fioravera, A., … Mohr, C. (2016). Stability of right visual field advantage in an international lateralized lexical decision task irrespective of participants' sex, handedness or bilingualism. *Laterality: Asymmetries of Body, Brain and Cognition*, *21*, 502–524. http://doi.org/10.1080/1357650X.2015.1130716

Witzel, C., & Gegenfurtner, K. R. (2011). Is there a lateralized category effect for color? *Journal of Vision*, *11*, 1–25. http://doi.org/10.1167/11.12.16

Yovel, G., Wilmer, J. B., & Duchaine, B. (2014). What can individual differences reveal about face processing? *Frontiers in Human Neuroscience*, *8*, 1–9. http://doi.org/10.3389/fnhum.2014.00562

Yovel, G., Yovel, I., & Levy, J. (2001). Hemispheric asymmetries for global and local visual perception: Effects of stimulus and task factors. *Journal of Experimental Psychology: Human Perception and Performance*, *27*, 1369–1385. http://doi.org/10.1037//0096-1523.27.6.1369

Zago, L., Petit, L., Mellet, E., Jobard, G., Crivello, F., Joliot, M., … Tzourio-Mazoyer, N. (2015). The association between hemispheric specialization for language production and for spatial attention depends on left-hand preference strength. *Neuropsychologia*, *93*, 349–406. http://doi.org/10.1016/j.neuropsychologia.2015.11.018