

Neural correlates of unconscious processing in functional magnetic resonance imaging: does brain activity contain more information than can be consciously reported?

Joaquim Streicher ^{1,2,3,*}, Sascha Meyen ⁴, Volker H. Franz ^{4,†}, Timo Stein ^{5,†}

¹Département de neurosciences, Université de Montréal, Pavillon Paul-G.-Desmarais, 2960, chemin de la Tour, local 111, Montréal, Québec H3T 1J4, Canada

²Center for Advanced Research in Sleep Medicine, Hôpital du Sacré-Cœur de Montréal, CIUSSS du Nord-de-l'île-de-Montréal, 5400 Boulevard Gouin Ouest, 5th floor, Wing J, Door 5080, Montréal, Québec H4J 1C5, Canada

³Integrated Trauma Center, Hôpital du Sacré-Cœur de Montréal, CIUSSS du Nord-de-l'île-de-Montréal, 5400, boulevard Gouin Ouest, Montréal, Québec H4J 1C5, Canada

⁴Department of Computer Science, University of Tübingen, Sand 14, 72076 Tübingen, Germany

⁵Department of Psychology, University of Amsterdam, Nieuwe Achtergracht 129-B, 1018 WT Amsterdam, The Netherlands

*Corresponding author. Département de neurosciences, Université de Montréal, Pavillon Paul-G.-Desmarais, C.P. 6128, succ. Centre-ville, Montréal, Québec H3C 3J7, Canada. E-mail: joaquim.streicher@umontreal.ca

†These authors contributed equally and shared last-authorship.

Abstract

A central question of consciousness research is which cognitive processes can occur unconsciously. To investigate this, researchers typically compare participants' ability to consciously discriminate a stimulus to their unconscious processing of the same stimulus (e.g. measured via reaction time or brain activity). If participants are not significantly different from chance in the awareness (or "direct") measure while nevertheless there is a significant effect in the processing (or "indirect") measure, researchers argue that there is no conscious processing of the stimulus, while the stimulus is nevertheless somehow processed, as indicated by the processing measure. In consequence researchers conclude that the stimulus has been processed unconsciously. Using neuroimaging techniques such as functional magnetic resonance imaging (fMRI), researchers then infer which brain regions are involved in unconscious versus conscious processing. However, this methodology is based on a fundamental statistical fallacy that has likely led to an overestimation of the scope of unconscious processing, regarding both its capacity and the brain areas involved. The key problem is that sensitivities in the two measures are never directly compared. Therefore, it is not appropriate to conclude that the processing measure had higher sensitivity than the awareness measure. We reanalyzed the results from 16 fMRI studies directly comparing the sensitivities of both measures in 80 experimental conditions. Our results show that, using this sensitivity comparison method, only eight experimental conditions provide evidence for unconscious processing. These results question the validity of the interpretations commonly drawn in the field.

Keywords: neural correlates of consciousness; fMRI; reanalysis; indirect task advantage; signal detection theory

Introduction

Consciousness is currently one of the most fascinating and challenging topics in cognitive neuroscience and related fields (Seth 2018, Michel et al. 2019, Dehaene et al. 2021). In the last decades, consciousness research has been focused on investigating brain activity associated with conscious processing, or neural correlates of consciousness (NCCs; Crick and Koch 1990, Koch et al. 2016, Rees et al. 2002). In this quest, functional magnetic resonance imaging (fMRI) has been and continues to be one of the most widely used neuroimaging methods, and fMRI studies constitute an important milestone in the consciousness science landscape (Hesselmann 2013). A common approach to investigate NCCs is to test the scope of unconscious processing (Kouider and Dehaene 2007, Lin and He 2008).

The standard reasoning to investigate unconscious processing

A widely used reasoning to isolate unconscious processing is to collect two measures: a measure of stimulus awareness, or *direct task*—e.g. visibility ratings or discrimination responses in a task with a binary response format—and a measure of stimulus processing, or *indirect task*—e.g. reaction time (RT) or brain activity (Hannula et al. 2005, Schmidt and Vorberg 2006, Simons et al. 2007, Sand and Nilsson 2016, Shanks et al. 2021). Note that here we employ the term "indirect task" for consistency with the behavioral priming literature, but that it could be used interchangeably with "indirect measure" or "processing measure," terms that may be more appropriate in the case of neuroimaging studies where participants often do not perform a separate task

Received 24 December 2024; revised 5 October 2025; accepted 9 October 2025

© The Author(s) 2025. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

but indirect measures are taken in parallel. Similarly, the “direct task” is sometimes referred to as “direct measure” or “awareness measure.”

If participants perform at chance when responding to the stimulus directly (typically tested by performing a *t*-test), any effect obtained through the indirect task is considered evidence that the stimulus was processed unconsciously (Stockart et al. 2024). This approach is sometimes referred to as the “double *t*-test” approach, as it consists in performing an initial *t*-test to assess participants’ awareness, followed by a second *t*-test to evaluate unconscious processing (Fig. 1a).

When the indirect task is performed using neuroimaging techniques such as fMRI, significant changes of brain activity or decoding performance serve as evidence that the stimulus was processed by the brain, whereas the initial awareness *t*-test reveals whether participants were able to discriminate the stimulus better than chance or not. One can then allegedly infer which brain regions are involved in the unconscious processing of the stimulus and, by comparison, deduce which brain regions are involved in its conscious processing (i.e. NCCs). Following that methodology, numerous stimuli with different level of cognitive complexity have been suggested to be processed unconsciously: Gabor patches and gratings, objects and tools, words and semantic objects, emotional faces, or body postures (Dehaene et al. 2001, Haynes and Rees 2005, Dannlowski et al. 2007, Hesselmann and Malach 2011, Prochnow et al. 2013, Imamoglu et al. 2014, Ulrich et al. 2014, Suslow et al. 2015, Torralbo et al. 2016, Tettamanti et al. 2017, Zhan et al. 2018, Sheikh et al. 2019). Critically, many studies suggest that high-level cognitive functions can operate without consciousness: memory, language abilities, high-level social skills, or cognitive control (Marois et al. 2004, Kouider et al. 2007, Van Gaal et al. 2010, Freeman et al. 2014, Axelrod et al. 2015, Züst et al. 2015, Rosenthal et al. 2016). The accumulation of results supporting the existence of unconscious perception and cognition in the last decades has shaped our current understanding of unconscious processing and the brain networks it involves, and by extension our definition of the NCCs. But there is a problem with this approach.

Fallacy in the double *t*-test approach

The limits of unconscious processing are highly debated, and a major consideration when testing those limits is the proper manipulation and assessment of participants’ (lack of) awareness (Reingold and Merikle 1990, Schmidt and Vorberg 2006, Stein and Peelen 2021). This is not trivial, and there exist many pitfalls for consciousness researchers: invalid assessment of awareness, biased experimental design, or lack of statistical power (Newell and Shanks 2014, Vadillo et al. 2020, 2022, Stein et al. 2021, 2024, Stein and Peelen 2021, Phillips 2021a,b). Crucially, the standard reasoning to rule out consciousness has been shown to be insufficient to conclude the existence of unconscious processing because it lacks the fundamental statistical test that would allow researchers to conclude that the indirect task outperforms the direct task: the measures of the two tasks are never actually compared (Eriksen 1960, Reingold and Merikle 1988, Schmidt and Vorberg 2006, Meyen et al. 2022).

In the framework of the standard reasoning, researchers rule out awareness when participants are not significantly different from chance when discriminating the stimuli (direct measure), and then assume that significant effects in the indirect measures (e.g. brain activity) were reflecting unconscious processing (cf. Fig. 1a). However, there are two major problems with this reasoning: (i) The finding that the direct task is not significantly

different from chance is not sufficient evidence for the absence of sensitivity (Fallacy 1 in Fig. 1a). To make things worse, the direct task is often severely underpowered and therefore likely to yield false negative results (Vadillo et al. 2016, 2020). (ii) A significant effect in the indirect task is not sufficient evidence to conclude that the indirect task showed relatively good sensitivity to the stimulus—better than the sensitivity in the direct task (Fallacy 2 in Fig. 1a). Quite to the contrary, it has been shown repeatedly that in studies using the standard reasoning, the sensitivity of the indirect task was just as poor as that of the direct task (Franz et al. 2024, Franz and von Luxburg 2015, Meyen et al. 2022, 2024, Schnepf et al. 2022, Zerweck et al. 2021). The underlying reason for all these problems is the fact that the procedure of performing two *t*-tests instead of a direct comparison is incorrect *per se* (e.g. Franz and Gegenfurtner 2008/Appendix B, Gelman and Stern 2006, Nieuwenhuis et al. 2011; similar problems also exist for Bayesian analyses, Palfi and Dienes 2020). All in all, without a statistical comparison of the sensitivities of the two tasks, one cannot conclude that the indirect task was more sensitive to stimulus information than the direct task, and therefore it cannot follow that there is evidence for unconscious processing.

The sensitivity comparison method

To replace the double *t*-test approach of the standard reasoning with a more appropriate approach, an alternative methodology has recently been advocated (Meyen et al. 2022, 2024, Zerweck et al. 2021) which is consistent with previous recommendations (Reingold and Merikle 1988, Schmidt and Vorberg 2006). The goal of this *sensitivity comparison method* (Fig. 1b) is to directly compare the performances in the indirect and direct tasks. To achieve this, the sensitivity to the stimulus is calculated for each task (d' values). This allows testing whether the *sensitivity of the indirect task really outperforms the sensitivity of the direct task*, a situation that was dubbed indirect task advantage (ITA, Meyen et al. 2022).

Note that establishing an ITA is a necessary condition for all the inferences about conscious and unconscious processing that are typically drawn (this purely empirical condition is therefore called Step 1 in Fig. 1b). Only if an ITA is established, we can then go on to draw further-reaching inferences about consciousness and NCCs (Step 2 in Fig. 1b). For example, in this second step, one would need to assess whether the direct task adequately measured consciousness and the indirect task adequately measured unconscious processing. For the purposes of the current article, however, it is sufficient to focus exclusively on Step 1: Can we establish an ITA? We will show that the data of many studies do not seem to support this first step on which all the further-reaching inferences are built.

The sensitivity comparison method has already been practically applied to diverse datasets, namely, unconscious priming (Meyen et al. 2022), unconscious number processing (Zerweck et al. 2021), unconscious contextual cueing (Meyen et al. 2024), unconscious response inhibition (Huang et al. 2023), and event-related potentials analysis in the context of unconscious perception (Schnepf et al. 2022).

The sensitivity comparison method applied to fMRI studies

In this study, we extended the sensitivity comparison method to neuroimaging studies and reanalyzed the results of 16 fMRI studies. Our objective was twofold: investigating the validity of previous interpretations supporting unconscious processing in fMRI studies, and therefore their impact on our definition of the NCCs, and demonstrating that a systematic use of the sensitivity

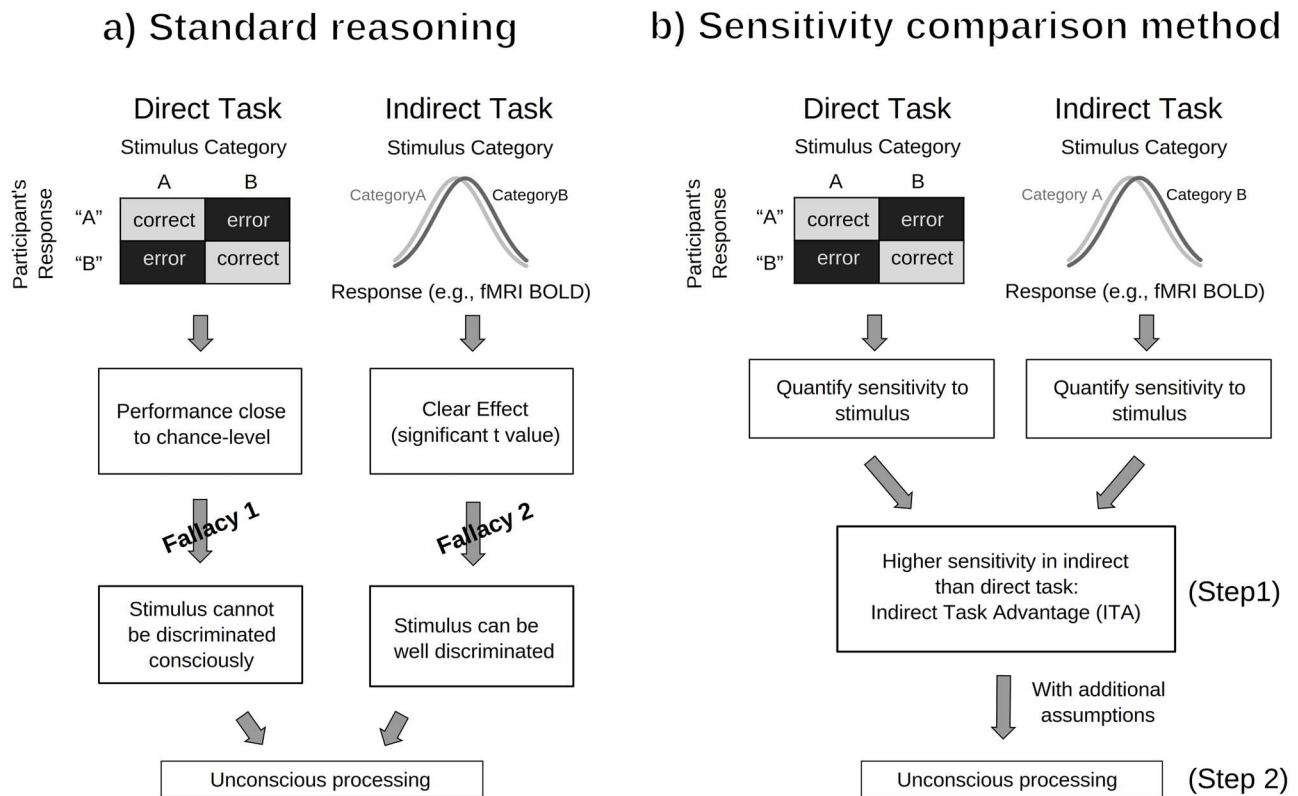


Figure 1. Inferring unconscious processing using neuroimaging: standard reasoning versus sensitivity comparison. Note. Demonstrating an ITA—higher sensitivity in indirect than direct responses—is a prerequisite for the typical, further-reaching inferences about conscious/unconscious processing. (a) In the standard reasoning, or double t-test approach, the direct task is performed to ensure participants' unawareness and any effect obtained through the indirect task serves as evidence that brain data is sensitive to stimulus information. However, there are two fundamental statistical fallacies here. Fallacy 1: the absence of a significant effect for the direct task cannot constitute evidence for the absence of awareness, especially as the direct task is often underpowered. Fallacy 2: a significant effect for the indirect task does not necessarily imply good sensitivity. (b) The appropriate analysis requires a sensitivity comparison and thus an explicit calculation of the indirect task's sensitivity. The sensitivities of the two measures can then be compared to determine if there is an ITA or not (Step 1). If an ITA is confirmed, one must justify that the observed ITA is attributable to unconscious processing rather than some other phenomenon (Step 2). Adapted from Meyen et al. (2022), used under the Creative Commons Attribution 4.0 International License.

comparison method would help future studies to overcome most of the biases currently undermining the field of consciousness research. We did not apply the sensitivity comparison method to the behavioral results of these studies because we wanted to focus solely on the fMRI results (note, however, that the behavioral results of some of those studies were already analyzed in Meyen et al. 2022).

We used a conservative *benefit-of-the-doubt* approach when reanalyzing the results from the selected studies, meaning that our methodological choices always gave an advantage to the indirect task over the direct task, and therefore favored finding an ITA (cf. Meyen et al. 2022). Because we found little evidence for ITAs, this benefit-of-the-doubt approach makes the case even stronger ("although we favored the indirect task, its sensitivity was not larger than that of the direct task"). In a similar vein, we also did not question the validity of the measures employed. That is, we fully accepted the original studies' assumptions that measures matched the intended constructs (e.g. that the direct task measured conscious processing and the indirect task unconscious processing). For example, if the stimulus feature at test in the direct task did not correspond to the feature at test in the indirect task, we still included the study, although the indirect task did not reflect processing of the feature of interest (for a detailed discussion of such problems, see Schmidt and Biafora 2023). Thus, we conducted a purely empirical reanalysis of the data allegedly supporting an ITA (Step 1 in Fig. 1b) and did not assess the validity

of attributing an ITA to unconscious processing (Step 2 in Fig. 1b, more on this in the Discussion section).

Materials and methods

Selection process

To select the studies for the reanalysis, we went through a bottleneck selection process (Fig. 2). We first searched for all fMRI consciousness studies investigating unconscious processing of visual stimuli using two databases (PubMed and Web of Science)—keywords are available in the supplementary material section. Duplicates and meta-analyses were excluded right away, either via the databases' selection criteria or by a manual screening of the articles' titles and abstracts. A small proportion of the articles were found through manual searching, meaning that they were cited in articles we found through the database searching. The filtering tools offered by the two databases partly failed to isolate the studies of interest, and we went through the articles' abstracts to exclude studies not meeting our criteria (no fMRI, no method for rendering a stimulus unconscious). We also decided to exclude patient studies from our corpus, although reanalyzing data from those studies could be interesting in the future. We next read through all the remaining articles to only keep studies with an "objective" measure of awareness (i.e. one that quantified behavioral performance with regard to an objectively measurable stimulus property, as opposed to "subjective" measures that

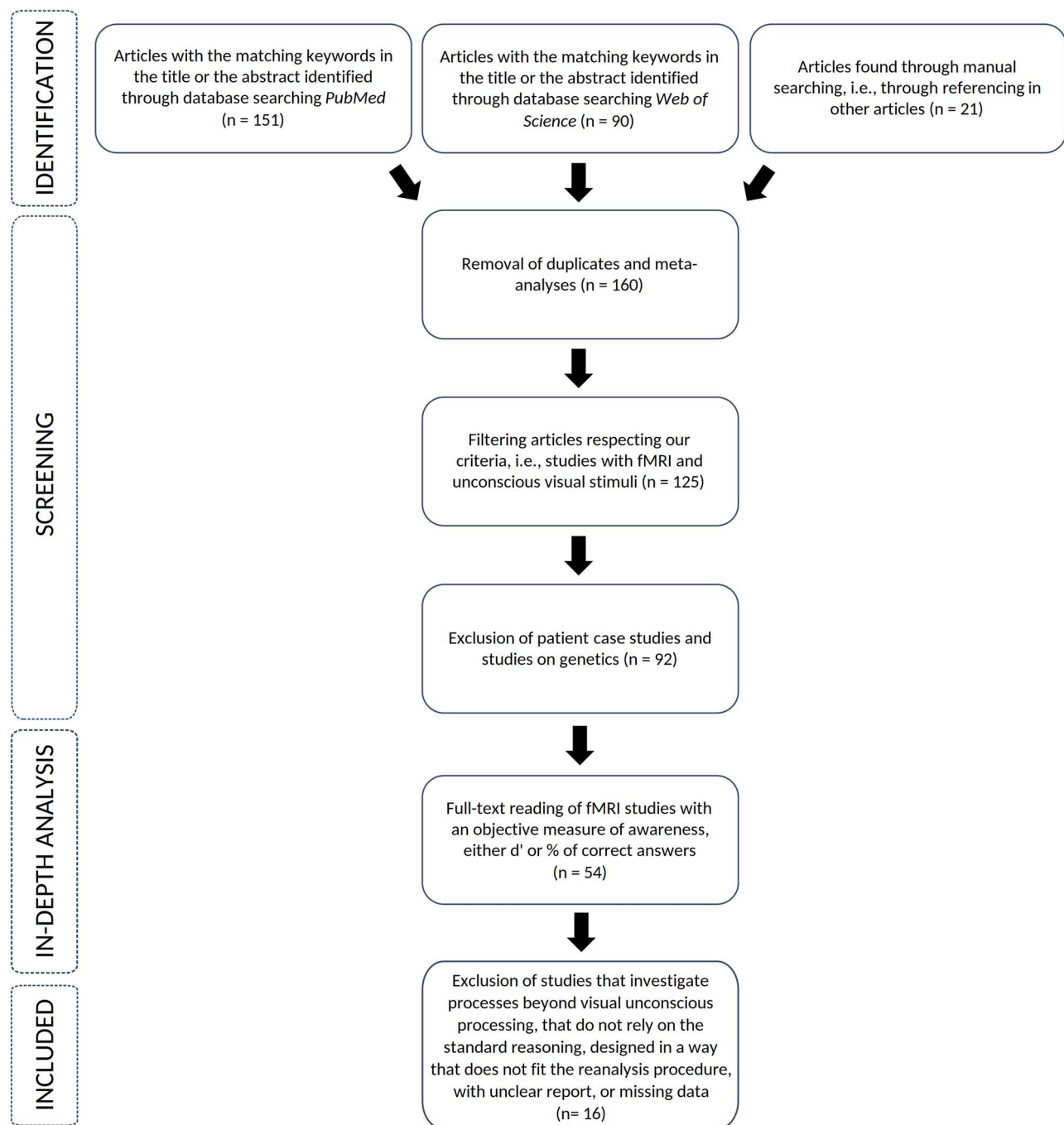


Figure 2. Selection of the studies for the reanalysis. Note. Pipeline for selecting studies for reanalysis. The number of studies remaining after each of the exclusion steps is reported in the figure (n). *Identification.* We used the two databases PubMed and web of science and a manual search to constitute a first corpus. Keywords are available in the supplementary material. *Screening.* We went through the titles and abstracts to remove duplicates, meta-analyses, and articles not fulfilling our main criteria. *In-depth analysis.* Full-text articles from studies with an objective measure of awareness were analyzed. *Included.* Sixteen studies were finally selected for the reanalysis. We excluded studies investigating processes beyond visual unconscious processing and visual perception (e.g. working memory or implicit learning), studies that did not solely base their conclusions on the standard reasoning, studies with a design not allowing us to perform the reanalysis (e.g. correlation analysis), studies from which the design or results were ambiguous, or with missing data necessary for the reanalysis.

rely on participant's unfalsifiable introspective reports such as "seen"/"unseen" judgments), as this was required to perform the reanalysis.

Studies fulfilling all the above-mentioned criteria were analyzed in depth and cut down to 16 studies (Axelrod et al. 2015, Dehaene et al. 2001, Fang et al. 2005, Fogelson et al. 2014, Freeman et al. 2014, Haynes & Rees 2005, Kouider et al. 2007, Kouider et al. 2009, Kouider et al. 2016, Moutoussis & Zeki 2002, Schurger et al.

2010, Stein et al. 2021, Sterzer et al. 2008, Ulrich & Kiefer 2016, van Gaal et al. 2010, Yang et al. 2012) based on five criteria: First, we focused our reanalysis on visual unconscious processing, and studies investigating other processes such as working memory, implicit learning, or negative compatibility effects were excluded. Based on this criterion, we excluded $n = 12$ studies (note that some studies were excluded based on more than one criterion). Second, studies had to claim unconscious processing following the

standard reasoning. Studies that used other rationales or methods (e.g. correlation analysis between the awareness assessment measure and brain activity), were excluded. We excluded $n=13$ additional studies based on this criterion. Third, studies had to present a design that allowed the application of our reanalysis method. Studies that had a behavioral measure that differed from the traditional binary response format (e.g. a direct task with four response alternatives), or that presented fMRI analysis beyond univariate or multivariate analysis (e.g. functional connectivity analysis) were excluded ($n=5$ additional exclusions). Fourth, the designs of the selected studies had to allow for a comparison between direct and indirect tasks, and values had to be reported unambiguously. Studies designed in a way that allowed reanalysis but made conclusions difficult to interpret (e.g. because there was an important conceptual mismatch between the stimulus feature tested in the direct task and the stimulus feature tested in the indirect task) were excluded ($n=3$ additional studies excluded). Fifth, studies had to report enough data to allow for our reanalysis (e.g. d' , number of trials, etc.). We contacted the authors of seven studies that otherwise fulfilled all our inclusion criteria (three studies failed also other criteria such that we did not contact the authors). We received only one reply. Unfortunately, such low return-rates are to be expected (Wicherts et al. 2006). After further deliberation, we managed to reanalyze results from one of these studies based on the summary statistics available online. We therefore excluded $n=5$ additional studies based on this criterion.

Importantly, many of the studies finally selected for the reanalysis presented methodological issues (participant and trial exclusions, inconsistent feature of discrimination, separate task for the direct measure) that probably had an impact on the results of our reanalysis. However, it is important to note that, in agreement with our benefit-of-the-doubt approach, these methodological caveats always biased our reanalysis results toward finding evidence for an ITA in favor of unconscious processing (see Discussion section). Note also that we reanalyzed the most promising results from the 16 selected studies, i.e. significant ROIs or results used by the authors to demonstrate unconscious processing. These 80 experimental conditions had the largest effects and therefore highest indirect sensitivities. We did not reanalyze results that, *a priori*, were not expected to yield ITAs, and we did not conduct additional analyses of alternative candidate ROIs to avoid straw-man arguments. However, different ROIs may potentially yield more ITAs in future reanalyses.

The sensitivity comparison method applied to fMRI data

The sensitivity comparison method (Meyen et al. 2022) consists in estimating a sensitivity value (here d') for the indirect task, such that this sensitivity can be compared to the sensitivity of the direct task (which is typically reported in studies). This makes it possible to determine whether there is an ITA (i.e. whether the sensitivity in the indirect task is superior to the sensitivity in the direct task). When using this method to reanalyze existing studies, one is faced with one problem, though: Typically, one would need the trial-by-trial data to estimate the sensitivity. Because those data are often not available anymore, Meyen et al. (2022) developed a method to estimate the sensitivities from the typically published summary statistics (e.g. the results of a t -test). This method requires only one additional parameter q^2 , which corresponds to the ratio of between-subject to within-subject variability. As such, it is independent of the number of participants and trials, N and K , which is particularly useful when reanalyzing

studies with different N and K . Practically, q^2 can be thought of as a single-trial reliability. It can be mathematically shown to be equal to the variance of participants' underlying individual sensitivities. This parameter q^2 can be estimated from similar studies for which trial-by-trial data are available. Here we used the data from Stein et al. (2021) to estimate this parameter for fMRI data and then used this estimate to reanalyze the selected studies. Stein et al. (2021) was selected to determine q^2 as it was the only study for which we had access to a complete dataset, but also because it seemed to provide us with a representative estimate: This study uses stimuli that are commonly used in research on unconscious perception, pictures of faces and houses. Moreover, the sample size was high ($N=43$) and, compared to the other studies that were selected for our reanalysis, this study presented fewer of the confounding biases we mentioned above.

To estimate the parameter q^2 , we used each participant's observed, individual sensitivity that Stein et al. (2021) had obtained by decoding fMRI data of the lateral occipital cortex (LOC). We chose the LOC results because they seemed most promising to show an ITA. We conducted a Bayesian analysis to determine the highest posterior density interval (Hyndman 1996; Kruschke 2021) for q^2 . Starting with a Jeffreys' prior on q^2 , we employed a Markov Chain Monte Carlo algorithm because it can efficiently estimate posterior distributions with multiple variables making optimal use of each individual sensitivity instead of just the observed standard deviation. We found a posterior mean $q^2 = 0.0061$ ($q = 0.08$) with 95% HDI = [0.001, 0.019]. Note that a simpler, naïve plugin estimation consistently yielded a mean $q^2 = 0.0059$. This ratio for fMRI data is somewhat smaller than what we had estimated for RT data (with a conservative assumption, we had $q^2 = 0.0225$ there). Again, this decrease is plausible due to fMRI measurements incurring more trial-by-trial noise. Thus, we used for all our subsequent reanalyses of fMRI data the parameter estimate $q^2 = 0.006$. This corresponds to assuming that the standard deviation of individual, underlying sensitivities is $q = 0.08$ (observed sensitivities may vary more due to sampling noise). We also estimated q^2 based on the other conditions of Stein et al. (2021), namely fMRI measures at different time points during the trials (at 0, 1.6, 3.2, 4.8, 6.4, and 8.0 s after stimulus onset). Estimates of q^2 in the objectively invisible condition were on average $q^2 = 0.005$ and in the subjectively invisible condition $q^2 = 0.004$ (see Supplementary Fig. S1). The largest estimate we found in any of the conditions was $q^2 = 0.008$, which is likely an overestimate. Had we chosen this value instead of $q^2 = 0.006$ for our reanalyses, the estimated indirect task sensitivities would only change marginally (e.g. from $d' = 0.147$ to $d' = 0.152$ in the first reanalyzed condition of Dehaene et al. 2001).

With this new q^2 , we applied the sensitivity comparison method to the 16 fMRI consciousness studies that were selected for our reanalysis by estimating d' from published summary statistics given minimal assumptions. The following formula summarizes how d' was estimated for the indirect task:

$$d'_{\text{indirect}} = t \cdot c_{N,K,q^2} \quad (1)$$

where the d' value for the indirect task—here parameter estimates from a general linear model fitted to the fMRI data or multivariate pattern analysis (MVPA) values—is estimated based on the reported t -value of a paired t -test comparing two conditions with different neural activation. The constant c_{N,K,q^2} adjusts for the increase in t values that occurs with a greater number of

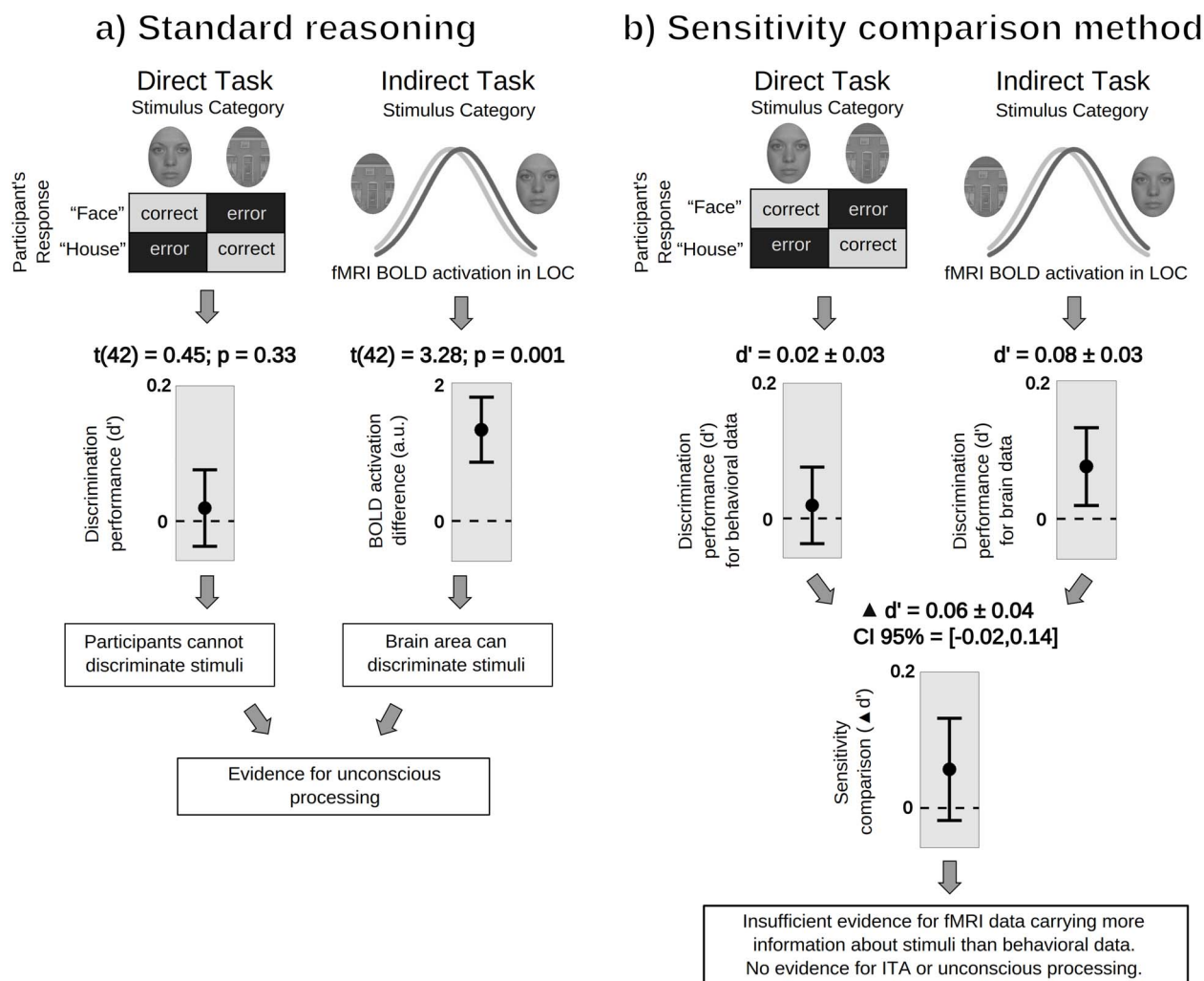


Figure 3. Standard reasoning versus sensitivity comparison method: Example data from Stein et al. (2021). Note. Example of the two procedures in Stein et al. (2021) for the univariate analysis of the LOC. (a) Erroneous double t-test approach (“standard reasoning”): Consciousness researchers usually test participants’ ability to discriminate stimuli, here faces and houses, by computing a d' value and testing it against zero. If d' is not significantly different from zero ($P > .05$), participants are considered unaware of the stimuli’s features of interest. On the other hand, a significant difference in brain activity, here BOLD variations in LOC, is then considered as evidence that the stimuli’s features of interest are processed unconsciously. This standard reasoning is, however, flawed (cf. Fig. 1). (b) Appropriate sensitivity comparison method: A d' value is computed for both behavioral and brain measures and the two values are directly compared. Here there is no significant difference between the two sensitivity values, suggesting that there is insufficient evidence for brain data being more sensitive to stimulus information than behavioral data. Note that the sensitivity for the indirect task is significantly above chance, but that this test alone is not enough to conclude that the sensitivity of the indirect task outperforms the sensitivity of the direct task. Brain data should contain more information relative to the stimulus than behavioral data for one to conclude that there is an ITA and, potentially, evidence for unconscious processing. Error bars depict 95% CIs. The BOLD activation plot in the top row is only for illustrative purposes and not based on data.

participants (N), a higher number of trials (K), and it accounts for the dependency on the ratio of between-subject to within-subject variance (q^2). This scaling factor c_{N,K,q^2} was computed for each reanalyzed study and condition separately, where we used the respective N and K but always the same $q^2 = 0.006$. When t values were missing, we transformed z scores using quantile mapping: We first computed the associated p -value and then the associated t value based on the number of degrees of freedom. When values were not clearly reported or when there seemed to be ambiguity, we always selected the value giving an advantage to the unconscious processing conclusion, in order to follow our benefit-of-the-doubt approach. Analyses can be performed by using our website (<http://www.ecogsci.cs.uni-tuebingen.de/ITAcaculator/>) or, for more flexibility, the R code available on OSF (<https://osf.io/wnfta/>).

Results

We first provide an example by applying the sensitivity comparison method to data from Stein et al. (2021)—see Fig. 3. We then give a summary of all results from our reanalysis (Fig. 4).

The sensitivity comparison method applied to Stein et al. (2021)

In the study of Stein et al. (2021), 43 participants were presented with masked house and face stimuli (200 trials per participant). Participants discriminated the presented stimulus in each trial (direct task). Stein and colleagues reported a d' of 0.02 ($SE = 0.038$) for the direct task and $t(42) = 3.28$ for MVPA decoding of faces versus houses in LOC (highest reported value for the indirect task) and—based on the standard reasoning—infer an ITA.

From this published t -value, we estimated for the indirect task a sensitivity of $d' = 0.08$ ($SE = 0.03$). This resulted in a d' difference between indirect and direct tasks of 0.06 ($SE = 0.04$), with a 95% confidence interval (CI) including zero (95% CI = $[-0.02, 0.14]$), thereby indicating that the sensitivity-difference did not deviate significantly from zero. That is, although the sensitivity of the indirect task was slightly larger than that of the direct task, this small difference can be explained by noise: There was not enough evidence to support an ITA based on the reported statistics. These results are depicted in detail in Fig. 3 and in a more compact fashion in Fig. 4. This is the analysis we applied to all included studies (see below).

To demonstrate that the sensitivity comparison method is in fact capable of detecting ITAs, we conducted an additional analysis where we contrasted conditions of Stein et al. (2021) for which it was *a priori* plausible to expect a difference. For this, we compared the LOC activity in the “objectively visible condition” ($t = 16.69$; $d' = 0.4$; $SE = 0.05$) with the direct task performance of the “objectively invisible condition” ($d' = 0.02$, $SE = 0.04$). As a result, we found a difference of 0.38 ($SE = 0.06$; 95% CI = $[0.26; 0.50]$). This shows that our method is in principle able to detect an ITA: If processing is sufficiently strong in the indirect task, sensitivity differences can be detected. But note that, in the “objectively visible condition,” participants’ direct task sensitivity was also higher with $d' = 4.5$ so that no evidence for unconscious processing is given by this comparison either.

The sensitivity comparison method applied to all studies

Next, we applied the sensitivity comparison method to all selected studies and depict the results in Fig. 4. Inspection of this figure shows that in most conditions, the sensitivity difference was close to zero and not significantly different from zero (error bars correspond to 95% CIs).

Only eight of the 80 experimental conditions (10%) showed a significant ITA, while two conditions (2.5%) showed a significant opposite effect (dubbed direct task advantage; DTA; in the figure). Note that one would *a priori* expect a false positive rate of 5%, and that we did not correct for multiple hypothesis testing. So, overall, there seems little evidence for an ITA across all studies, although all studies inferred unconscious processing based on their reported results. With that, a parsimonious explanation of most of the data is that weak, residual conscious processing underlies responses. In this majority of the cases, there is not enough evidence for processing with a sensitivity beyond that in participants’ direct (conscious) responses. There was insufficient empirical basis for interpretations about unconscious processing.

Nevertheless, it is instructive to have a closer look at the eight conditions that supported an ITA. They came from three studies (Haynes and Rees 2005, Sterzer et al. 2008, Freeman et al. 2014), including two studies using MVPA as the indirect task (Haynes and Rees 2005, Sterzer et al. 2008). We discuss one of them as an illustration in the following. A description of all reanalyzed studies, the data used for the reanalysis such as reported statistics, number of participants, and number of trials as well as all detailed numeric results are available in the supplementary materials (Supplementary Tables S1–S16).

Sterzer et al. (2008) reported a d' of -0.05 ($SE = 0.14$) for the direct task based on data from five participants (77 trials per participant) performing a binary discrimination task with face and house stimuli rendered invisible using continuous flash suppression (CFS). Based on the reported statistics from their decoding analysis, we estimated a d' of 0.69 ($SE = 0.18$) for the indirect

task (highest decoding accuracy was found when using both FFA and PPA ROIs data: 63.5%), resulting in a sensitivity difference of 0.74 ($SE = 0.23$)—this was the highest sensitivity difference we found across all experimental conditions of all studies. The 95% CI excluded zero (CI = $[0.11, 1.37]$), supporting the existence of an ITA.

Discussion

Our reanalysis evaluated evidence for unconscious processing and its neural correlates in 16 fMRI studies with 80 experimental conditions. We applied our sensitivity comparison reanalysis method on their summary statistics and thereby replaced the problematic standard reasoning (Fig. 1a) with a more appropriate method (Fig. 1b). In most conditions (87.5% of all conditions), there was no significant difference between sensitivities in indirect and direct tasks, 2.5% of conditions showed higher sensitivity for the direct task, and only 10% of conditions showed a larger sensitivity difference in the indirect task than in the direct task—that is an ITA (Fig. 4). However, results from all these conditions had been considered sufficient evidence for unconscious processing in the original studies. As an ITA is a necessary step for further-reaching inferences about conscious and non-conscious processing, a lack of clear evidence for an ITA constitutes a problem for such inferences.

These results suggest that the evidence for unconscious processing has often been overestimated in consciousness studies using fMRI. This is a similar situation as has been found for behavioral priming (Zerweck et al. 2021, Meyen et al. 2022), priming in EEG (Schnepf et al. 2022), implicit learning (Meyen et al. 2024), and lie detection (Franz and von Luxburg 2015, Franz et al. 2024). Note that some of the here reanalyzed fMRI studies also presented behavioral data. The analysis of these behavioral data (performed in Meyen et al. 2022) is in agreement with our results from reanalyzing the fMRI data here: There is little evidence for ITAs in either measure (Meyen et al. 2022). Without establishing an ITA one cannot infer that there were unconscious processing of a stimulus beyond what participants can consciously report. In other words, based on our results it seems likely that brain activity has often been prematurely associated with unconscious processing, which in turn has consequences for our definition and understanding of the NCCs. In the following we discuss some further, open issues.

Methodological biases and their impact on ITAs

While our reanalysis did not indicate convincing evidence for an ITA in most studies and conditions, we now want to discuss those few conditions where our reanalysis did find evidence for an ITA. What can we conclude for those conditions? Do these ITAs reflect unconscious processing (Step 2 in Fig. 1b)? Unfortunately, some common methodological flaws might bias consciousness studies toward finding evidence for unconscious processing. For example, there often exists a statistical power asymmetry between the two tasks, the direct task being performed with less participants or trials. In fMRI studies, this asymmetry is typically exacerbated by the high number of data points (voxels) in fMRI as well as by the nature of the data: continuous responses for the brain (e.g. BOLD activation of the amygdala) are often compared to binary responses for the participants (e.g. “fearful” versus “neutral”), thus giving the indirect task an advantage in terms of statistical power compared to the direct task.

Additionally, it is common in fMRI studies to have participants perform the two tasks at different points in time, in different contexts, e.g. inside and outside the scanner, or even to have

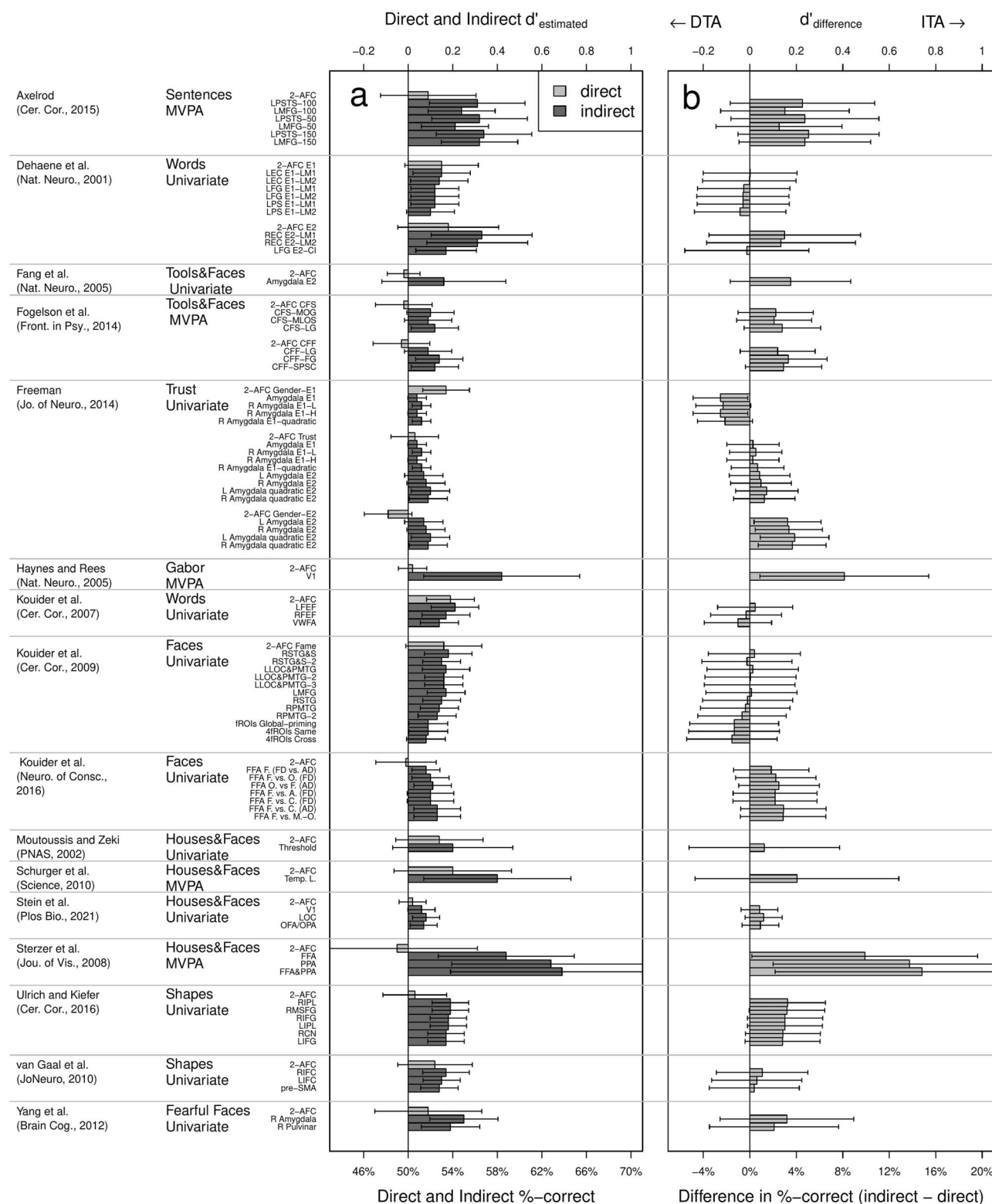


Figure 4. Sensitivity comparison in 16 fMRI studies investigating unconscious processing. *Note.* (a) Sensitivity for the direct and indirect tasks. (b) Sensitivity comparison between the two tasks. The difference between the sensitivities of the two tasks is reported on the horizontal axis. 95% CIs are reported using error bars. The data used to perform the reanalysis is available in [Supplementary Tables S1–S16](#). DTA is short for direct task advantage, i.e. when there is evidence that the direct task outperforms the indirect task. Acronyms and abbreviations from the different conditions are available in the supplementary material.

different participants for the two measures. Although this might not always give a spurious advantage to the indirect task, using dissimilar data collection procedures is problematic if one intends to compare the two tasks. Moreover, participants typically have to respond at a designated time window after stimulus presentation, while the brain is measured at the time of stimulus presentation. This creates another asymmetry: in the direct task, participants are burdened by short-term memory demands because information has to be maintained until report, while the indirect task occurs simultaneously (and sometimes the best time points are selected for reporting results).

Some studies use MVPA, an approach in which the information available in brain activity is used in an optimal way to classify stimuli. Comparing these classifications with participants' direct task classifications requires some additional considerations: First, training a classifier after all trials have been recorded (offline) grants an advantage over participants in the direct task, who have to give responses in an online fashion—participants have to set internal decision thresholds on the fly, which is more difficult than when the threshold is set *post hoc* based on all trials. Second, determining which features of the stimuli allow for optimal classification may only be learned by participants during the experiment, while a trained classifier can optimally weight the neural activity representing those features *post hoc*. Finally, in all fMRI studies, the exclusion of participants or scans, or the formation of two groups *post hoc*, can lead to regression to the mean effects (Barnett et al. 2005, Schmidt 2015, Shanks 2017).

Three studies which did not pass our criteria for inclusion in the reanalysis deserve to be mentioned here, as they complete the overall picture. One study by Cao et al. (2021) did perform a sensitivity comparison (and was therefore not included because no reanalysis was necessary). The authors investigated unconscious processing of facial identity. They presented two faces to participants—either two actors' faces or two familiar faces—morphed to different degrees and masked by CFS. Facial identity was decoded from fMRI activity in the right fusiform face area and directly compared to participants' face identity recognition performance (direct task). The authors found only anecdotal evidence for an ITA ($t(15) = 1.85$, $P = .042$ in a one-tailed test, $BF = 2.033$; Cao et al. 2021). A replication of these results would be helpful to assess the robustness of this effect and to clarify whether facial identity might be a promising line of future research on unconscious processing. Two other studies investigated the unconscious processing of object-context relations (Faivre et al. 2019) and emotional words (Hoffmann et al. 2015). Both studies followed the standard reasoning. However, both reported null findings and therefore did not even claim unconscious processing (this is why we did not include them in our reanalysis). Taken together, their results also indicate that the evidence for ITAs is scarce.

In our reanalysis, studies with and without confirmed ITA are affected by some of the above-mentioned caveats and may thereby result in a biased comparison between direct and indirect tasks in favor of the latter one. This gives more weight to our results. Although studies were affected by these biases (for which we did not correct) inflating evidence for ITAs, we still found little evidence for ITAs overall.

The role of statistics in neuroimaging studies investigating unconscious processing

A frequent objection of consciousness researchers to the sensitivity comparison method applied in the present reanalysis is that it were underpowered (Stockart et al. 2024). However, demonstrating unconscious processing based on the results of two tasks logically

requires an appropriate comparison. If this comparison then turns out to lack statistical power then we should not return to the problematic standard reasoning. Instead, we must first reevaluate the strength of evidence from previous studies and second aim to increase power in future studies. In many of our reanalyses, the direct task sensitivity lacked precision (larger direct than indirect error bars in Fig. 4a). But the precision of a comparison between two measures is always limited by the lower precision of the two measures. Thus, the demand for higher power directly translates into increasing the number of participants and trials in the direct task. Note that this improvement would even be relatively cheap, given that we are talking about the behavioral task not the indirect task, where imaging is involved. Another option for future studies is to include continuous measures in the direct task to then compare the information in both tasks based on their continuous responses.

Limitations of the present work

There are some limitations inherent to the present reanalysis. First, our reanalysis procedure requires objective measures of awareness, and we did not consider the case of subjective awareness measures (Seth et al. 2008, Kiefer et al. 2023). Claims about subjectively-unconscious processing, if established with appropriate subjective measures, are not targeted by our sensitivity comparison critique. Second, the estimate of the parameter q^2 was based on a single fMRI study (Stein et al. 2021), which might not be representative of the whole field. Our study provides a first estimate that might evolve in future studies. However, since Stein et al. (2021) had a comparatively large sample size ($N = 43$), it is likely to constitute an appropriate estimation. Third, fMRI data is complex, and we focused on studies with a traditional design (2-class MVPA, 2-conditions t values or F values from ANOVAs). Since this is the first time such a reanalysis has been performed with fMRI data, extending to all types of design would require additional adaptations of the reanalysis procedure. Similarly, extensions of the sensitivity comparison method should include direct tasks with more than two response alternatives. Fourth, we decided to focus our reanalysis on results interpreted by the authors of the reanalyzed studies as evidence for unconscious processing. Therefore, we did not reanalyze results that, *a priori*, were not expected to yield ITAs. The specificity of our method could thus be further validated in the future by conducting additional analyses of alternative candidate ROIs. Fifth, our approach only used the reported summary statistics—due to the well-known problems of obtaining raw-data from published studies (Wicherts et al. 2006). Performing a sensitivity comparison based on the individual trial-by-trial data could increase statistical power. Sixth, one may criticize that transforming the continuous indirect task responses into binary predictions for each trial in order to compute sensitivity from hit rates and false alarm rates artificially reduces statistical power. Indeed, dichotomization reduces power. But any comparison is necessarily limited by the weakest comparand: If the direct task is by design based on binary responses, an appropriate comparison must use dichotomized indirect task responses. Our argument is that, if the original studies infer unconscious processing based on contrasting the indirect task results to the direct task data measured on a binary scale, the indirect task data should show a difference when brought to that scale. Seventh, some studies did not test the right feature of discrimination, i.e. brain and behavioral datasets were not tested on the same information, such that comparing their sensitivity can be problematic (Reingold and Merikle 1988, Schmidt and Vorberg 2006, Schmidt and Biafora

2023). We recommend future studies to avoid such direct–indirect mismatch by testing the feature of interest in both the direct and indirect task and follow the criteria we mentioned in the previous section (same number of participants and trials, no *post hoc* exclusion, experimental design as similar as possible) to avoid an unfair comparison between direct and indirect tasks (in line with recent recommendations, Stockart et al. 2024). Although some exclusion criteria are clear-cut (e.g. missing data necessary for the reanalysis), others might seem debatable. For example, we decided to focus on visual unconscious processing but future studies could include studies investigating more complex cognitive processes, such as working memory or implicit learning. We also excluded studies with what we considered too heavy methodological biases or for which values were not clearly reported to avoid obtaining non-interpretable or invalid results from the reanalysis. Future reanalyses might offer complementary findings by reanalyzing different sets of studies based on different selection criteria. Finally, although it seems likely that the estimated q^2 value will be within a reasonable range, it might still vary across laboratories and designs. Future work should estimate q^2 using a diversity of fMRI datasets and designs to further improve the generalizability of our conclusions.

Future directions: investigating the NCCs using the sensitivity comparison method

Our approach can be seen as providing a focus on the most promising designs and methods: Based on this first attempt of reanalysis of a very limited corpus, it appears that studies using MVPA analysis in visual areas (V1, FFA, PPA) were the most potent in confirming an ITA (stimulus type: faces/houses, gratings; suppression method: CFS, sandwich masking). Future studies that will survive the sensitivity comparison will constitute good candidates to define the extent of unconscious processing and identify neural correlates of consciousness.

Current trends in the field as reported in the UnconTrust database (in prep, <https://uncontrustdb.tau.ac.il/>) suggest that, although there has been an increase in good practices in recent years, suboptimal methodological choices remain a widespread problem (e.g. post-experiment awareness measure, direct–indirect mismatch). Apart from following the criteria we discussed earlier in this work (see also Rothkirch and Hesselmann 2017, Stein et al. 2024, Stockart et al. 2024) we recommend future consciousness studies using fMRI or other neuroimaging techniques to quantitatively compare brain data to behavioral data directly on the same scale—e.g. with d' values, but other tools could be used to compare the two tasks, for instance tools from information theory (Ince et al. 2017, Mediano et al. 2022, Meyen 2022). A systematic use of the sensitivity comparison method will lead to better practice and improved methodology within consciousness research. For example, it is likely that researchers will be more attentive to having enough statistical power in their measure of stimulus awareness, and that more effort will be made to have the two measures resulting from procedures as similar as possible, so that they are comparable. Having brain activity as an indirect measure makes the comparison of sensitivity less straightforward than with behavioral priming (as in Meyen et al. 2022), and having two different quantitative data types (binomial, univariate or multivariate analyses) on two different scales also makes the comparison difficult. The fact that data type and statistical tests vary between the two measures reinforces the need for a comparison of sensitivity on the same scale. Therefore, it will be important to develop additional methods to directly compare brain and behavioral data (e.g. Ince et al. 2017). As a

final note, we stress once again that an ITA alone is not sufficient to conclude unconscious processing but that it only represents a necessary first step. Other aspects of validity must be addressed in a second step to make claims about unconscious processing. Future studies are needed to confirm the results from this first reanalysis, and a systematic meta-analysis using the sensitivity comparison could be informative regarding the types of stimuli that yield an effect supporting unconscious processing, as well as the brain networks involved.

Conclusions

By showing that a large proportion of the fMRI consciousness studies we reanalyzed fail in providing evidence for an ITA, we call for a methodological update of the field of consciousness science. After Meyen et al. (2022) had shown that behavioral priming effects from many consciousness studies are most likely due to statistical artifacts, we expanded this reasoning to studies using fMRI data as an indirect measure. We argue that, to determine that brain activity reflects unconscious stimulus processing, one should provide evidence that brain data is more sensitive to stimulus information than behavioral data from the awareness assessment. However, based on our results, we argue that brain data rarely outperforms behavioral data. Together with the additional methodological flaws undermining the field, the results from our reanalysis put common interpretations in consciousness research into question. We recommend future studies investigating unconscious processing using fMRI and other neuroimaging techniques to support their claims with a direct comparison (of sensitivities) between brain and behavioral data. Such profound methodological changes in the field would contribute to a better understanding of unconscious processing and would provide us with a more accurate working definition of the neural correlates of consciousness.

Author contributions

Joaquim Streicher (Conceptualization [equal], Data curation [lead], Formal analysis, Investigation, Methodology, Project administration [equal], Visualization [lead], Writing—original draft, Writing—review & editing [equal]), Sascha Meyen (Conceptualization [equal], Data curation [supporting], Formal analysis, Methodology [equal], Visualization [supporting], Writing—review & editing [equal]), Volker H. Franz (Conceptualization, Methodology, Resources, Supervision, Writing—review & editing [equal]), and Timo Stein (Conceptualization, Methodology, Supervision, Writing—review & editing [equal])

Supplementary data

Supplementary data is available at *Neuroscience of Consciousness* online.

Conflict of interest

None declared.

Funding

This work was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)—project number 276693517—SFB 1233 (Robust Vision); VHF is a member of the

Machine Learning Cluster of Excellence, EXC number 2064/1—project number 390727645.

Data availability

Full results from the reanalysis are available in the supplementary materials. The analysis code with the data underlying this article is available in a repository on Open Science Framework, at <https://osf.io/wnfta/>.

References

- Axelrod V, Bar M, Rees G et al. Neural correlates of subliminal language processing. *Cereb Cortex* 2015;**25**:2160–9. <https://doi.org/10.1093/CERCOR/BHU022>
- Barnett AG, Van Der Pols JC, Dobson AJ. Regression to the mean: what it is and how to deal with it. *Int J Epidemiol* 2005;**34**:215–20. <https://doi.org/10.1093/ije/dyh299>
- Cao R, Qian C, Ren S et al. Visual adaptation and 7T fMRI reveal facial identity processing in the human brain under shallow interocular suppression. *Neuro Image* 2021;**244**:118622. <https://doi.org/10.1016/j.neuroimage.2021.118622>
- Crick F, Koch C. Towards a neurobiological theory of consciousness. *Semin Neurosci* 1990;**2**:263–75.
- Dannlowski U, Ohrmann P, Bauer J et al. Amygdala reactivity predicts automatic negative evaluations for facial emotions. *Psychiatry Res Neuroimaging* 2007;**154**:13–20. <https://doi.org/10.1016/j.psychres.2006.05.005>
- Dehaene S, Naccache L, Cohen L et al. Cerebral mechanisms of word masking and unconscious repetition priming. *Nat Neurosci* 2001;**4**:752–8. <https://doi.org/10.1038/89551>
- Dehaene S, Lau H, Kouider S. What is consciousness, and could machines have it? In: von Braun JS, Archer M, Reichberg GM, et al. (eds.), *Robotics, AI, and Humanity*. Cham: Springer, 2021, 43–56. https://doi.org/10.1007/978-3-030-54173-6_4
- Eriksen CW. Discrimination and learning without awareness: a methodological survey and evaluation. *Psychol Rev* 1960;**67**:279–300. <https://doi.org/10.1037/h0041622>
- Faivre N, Dubois J, Schwartz N et al. Imaging object-scene relations processing in visible and invisible natural scenes. *Sci Rep* 2019;**9**:4567. <https://doi.org/10.1038/s41598-019-38654-z>
- Fang F, He S. Cortical responses to invisible objects in the human dorsal and ventral pathways. *Nat Neurosci* 2005;**8**:1380–5. <https://doi.org/10.1038/nm1537>
- Franz VH, Gegenfurtner KR. Grasping visual illusions: consistent data and no dissociation. *Cogn Neuropsychol* 2008;**25**:920–50. <https://doi.org/10.1080/02643290701862449>
- Franz VH, Meyen S, von Luxburg U. Technical comment on Gunderson, ten Brinke, and Sokol-Hessner (2023). When the body knows: interoceptive accuracy enhances physiological but not explicit differentiation between liars and truth-tellers. *Personality and individual differences*, 204, 112039. *Personal Individ Differ* 2024;**217**:112439. <https://doi.org/10.1016/j.paid.2023.112439>
- Franz VH, von Luxburg U. No evidence for unconscious lie detection: a significant difference does not imply accurate classification. *Psychol Sci* 2015;**26**:1646–8. <https://doi.org/10.1177/0956797615597333>
- Freeman JB, Stoller RM, Ingbreten ZA et al. Amygdala responsivity to high-level social information from unseen faces. *J Neurosci* 2014;**34**:10573–81. <https://doi.org/10.1523/JNEUROSCI.5063-13.2014>
- Gelman A, Stern H. The difference between “significant” and “not significant” is not itself statistically significant. *Am Stat* 2006;**60**:328–31. <https://doi.org/10.1198/000313006X152649>
- Hannula DE, Simons DJ, Cohen NJ. Imaging implicit perception: promise and pitfalls. *Nat Rev Neurosci* 2005;**6**:247–55. <https://doi.org/10.1038/nrn1630>
- Haynes J-D, Rees G. Predicting the orientation of invisible stimuli from activity in human primary visual cortex. *Nat Neurosci* 2005;**8**:686–91. <https://doi.org/10.1038/nrn1445>
- Hesselmann G. Dissecting visual awareness with fMRI. *Neuroscientist* 2013;**19**:495–508. <https://doi.org/10.1177/1073858413485988>
- Hesselmann G, Malach R. The link between fMRI-BOLD activation and perceptual awareness is “stream-invariant” in the human visual system. *Cereb Cortex* 2011;**21**:2829–37. <https://doi.org/10.1093/CERCOR/BHR085>
- Hoffmann M, Mothes-Lasch M, Miltner WH et al. Brain activation to briefly presented emotional words: effects of stimulus awareness. *Hum Brain Mapp* 2015;**36**:655–65. <https://doi.org/10.1002/hbm.22654>
- Huang Q, Jiang R, Mao X et al. Does response inhibition occur unconsciously? A systematic review and Bayesian meta-analysis. *Conscious Cogn* 2023;**115**:103570. <https://doi.org/10.1016/j.concog.2023.103570>
- Hyndman RJ. Computing and graphing highest density regions. *Am Stat* 1996;**50**:120–6. <https://doi.org/10.1080/00031305.1996.10474359>
- Imamoglu F, Heinze J, Imfeld A et al. Activity in high-level brain regions reflects visibility of low-level stimuli. *Neuro Image* 2014;**102**:688–94. <https://doi.org/10.1016/j.neuroimage.2014.08.045>
- Ince RA, Giordano BL, Kayser C et al. A statistical framework for neuroimaging data analysis based on mutual information estimated via a Gaussian copula. *Hum Brain Mapp* 2017;**38**:1541–73. <https://doi.org/10.1002/hbm.23471>
- Kiefer M, Frühauf V, Kammer T. Subjective and objective measures of visual awareness converge. *PLoS One* 2023;**18**:e0292438. <https://doi.org/10.1371/journal.pone.0292438>
- Koch C, Massimini M, Boly M et al. Neural correlates of consciousness: progress and problems. *Nat Rev Neurosci* 2016;**17**:307–21. <https://doi.org/10.1038/nrn.2016.22>
- Kouider S, Barbot A, Madsen KH et al. Task relevance differentially shapes ventral visual stream sensitivity to visible and invisible faces. *Neurosci Conscious* 2016;**2016**:niw021. <https://doi.org/10.1093/nc/niw021>
- Kouider S, Dehaene S. Levels of processing during non-conscious perception: a critical review of visual masking. *Philos Trans R Soc B Biol Sci* 2007;**362**:857–75. <https://doi.org/10.1098/rstb.2007.2093>
- Kouider S, Dehaene S, Jobert A et al. Cerebral bases of subliminal and supraliminal priming during reading. *Cereb Cortex* 2007;**17**:2019–29. <https://doi.org/10.1093/CERCOR/BHL110>
- Kruschke JK. Bayesian analysis reporting guidelines. *Nat Hum Behav* 2021;**5**:1282–91. <https://doi.org/10.1038/s41562-021-01177-7>
- Lin Z, He S. Seeing the invisible: the scope and limits of unconscious processing in binocular rivalry. *Nat Prec* 2008;1–1. <https://doi.org/10.1038/npre.2008.2246.1>
- Marois R, Yi D-J, Chun MM. The neural fate of consciously perceived and missed events in the attentional blink. *Neuron* 2004;**41**:465–72. [https://doi.org/10.1016/S0896-6273\(04\)00012-1](https://doi.org/10.1016/S0896-6273(04)00012-1)
- Mediano PA, Rosas FE, Bor D et al. The strength of weak integrated information theory. *Trends Cogn Sci* 2022;**26**:646–55. <https://doi.org/10.1016/j.tics.2022.04.008>
- Meyen S. Information-theoretic perspectives on unconscious priming and group decisions: Retrieving maximum information from human

- responses Ph.D. Thesis, Universität Tübingen, 2022. <https://ub01.uni-tuebingen.de/xmlui/handle/10900/125321>.
- Meyen S, Vadillo MA, von Luxburg U et al. No evidence for contextual cueing beyond explicit recognition. *Psychon Bull Rev* 2024;**31**: 907–30. <https://doi.org/10.3758/s13423-023-02358-3>
- Meyen S, Zerweck IA, Amado C et al. Advancing research on unconscious priming: when can scientists claim an indirect task advantage? *J Exp Psychol Gen* 2022;**151**:65–81. <https://doi.org/10.1037/xge0001065>
- Michel M, Beck D, Block N et al. Opportunities and challenges for a maturing science of consciousness. *Nat Hum Behav* 2019;**3**:104–7. <https://doi.org/10.1038/s41562-019-0531-8>
- Moutoussis K, Zeki S. The relationship between cortical activation and perception investigated with invisible stimuli. *Proc Natl Acad Sci* 2002;**99**:9527–32. <https://doi.org/10.1073/pnas.142305699>
- Newell BR, Shanks DR. Unconscious influences on decision making: a critical review. *Behav Brain Sci* 2014;**37**:1–19. <https://doi.org/10.1017/S0140525X12003214>
- Nieuwenhuis S, Forstmann BU, Wagenmakers E-J. Erroneous analyses of interactions in neuroscience: a problem of significance. *Nat Neurosci* 2011;**14**:Article 9. <https://doi.org/10.1038/nn.2886>
- Palfi B, Dienes Z. Why Bayesian “evidence for H_1 ” in one condition and Bayesian “evidence for H_0 ” in another condition does not mean good-enough Bayesian evidence for a difference between the conditions. *Adv Methods Pract Psychol Sci* 2020;**3**:300–8. <https://doi.org/10.1177/2515245920913019>
- Phillips I. Blindsight is qualitatively degraded conscious vision. *Psychol Rev* 2021a;**128**:558–84. <https://doi.org/10.1037/rev0000254>
- Phillips I. Scepticism about unconscious perception is the default hypothesis. *J Conscious Stud* 2021b;**28**:186–205.
- Prochow D, Kossack H, Brunheim S et al. Processing of subliminal facial expressions of emotion: a behavioral and fMRI study. *Soc Neurosci* 2013;**8**:448–61. <https://doi.org/10.1080/17470919.2013.812536>
- Rees G, Kreiman G, Koch C. Neural correlates of consciousness in humans. *Nat Rev Neurosci* 2002;**3**:261–70. <https://doi.org/10.1038/nrn783>
- Reingold EM, Merikle PM. Using direct and indirect measures to study perception without awareness. *Percept Psychophys* 1988;**44**:563–75. <https://doi.org/10.3758/BF03207490>
- Reingold EM, Merikle PM. On the inter-relatedness of theory and measurement in the study of unconscious processes. *Mind Language* 1990;**5**:9–28. <https://doi.org/10.1111/j.1468-0017.1990.tb00150.x>
- Rosenthal CR, Andrews SK, Antoniadis CA et al. Learning and recognition of a non-conscious sequence of events in human primary visual cortex. *Curr Biol* 2016;**26**:834–41. <https://doi.org/10.1016/j.cub.2016.01.040>
- Rothkirch M, Hesselmann G. What we talk about when we talk about unconscious processing—a plea for best practices. *Front Psychol* 2017;**8**:835. <https://doi.org/10.3389/fpsyg.2017.00835>
- Sand A, Nilsson ME. Subliminal or not? Comparing null-hypothesis and Bayesian methods for testing subliminal priming. *Conscious Cogn* 2016;**44**:29–40. <https://doi.org/10.1016/j.concog.2016.06.012>
- Schmidt T. Invisible stimuli, implicit thresholds: why invisibility judgments cannot be interpreted in isolation. *Adv Cogn Psychol* 2015;**11**:31–41. <https://doi.org/10.5709/acp-0169-3>
- Schmidt T, Biafora M. A theory of visibility measures in the dissociation paradigm. *Psychon Bull Rev* 2023;**31**:65–88. <https://doi.org/10.3758/s13423-023-02332-z>
- Schmidt T, Vorberg D. Criteria for unconscious cognition: three types of dissociation. *Percept Psychophys* 2006;**68**:489–504. <https://doi.org/10.3758/BF03193692>
- Schnepf IA, Friedrich F, Hepting C et al. Neural mechanisms of response priming do not support veridical unconscious processing. *Conscious Cogn* 2022;**102**:103348. <https://doi.org/10.1016/j.concog.2022.103348>
- Schurger A, Pereira F, Treisman A et al. Reproducibility distinguishes conscious from nonconscious neural representations. *Science* 2010;**327**:97–9. <https://doi.org/10.1126/science.1180029>
- Seth AK. Consciousness: the last 50 years (and the next). *Brain Neurosci Adv* 2018;**2**:2398212818816019. <https://doi.org/10.1177/2398212818816019>
- Seth AK, Dienes Z, Cleeremans A et al. Measuring consciousness: relating behavioural and neurophysiological approaches. *Trends Cogn Sci* 2008;**12**:314–21. <https://doi.org/10.1016/j.tics.2008.04.008>
- Shanks DR. Regressive research: the pitfalls of post hoc data selection in the study of unconscious mental processes. *Psychon Bull Rev* 2017;**24**:752–75. <https://doi.org/10.3758/s13423-016-1170-y>
- Shanks DR, Malejka S, Vadillo MA. The challenge of inferring unconscious mental processes. *Exp Psychol* 2021;**68**:113–29. <https://doi.org/10.1027/1618-3169/a000517>
- Sheikh UA, Carreiras M, Soto D. Decoding the meaning of unconsciously processed words using fMRI-based MVPA. *Neuro Image* 2019;**191**:430–40. <https://doi.org/10.1016/J.NEUROIMAGE.2019.02.010>
- Simons DJ, Hannula DE, Warren DE et al. Behavioral, neuroimaging, and neuropsychological approaches to implicit perception. In: Zelano P, Moscovitch M, Thompson E (eds.), *Cambridge Handbook of Consciousness*, pp. 207–50. New York: Cambridge University Press, 2007 <https://doi.org/10.1017/CBO9780511816789.010>.
- Stein T, Kaiser D, Fahrenfort JJ et al. The human visual system differentially represents subjectively and objectively invisible stimuli. *PLoS Biol* 2021;**19**:e3001241. <https://doi.org/10.1371/journal.pbio.3001241>
- Stein T, Peelen MV. Dissociating conscious and unconscious influences on visual detection effects. *Nat Hum Behav* 2021;**5**:612–24. <https://doi.org/10.1038/s41562-020-01004-5>
- Stein T, van Gaal S, Fahrenfort JJ. How (not) to demonstrate unconscious priming: overcoming issues with post-hoc data selection, low power, and frequentist statistics. *Conscious Cogn* 2024;**119**:103669. <https://doi.org/10.1016/j.concog.2024.103669>
- Sterzer P, Haynes J-D, Rees G. Fine-scale activity patterns in high-level visual areas encode the category of invisible objects. *J Vis* 2008;**8**:10.1–12. <https://doi.org/10.1167/8.15.10>
- Stockart F, Schreiber M, Amerio P et al. Studying unconscious processing: Contention and consensus. *Behavioral and Brain Sciences*, 1–77. <https://doi.org/10.1017/S0140525X25101489>
- Suslow T, Ihme K, Quirin M et al. Implicit affectivity and rapid processing of affective body language: an fMRI study. *Scand J Psychol* 2015;**56**:545–52. <https://doi.org/10.1111/SJOP.12227>
- Tettamanti M, Conca F, Falini A et al. Unaware processing of tools in the neural system for object-directed action representation. *Soc Neurosci* 2017;**37**:10712–24. <https://doi.org/10.1523/JNEUROSCI.1061-17.2017>
- Torralbo A, Kelley TA, Rees G et al. Attention induced neural response trade-off in retinotopic cortex under load. *Sci Rep* 2016;**6**:1–10. <https://doi.org/10.1038/srep33041>
- Ulrich M, Adams SC, Kiefer M. Flexible establishment of functional brain networks supports attentional modulation of unconscious cognition. *Hum Brain Mapp* 2014;**35**:5500–16. <https://doi.org/10.1002/hbm.22566>
- Ulrich M, Kiefer M. The neural signature of subliminal visuomotor priming: brain activity and functional connectivity profiles. *Cereb Cortex* 2016;**26**:2471–82. <https://doi.org/10.1093/cercor/bhv070>

- Vadillo MA, Konstantinidis E, Shanks DR. Underpowered samples, false negatives, and unconscious learning. *Psychon Bull Rev* 2016;**23**:87–102. <https://doi.org/10.3758/s13423-015-0892-6>
- Vadillo MA, Linssen D, Orgaz C et al. Unconscious or underpowered? Probabilistic cuing of visual attention. *J Exp Psychol Gen* 2020;**149**:160–81. <https://doi.org/10.1037/xge0000632>
- Vadillo MA, Malejka S, Lee DYH et al. Raising awareness about measurement error in research on unconscious mental processes. *Psychon Bull Rev* 2022;**29**:21–43. <https://doi.org/10.3758/s13423-021-01923-y>
- Van Gaal S, Ridderinkhof KR, Scholte HS et al. Unconscious activation of the prefrontal no-go network. *J Neurosci* 2010;**30**:4143–50. <https://doi.org/10.1523/JNEUROSCI.2992-09.2010>
- Wicherts JM, Borsboom D, Kats J et al. The poor availability of psychological research data for reanalysis. *Am Psychol* 2006;**61**:726–8. <https://doi.org/10.1037/0003-066X.61.7.726>
- Yang J, Cao Z, Xu X et al. The amygdala is involved in affective priming effect for fearful faces. *Brain Cogn* 2012;**80**:15–22. <https://doi.org/10.1016/j.bandc.2012.04.005>
- Zerweck IA, Kao C-S, Meyen S et al. Number processing outside awareness? Systematically testing sensitivities of direct and indirect measures of consciousness. *Atten Percept Psychophys* 2021;**83**:2510–29. <https://doi.org/10.3758/s13414-021-02312-2>
- Zhan M, Goebel R, de Gelder B et al. Ventral and dorsal pathways relate differently to visual awareness of body postures under continuous flash suppression. *eNeuro* 2018;**5**:ENEURO.0285-17.2017. <https://doi.org/10.1523/ENEURO.0285-17.2017>
- Züst MA, Colella P, Reber TP et al. Hippocampus is place of interaction between unconscious and conscious memories. *PLoS One* 2015;**10**:e0122459. <https://doi.org/10.1371/JOURNAL.PONE.0122459>

Supplementary Material

1 Keywords

1.1 PubMed (first search)

fmri[Title/Abstract] AND (unconscious[Title/Abstract] OR nonconscious[Title/Abstract] OR non-conscious[Title/Abstract] OR invisible[Title/Abstract] OR 'outside awareness'[Title/Abstract] OR unseen[Title/Abstract] OR 'without awareness'[Title/Abstract] OR subliminal[Title/Abstract] OR implicit[Title/Abstract]) AND (mask*[Title/Abstract] OR 'binocular rivalry'[Title/Abstract] OR 'binocular suppression'[Title/Abstract] OR 'interocular suppression'[Title/Abstract] OR 'continuous suppression'[Title/Abstract] OR 'flash suppression'[Title/Abstract])

1.2 Web of Science

TI=(fmri AND (unconscious OR nonconscious OR non-conscious OR invisible OR 'outside awareness' OR unseen OR 'without awareness' OR subliminal OR implicit) AND (mask* OR 'binocular rivalry' OR 'binocular suppression' OR 'interocular suppression' OR 'continuous suppression' OR 'flash suppression')) OR AB=(fmri AND (unconscious OR nonconscious OR non-conscious OR invisible OR 'outside awareness' OR unseen OR 'without awareness' OR subliminal OR implicit) AND (mask* OR 'binocular rivalry' OR 'binocular suppression' OR 'interocular suppression' OR 'continuous suppression' OR 'flash suppression')) OR (fmri AND ('attentional blink'[Title/Abstract] OR 'inattentional blindness'[Title/Abstract] OR 'change blindness'[Title/Abstract])))

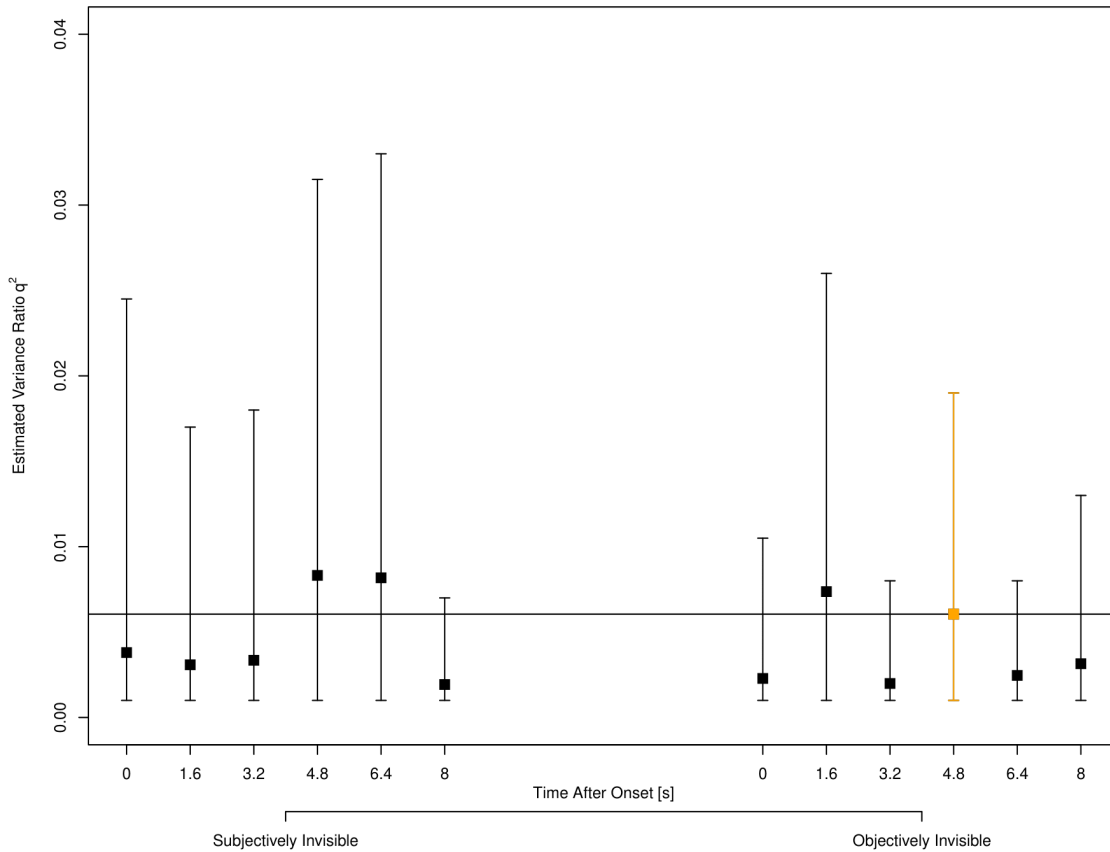
1.3 PubMed (second search)

fmri[Title/Abstract] AND ('attentional blink'[Title/Abstract] OR 'inattentional blindness'[Title/Abstract] OR 'change blindness'[Title/Abstract])

2 Estimation of q^2 from other experimental conditions

Figure S1

Estimates of the Variance Ratio q^2 from Stein et al. (2021)



Note. Stein et al. (2021) computed decoding sensitivities d' at different time points after stimulus onset (x-axis) for the subjectively and objectively invisible conditions. Based on these values, estimates for the variance ratio q^2 were derived together with 95% Highest Density Intervals (shown as error bars). The objectively invisible condition at time point 4.8 (colored in orange) yielded the highest decoding sensitivity d' . Therefore, we took the q^2 estimate from that condition for our reanalyses, $q^2 = 0.006$ (horizontal line). On average, q^2 estimates were lower with $q^2 = 0.004$ and $q^2 = 0.005$ in the subjectively and objective invisible condition, respectively. Thus, the q^2 estimate we used is likely an overestimate and increases the chances of finding an Indirect Task Advantage (ITA) in our reanalyses.

3 Original data and reanalysis results

Axelrod et al. (2015) In this study, researchers used continuous flash suppression (CFS) to mask a succession of words, either forming a meaningful or non-meaningful sentence. A localizer with visible words was used to define regions of interest (ROI; LPSTS: left posterior superior temporal sulcus, LMFG: left middle frontal gyrus). Decoding analyses were run on different numbers of voxels (100, 50 and 150). Four sentences were displayed per block, and participants were asked whether these were meaningful sentences at the end of each block. Participants performed distinct direct tasks for the two stimulus types. Both resulted in the same mean accuracy M . We picked the lower standard error (SE) value to follow our benefit-of-the-doubt approach and favor finding an ITA. Only ‘guessed’ trials were included in the original analysis.

Standard Reasoning: “After each block of either sentences or nonwords participants reported whether they had been aware of even a single word—a procedure which ensured that data analyses were conducted only on blocks judged invisible by participants. To discriminate between neural activity elicited by the 2 conditions we used multivoxel pattern classification analyses (MVPA) focusing on the language network [...] The principal goal of our research was to test whether the frontal lobes were involved in any unconscious processing of language.” (p. 2160 and 2161)

Table S1: Reported results from Axelrod et al. (2015)

	Indirect measure			Direct measure			Sensitivity Comparison	
	N	K	$d' \pm SE$	N	K	$d' \pm SE$	$d'_{\text{diff}} \pm SE$	CI 95%
LPSTS-100	15	77	0.31 ± 0.10	15	77	0.09 ± 0.10	0.23 ± 0.14	[-0.08, 0.54]
LMFG-100	15	77	0.24 ± 0.07	15	77	0.09 ± 0.10	0.15 ± 0.13	[-0.13, 0.43]
LPSTS-50	15	77	0.32 ± 0.10	15	77	0.09 ± 0.10	0.24 ± 0.15	[-0.08, 0.56]
LMFG-50	15	77	0.21 ± 0.07	15	77	0.09 ± 0.10	0.13 ± 0.13	[-0.14, 0.40]
LPSTS-150	15	77	0.34 ± 0.10	15	77	0.09 ± 0.10	0.25 ± 0.14	[-0.05, 0.56]
LMFG-150	15	77	0.32 ± 0.08	15	77	0.09 ± 0.10	0.24 ± 0.13	[-0.05, 0.52]

Note. LPSTS-n: left posterior superior temporal sulcus (decoding using ROI of n voxels); LMFG: left middle frontal gyrus.

Dehaene et al. (2001). The study consists of two experiments. In E1 the ITA referred to the presence vs. absence of the masked word, while in E2 the ITA referred to a congruency effect (repeated vs. different words). In E1, visibility of words was evaluated using different measures (detection/naming, recognition memory, forced-choice test); for this reanalysis we could only include the forced-choice test. In E1, neural activity was measured using both EEG and fMRI; here we focused on fMRI results. Authors reported Z values at local maxima. We converted these values into t values using quantile mapping (see Methods section).

Standard Reasoning: “In the forced-choice test, on each of 37 trials, a short stream comprising a single masked word was presented. Participants were told about the presence of a hidden word and were asked to select it among two choice words presented left and right of fixation. The success rate of 52.9% did not differ from the 50% value expected by chance [...]” (p. 752-753). “Behaviorally, participants again denied seeing the primes and were unable to select them in a two-alternative forced-choice test [...] brain activation was reduced in extrastriate, fusiform and precentral regions similar to those observed in experiment 1. This shows that the repetition suppression phenomenon, which was previously obtained with consciously visible stimuli, can be replicated with unseen masked primes [...]” (p. 755-756).

Table S2: Reported results from Dehaene et al. (2001)

	Indirect measure			Direct measure			Sensitivity Comparison	
	<i>N</i>	<i>K</i>	$d' \pm SE$	<i>N</i>	<i>K</i>	$d' \pm SE$	$d'_{\text{diff}} \pm SE$	CI 95%
LEC E1-LM1	15	150	0.15 ± 0.06	27	37	0.15 ± 0.08	0.00 ± 0.10	[-0.20, 0.20]
LEC E1-LM2	15	150	0.14 ± 0.06	27	37	0.15 ± 0.08	-0.00 ± 0.10	[-0.20, 0.20]
LFG E1-LM1	15	150	0.12 ± 0.05	27	37	0.15 ± 0.08	-0.03 ± 0.10	[-0.22, 0.17]
LFG E1-LM2	15	150	0.12 ± 0.05	27	37	0.15 ± 0.08	-0.03 ± 0.10	[-0.23, 0.17]
LPS E1-LM1	15	150	0.12 ± 0.05	27	37	0.15 ± 0.08	-0.03 ± 0.10	[-0.23, 0.17]
LPS E1-LM2	15	150	0.10 ± 0.05	27	37	0.15 ± 0.08	-0.04 ± 0.10	[-0.24, 0.16]
REC E2-LM1	10	240	0.33 ± 0.10	10	64	0.18 ± 0.10	0.15 ± 0.14	[-0.18, 0.48]
REC E2-LM2	10	240	0.31 ± 0.10	10	64	0.18 ± 0.10	0.13 ± 0.14	[-0.19, 0.45]
LFG E2-CI	10	480	0.17 ± 0.06	10	64	0.18 ± 0.10	-0.01 ± 0.12	[-0.28, 0.25]

Note. LEC: left extrastriate cortex; LFG: left fusiform gyrus; LPS: left precentral sulcus; REC: right extrastriate cortex; E1: experiment 1; E2: experiment 2; LM1: local maximum 1; LM2: local maximum 2; CI: case independent.

Fang et al. (2005). The study reports two experiments. The first experiment does not fit the scope of this reanalysis and was therefore excluded. In E2, tools and faces were rendered invisible using binocular rivalry. In the direct measure, participants were asked to determine in which interval the intact (vs. scrambled) object was presented. In parallel, their BOLD activation in ventral and dorsal ROIs were measured (tools vs. faces; indirect measure). For the indirect task, we estimated the number of trials *K* based on our best understanding of the methods section and following our benefit-of-the-doubt approach. Note that we used the *F* value from the main effect incorporating visible and invisible trials. Although we are only interested in the invisible trials (there was no *F* value for only the invisible condition), the effect is certainly larger in the visible condition and including these trials follows our benefit-of-the-doubt approach. Note also that here the direct measure compares intact vs. scrambled while the indirect measure compares tools vs. faces. This methodological discrepancy between the two measures greatly limits any conclusion that could be made from a confirmed ITA.

Standard Reasoning: “[...] although the subjects could not tell whether the images were faces or tools or even if they were intact, their dorsal cortical neurons still reacted differently to different invisible object images: images of tools induced much stronger BOLD signals in the dorsal ROIs than did images of faces” (p. 1383).

Table S3: Reported results from Fang et al. (2005)

	Indirect measure			Direct measure			Sensitivity Comparison	
	<i>N</i>	<i>K</i>	$d' \pm SE$	<i>N</i>	<i>K</i>	$d' \pm SE$	$d'_{\text{diff}} \pm SE$	CI 95%
Amygdala E2	5	520	0.16 ± 0.10	7	500	-0.02 ± 0.03	0.18 ± 0.11	[-0.08, 0.43]

Note. E2: experiment 2.

Fogelson et al. (2014). In this study, faces and tools were rendered invisible using continuous flash suppression (CFS) or chromatic flicker fusion (CFF). Participants were tested on their ability to discriminate stimuli for both methods (direct measure). MVPA was performed on fMRI data using a linear SVM classifier (indirect measure). We should emphasize that one of the conditions (blinding method = CFF; ROI = fusiform gyrus) almost yielded an ITA (CI 95% = [-0.00044, 0.33251]).

Standard Reasoning: “Behavioral data collected during scanning show that subjects were at chance when guessing stimulus category during both invisible conditions [...] A subset of these regions also showed significant category classification in the absence of stimulus awareness. [...]” (p. 5-6).

Table S4: Reported results from Fogelson et al. (2014)

	Indirect measure			Direct measure			Sensitivity Comparison	
	<i>N</i>	<i>K</i>	$d' \pm SE$	<i>N</i>	<i>K</i>	$d' \pm SE$	$d'_{diff} \pm SE$	CI 95%
CFS-MOG	17	128	0.10 ± 0.05	17	128	-0.02 ± 0.06	0.11 ± 0.08	[-0.05, 0.27]
CFS-MOLS	17	128	0.09 ± 0.05	17	128	-0.02 ± 0.06	0.10 ± 0.08	[-0.06, 0.27]
CFS-LG	17	128	0.12 ± 0.05	17	128	-0.02 ± 0.06	0.14 ± 0.08	[-0.03, 0.30]
CFF-Ig	17	128	0.09 ± 0.05	17	128	-0.03 ± 0.06	0.12 ± 0.08	[-0.04, 0.28]
CFF-FG	17	128	0.14 ± 0.05	17	128	-0.03 ± 0.06	0.17 ± 0.08	[-0.00, 0.33]
CFF-SPS	17	128	0.12 ± 0.05	17	128	-0.03 ± 0.06	0.15 ± 0.08	[-0.02, 0.31]

Note. CFS: continuous flash suppression; CFF: chromatic flicker fusion; MOG: middle occipital gyrus; MOLS: middle occipital and lunate sulci; LG: lingual gyrus; FG: fusiform gyrus; SPS: superior precentral sulcus.

Freeman et al. (2014). Freeman et al. (2012) looked at BOLD signal variations in the amygdala when participants were shown faces with varying degrees of trustworthiness. Real and computer-generated faces with different degrees of facial trustworthiness were presented to participants using backward masking. The study reports two experiments. In E1, the authors used a blocked design to present three levels of facial trustworthiness (low, average, high). In E2, an event-related design allowed the presentation of a continuous range of facial trustworthiness. The two first runs of E2 were subliminal, and the two others were supraliminal. We focused the reanalysis on subliminal runs and values which relied on collapsed conditions (subliminal and supraliminal) were excluded. The authors report the results from one direct measure per experiment (gender discrimination) and from a separate task with different participants (trustworthiness discrimination). We decided to report both. The indirect measure corresponds to difference in BOLD activation in the amygdala between conditions (low trustworthiness vs average). Note that our estimated SE was lower than what authors reported, i.e., it appears that there was more variability in the observed values than we expected. This follows our benefit-of-the-doubt approach, as underestimating the SE will lead to narrower confidence intervals and more confirmed ITAs.

Four analyses out of 16 led to a confirmed ITA, all in experiment 2, namely in the left (d' difference = 0.16 ± 0.07 ; 0.19 ± 0.07 for the quadratic effect) and right amygdala (0.17 ± 0.07 and 0.18 ± 0.07 for the quadratic effect). Importantly, we only obtained a confirmed ITA when using the gender discrimination task as a direct measure, not the trustworthiness discrimination task (separate task with different participants). Note that reanalysis results from quadratic analysis are not as straightforward to interpret, but as they were considered as highly valuable evidence for unconscious processing by the authors, we finally decided to include them in the reanalysis.

Standard Reasoning: “ d' overall was quite low ($M=0.17$, $SE=0.11$), ensuring that the masked stimuli were below subjects’ awareness. [...] Low-trustworthy targets elicited stronger activation than average-trustworthy targets [...] Thus, regions in the bilateral amygdala exhibited especially strong activation for low-trustworthy faces when presented subliminally.” (p. 10576-10577).

Table S5: Reported results from Freeman et al. (2014)

	Indirect measure			Direct measure			Sensitivity Comparison	
	<i>N</i>	<i>K</i>	$d' \pm SE$	<i>N</i>	<i>K</i>	$d' \pm SE$	$d'_{\text{diff}} \pm SE$	CI 95%
Amygdala E1-Low-GenderDT	19	1440	0.04 ± 0.02	19	144	0.17 ± 0.05	-0.13 ± 0.06	[-0.24, -0.01]
Amygdala E1-Low-TrustDT	19	1440	0.04 ± 0.02	16	211	0.03 ± 0.05	0.01 ± 0.05	[-0.10, 0.13]
R Amygdala E1-Low-GenderDT	19	1440	0.06 ± 0.02	19	144	0.17 ± 0.05	-0.11 ± 0.06	[-0.23, 0.00]
R Amygdala E1-Low-TrustDT	19	1440	0.06 ± 0.02	16	211	0.03 ± 0.05	0.03 ± 0.05	[-0.09, 0.14]
R Amygdala E1-High-GenderDT	19	1440	0.04 ± 0.02	19	144	0.17 ± 0.05	-0.13 ± 0.06	[-0.24, -0.01]
R Amygdala E1-High-TrustDT	19	1440	0.04 ± 0.02	16	211	0.03 ± 0.05	0.01 ± 0.05	[-0.10, 0.13]
R Amygdala E1-quadratic-GenderDT	19	1440	0.06 ± 0.02	19	144	0.17 ± 0.05	-0.11 ± 0.06	[-0.23, 0.01]
R Amygdala E1-quadratic-TrustDT	19	1440	0.06 ± 0.02	16	211	0.03 ± 0.05	0.03 ± 0.05	[-0.08, 0.15]
L Amygdala E2-GenderDT	15	320	0.07 ± 0.04	15	160	-0.09 ± 0.05	0.16 ± 0.07	[0.02, 0.31]
L Amygdala E2-TrustDT	15	320	0.07 ± 0.04	16	211	0.03 ± 0.05	0.04 ± 0.06	[-0.09, 0.17]
R Amygdala E2-GenderDT	15	320	0.08 ± 0.04	15	160	-0.09 ± 0.05	0.17 ± 0.07	[0.02, 0.31]
R Amygdala E2-TrustDT	15	320	0.08 ± 0.04	16	211	0.03 ± 0.05	0.05 ± 0.06	[-0.08, 0.18]
L Amygdala-quadratic-E2-GenderDT	15	320	0.10 ± 0.04	15	160	-0.09 ± 0.05	0.19 ± 0.07	[0.05, 0.34]
L Amygdala-quadratic-E2-TrustDT	15	320	0.10 ± 0.04	16	211	0.03 ± 0.05	0.07 ± 0.06	[-0.06, 0.21]
R Amygdala-quadratic-E2-GenderDT	15	320	0.09 ± 0.04	15	160	-0.09 ± 0.05	0.18 ± 0.07	[0.04, 0.33]
R Amygdala-quadratic-E2-TrustDT	15	320	0.09 ± 0.04	16	211	0.03 ± 0.05	0.06 ± 0.06	[-0.07, 0.20]

Note. GenderDT: gender discrimination (direct task); TrustDT: trustworthiness discrimination (direct task); Low: low vs. average trustworthiness; High: high vs. average trustworthiness; E1: experiment 1; E2: experiment 2; L: left; R: right.

Haynes & Rees (2005). In this study, two experiments are reported. In E1, stimuli are visible and therefore the experiment was excluded from the reanalysis. In E2, Gabor patches are rendered invisible using a masking technique. Participants are asked to report the orientation at the end of each block by pressing on a button (direct measure). A classifier is trained on fMRI data to decode patches' orientation (indirect measure). Note that we estimated decoding values (% accuracy and SE) from Figure 3 because they were not reported in the main text.

Haynes and Rees (2005) reported an accuracy of 50.3% (SE = 0.4%) when four participants attempted to discriminate the orientation of a grating (1170 trials), corresponding to a d' of 0.02 (SE = 0.01). From their Figure 3, we estimated the maximum decoding performance of gratings' orientation in V1 at approximately 58.35% (SE = 2.25%). We considered the data point with the highest decoding accuracy available (100 voxels decoding) in order to follow our benefit-of-the-doubt approach. This decoding performance can be estimated as a d' value of 0.42 (SE = 0.11), resulting in a sensitivity difference of 0.41 (SE = 0.11) and a 95% confidence interval that excludes zero (CI = [0.04-0.77]).

Standard Reasoning: “[...] participants were completely unaware of the orientation of the masked gratings and were at chance performance ($50.3 \pm 0.4\%$, s.e.m.) in discriminating their orientation. [...] we found that single volumes could be classified with an accuracy that was significantly above chance for each of the four participants [...] Thus, even when participants' conscious reports indicated that they themselves could not distinguish the orientation of a masked grating, their brain state contained information that could permit such discrimination.” (p. 688).

Table S6: Reported results from Hayne & Rees (2005)

	Indirect measure			Direct measure			Sensitivity Comparison	
	<i>N</i>	<i>K</i>	$d' \pm SE$	<i>N</i>	<i>K</i>	$d' \pm SE$	$d'_{diff} \pm SE$	CI 95%
V1	4	900	0.42 ± 0.11	4	1170	0.02 ± 0.02	0.41 ± 0.11	[0.04, 0.77]

Note. V1: primary visual area.

Kouider et al. (2007). This study reports a subliminal and a supraliminal condition. In the subliminal condition, participants were presented with a masked prime, either unrelated to the target, orthographically similar to the target, or both orthographically similar and homophonic with the target. While participants were performing the task, fMRI data was recorded and the authors tested for a suppression effect, i.e., decrease of BOLD activation in some areas when prime and target were congruent (indirect measure). Immediately after the scanning session, participants' ability to discriminate the prime was tested (direct measure). We did our best to estimate *K* for the indirect measure based on our understanding of the methods section. Note that we included *Z* values that were not significant (left and right FEF) or uncorrected (VWFA) because these results nevertheless served the narrative of the reanalyzed study.

Standard Reasoning: "Data from the forced-choice prime identification task were used to evaluate prime visibility. Measures of *d'* values for each subject confirmed that they were unable to consciously perceive the primes in the subliminal condition [...] the VWFA previously reported in several studies of subliminal priming during reading (Dehaene et al. 2001, 2004; Devlin et al. 2004), also showed a small repetition suppression effect" (p. 2022-2023).

Table S7: Reported results from Kouider et al. (2007)

	Indirect measure			Direct measure			Sensitivity Comparison	
	<i>N</i>	<i>K</i>	$d' \pm SE$	<i>N</i>	<i>K</i>	$d' \pm SE$	$d'_{diff} \pm SE$	CI 95%
LFEF	15	448	0.21 ± 0.05	15	180	0.19 ± 0.05	0.02 ± 0.08	[-0.14, 0.18]
RFEF	15	448	0.17 ± 0.05	15	180	0.19 ± 0.05	-0.02 ± 0.07	[-0.17, 0.14]
VWFA	15	448	0.14 ± 0.04	15	180	0.19 ± 0.05	-0.05 ± 0.07	[-0.20, 0.09]

Note. LFEF: left frontal eye field; RFEF: right frontal eye field; VWFA: visual word form area.

Kouider et al. (2009). In this study, participants were presented with familiar and unfamiliar faces and were asked to perform a fame-judgment task. Target faces were preceded by a masked face, either the same picture (same-view), the same face but viewed from a different angle (cross-view), or a different face. BOLD variations were measured and repetition suppression effects were obtained (indirect measure 1). In another phase of the experiment, an ANOVA was performed on four selected face-responsive ROIs (indirect measure 2). For this second indirect measure, we interpreted *N* to be 13 instead of 16 to be in agreement with our benefit-of-the-doubt approach. In a separate task, participants were asked to identify if the prime face was familiar or unfamiliar (direct measure). Note that data from another direct measure was available but we decided to reanalyze the fame-judgment task, as it was performed with the same participants and as fame was the feature of interest of the main experiment. Moreover, the prime fame-judgment task resulted in a lower *d'* than the other task (*d'* = 0.34). Selecting the fame-judgment task as a direct measure therefore follows our benefit-of-the-doubt approach.

Standard Reasoning: "The forced-choice fame-judgment task on the primes confirmed that our

masking method rendered the primes largely invisible, as performance was close to chance [...]” (p. 16). [...] we found evidence of repetition-related hemodynamic response decreases (i.e., repetition suppression) in several regions of the occipitotemporal cortex. [...] These data provide evidence that face processing can occur in face-processing regions within the ventral visual stream in the absence of perceptual awareness.” (p. 18).

Table S8: Reported results from Kouider et al. (2009)

	Indirect measure			Direct measure			Sensitivity Comparison	
	<i>N</i>	<i>K</i>	$d' \pm SE$	<i>N</i>	<i>K</i>	$d' \pm SE$	$d'_{diff} \pm SE$	CI 95%
RSTG&S	16	480	0.18 ± 0.05	16	64	0.16 ± 0.08	0.02 ± 0.09	[-0.18, 0.22]
RSTG&S-2	16	480	0.15 ± 0.04	16	64	0.16 ± 0.08	-0.01 ± 0.09	[-0.21, 0.18]
LLOC&PMTG	16	480	0.17 ± 0.05	16	64	0.16 ± 0.08	0.01 ± 0.09	[-0.18, 0.21]
LLOC&PMTG-2	16	480	0.16 ± 0.04	16	64	0.16 ± 0.08	0.00 ± 0.09	[-0.19, 0.20]
LLOC&PMTG-3	16	480	0.16 ± 0.04	16	64	0.16 ± 0.08	-0.00 ± 0.09	[-0.20, 0.19]
LMFG	16	480	0.17 ± 0.04	16	64	0.16 ± 0.08	0.01 ± 0.09	[-0.19, 0.20]
RSTG	16	480	0.15 ± 0.04	16	64	0.16 ± 0.08	-0.01 ± 0.09	[-0.20, 0.18]
RPMTG	16	480	0.14 ± 0.04	16	64	0.16 ± 0.08	-0.02 ± 0.09	[-0.21, 0.17]
RPMTG-2	16	480	0.13 ± 0.04	16	64	0.16 ± 0.08	-0.03 ± 0.09	[-0.22, 0.16]
4FROIS Global priming	13	480	0.09 ± 0.04	16	64	0.16 ± 0.08	-0.07 ± 0.09	[-0.26, 0.12]
4FROIS Same	13	320	0.09 ± 0.04	16	64	0.16 ± 0.08	-0.07 ± 0.09	[-0.26, 0.13]
4FROIS Cross	13	320	0.08 ± 0.04	16	64	0.16 ± 0.08	-0.08 ± 0.09	[-0.27, 0.12]

Note. RSTG&S: right superior temporal gyrus and sulcus; LLOC&PMTG: left lateral occipital complex and posterior middle temporal gyrus; LMFG: left mid-fusiform gyrus; RSTG: right superior temporal gyrus; RPMTG: right posterior middle temporal gyrus; 4FROIS: 4 functional regions of interest they selected for the omnibus ANOVA; Same: same-view faces; Cross: cross-view faces.

Kouider et al. (2016). In this study, twenty participants were alternately asked to detect either faces, flowers or watches, meaning that these stimuli could be either task-relevant or task-irrelevant. Stimuli were masked and presentation duration was either 200ms (visible) or 33ms (invisible). Ten different participants performed a discrimination task outside of the scanner to assess stimulus visibility (direct measure). Note that stimulus awareness was also tested during the main experiment, but the separate task was considered a more valid way to assess awareness by the authors. An ANOVA (stimulus category x task relevance) and t-test analyses were performed on the BOLD activation differences in FFA between the different conditions (indirect measure).

Standard Reasoning: “[...] the stimuli in the invisible conditions could not be discriminated not only when responding solely to the target category ($d' = 0.061$, $t < 1$), but also when performing a forced-choice on each trial ($d' = -0.013$, $t < 1$).” (p. 4). “Further analysis revealed that FFA responses to faces were enhanced during face relevance blocks relative to alternative relevance blocks [...] the FFA responded more to invisible faces than to the other invisible objects in the face detection blocks [...] the activity for invisible faces was amplified relative to either invisible alternative objects ($t(19) = 2.246$, $P = 0.037$) or to invisible control objects ($t(19) = 2.237$, $P = 0.038$).” (p. 5) “Our study goes further by revealing that selective attention, in the absence of awareness, involves not only response enhancement for task-relevant information, but also the active filtering of distracting information in visual cortex [...] The current study provides further evidence for the flexibility of non-conscious perceptual processes, by showing that sensory regions can in turn be modulated by task relevance.” (p. 7).

Table S9: Reported results from Kouider et al. (2016)

	Indirect measure			Direct measure			Sensitivity Comparison	
	<i>N</i>	<i>K</i>	$d' \pm SE$	<i>N</i>	<i>K</i>	$d' \pm SE$	$d'_{diff} \pm SE$	CI 95%
FFA Faces (FD vs. AD)	20	360	0.08 ± 0.03	10	180	-0.01 ± 0.06	0.09 ± 0.07	[-0.07, 0.25]
FFA Faces vs. Objects (FD)	20	180	0.10 ± 0.04	10	180	-0.01 ± 0.06	0.11 ± 0.08	[-0.06, 0.28]
FFA Objects vs Faces (AD)	20	180	0.11 ± 0.04	10	180	-0.01 ± 0.06	0.13 ± 0.08	[-0.05, 0.30]
FFA Faces vs. Alternative (FD)	20	120	0.10 ± 0.05	10	180	-0.01 ± 0.06	0.11 ± 0.08	[-0.07, 0.29]
FFA Faces vs. Control (FD)	20	120	0.10 ± 0.05	10	180	-0.01 ± 0.06	0.11 ± 0.08	[-0.07, 0.29]
FFA Faces vs. Control (AD)	20	120	0.13 ± 0.05	10	180	-0.01 ± 0.06	0.14 ± 0.08	[-0.04, 0.33]
FFA Faces vs. Mask-only	20	120	0.13 ± 0.05	10	180	-0.01 ± 0.06	0.14 ± 0.08	[-0.04, 0.33]

Note. FFA: fusiform face area; FD: face detection task; AD: alternative object detection task.

Moutoussis & Zeki (2002). In this study, the authors use a method called dichoptic fusion to render stimuli (houses and faces) invisible. Visibility was assessed before scanning using a discrimination task (direct measure). Crucially, here we reanalysed the *t* threshold for single-voxel significance as defined by the authors ($t = 3.12$; indirect measure). It is important to stress that some results from which the values were not reported in the article might have been non-negligibly above this threshold, potentially yielding an ITA. However, results based on this significance threshold reported by the authors are not sufficient to support their claims.

Standard Reasoning: “[...] in the opposite conditions, the 2AFC gave an average performance of 52.7% (SD = 4.2); 4/7 subjects scored above 50%, the highest score being 59.4%.” “The resultant parameter estimates for each regressor at each voxel were compared by using *t* tests to determine whether significant activation resulting from a comparison of conditions had occurred. [...] Although the extent of activation was not as widespread as with perceived stimuli, it still is surprising that many “higher,” binocularly driven areas of the brain are activated by these invisible stimuli when compared with the (perceptually equivalent) uniform controls.” (p. 9528-9529).

Table S10: Reported results from Moutoussis & Zeki (2002)

	Indirect measure			Direct measure			Sensitivity Comparison	
	<i>N</i>	<i>K</i>	$d' \pm SE$	<i>N</i>	<i>K</i>	$d' \pm SE$	$d'_{diff} \pm SE$	CI 95%
Significance threshold	7	122	0.20 ± 0.11	7	32	0.14 ± 0.08	0.07 ± 0.13	[-0.26, 0.39]

Schurger et al. (2010). In this study, participants were presented with face and house stimuli which were masked using dichoptic-color masking. Participants’ awareness was tested on each trial using a discrimination task (direct measure). A Gaussian naïve Bayes classifier was trained on fMRI data (temporal lobes; indirect measure). Note that the authors also used wagering (confidence in one’s answer) as a collateral index of one’s awareness.

Standard Reasoning: “For visible stimuli, performance was at or near 100% correct for all 12 subjects, and all wagers were high. For invisible stimuli, task performance was only marginally different from chance ($54 \pm 2.5[\text{SEM}]\%$ correct; $P < 0.06$, one-tailed *t* test) [...] We used multivariate pattern analysis to ascertain how the encoding of perceptual information differs depending on whether or not that information is present in subjective experience (17). Thus, in our analyses we focused specifically on the patterns of activation corresponding to the perceptual information of which the subject was or was not aware: the category of the object.” (p. 97).

Table S11: Reported results from Schurger et al. (2010)

	Indirect measure			Direct measure			Sensitivity Comparison	
	<i>N</i>	<i>K</i>	$d' \pm SE$	<i>N</i>	<i>K</i>	$d' \pm SE$	$d'_{diff} \pm SE$	CI 95%
Temporal lobes	12	288	0.40 ± 0.15	12	288	0.20 ± 0.12	0.20 ± 0.20	[-0.24, 0.64]

Stein et al. (2021). Participants were asked to discriminate masked house and face stimuli after each trial, while lying in the scanner (direct measure). The authors examined BOLD activity patterns in four regions of interest (V1, LOC, OFA/OPA, FFA/PPA; indirect measure).

Standard Reasoning: “Importantly, in obj-inv trials, discrimination performance ($M = 0.02$, $SD = 0.25$) did not differ significantly from chance (Fig 1E), with moderate evidence for the null hypothesis of chance level discrimination [...]” (p. 5). “Activity patterns in LOC discriminated between faces and houses with above-chance accuracy in all visibility conditions [...] and also in obj-inv ($t(42) = 3.28$, $p = 0.001$, $dz = 0.50$, $BF+0 = 31.19$). Thus, LOC contained category information for both subjectively and objectively invisible stimuli.” (p. 6).

Table S12: Reported results from Stein et al. (2021)

	Indirect measure			Direct measure			Sensitivity Comparison	
	<i>N</i>	<i>K</i>	$d' \pm SE$	<i>N</i>	<i>K</i>	$d' \pm SE$	$d'_{diff} \pm SE$	CI 95%
V1	43	200	0.06 ± 0.03	43	200	0.02 ± 0.03	0.04 ± 0.04	[-0.04, 0.12]
LOC	43	200	0.08 ± 0.03	43	200	0.02 ± 0.03	0.06 ± 0.04	[-0.02, 0.14]
OFA/OPA	43	200	0.07 ± 0.03	43	200	0.02 ± 0.03	0.05 ± 0.04	[-0.03, 0.12]

Note. V1: primary visual area; LOC: lateral occipital complex; OFA/OPA: occipital face area/occipital place area.

Sterzer et al. (2008). In this study, MVPA was used to decode house and face stimuli rendered invisible using continuous flash suppression (CFS). Participants’ awareness was tested using both a discrimination task (direct measure) and a subjective assessment. Classification was performed on the fMRI data from two main ROIs (FFA and PPA) using linear support vector machines (indirect measure). 4.1% of the trials were excluded because they were subjectively visible. One run was excluded because the participant reported being unable to binocularly fuse the stimuli due to sleepiness.

Standard Reasoning: “Discrimination of invisible faces and houses was tested in a 2-alternative forced-choice task performed directly after each block; performance was at chance level in all participants (average $d' = -0.05 \pm 0.14$ SEM, $p = 0.71$, one-sample t-test). [...] Strikingly, prediction accuracy for invisible stimuli (where univariate analyses had failed to show differences between activity evoked by face and house stimuli) was also significantly above chance level (FFA: $58.8\% \pm 2.3$ SEM, $t(4) = 3.8$, $p = 0.019$; PPA: $62.5\% \pm 3.3$ SEM, $t(4) = 3.8$, $p = 0.019$; FFA + PPA: $63.5\% \pm 3.7$ SEM, $t(4) = 3.6$, $p = 0.022$).” (p. 6-7). “Our data demonstrate that activity patterns in the FFA and the PPA differentiate two categories of object stimuli (faces and houses) even when the stimuli are rendered invisible by interocular suppression.” (p. 9).

Table S13: Reported results from Sterzer et al. (2008)

	Indirect measure			Direct measure			Sensitivity Comparison	
	<i>N</i>	<i>K</i>	$d' \pm SE$	<i>N</i>	<i>K</i>	$d' \pm SE$	$d'_{diff} \pm SE$	CI 95%
FFA	5	77	0.44 ± 0.11	5	77	-0.05 ± 0.13	0.49 ± 0.18	[0.01, 0.98]
PPA	5	77	0.64 ± 0.16	5	77	-0.05 ± 0.13	0.69 ± 0.21	[0.10, 1.27]
FFA&PPA	5	77	0.69 ± 0.18	5	77	-0.05 ± 0.13	0.74 ± 0.23	[0.11, 1.37]

Note. FFA: fusiform face area; PPA: parahippocampal place area.

Ulrich & Kiefer (2016). Participants were presented with white geometrical shapes (circle, diamond, square, ellipsoid). Two of these shapes were mapped to the left index finger (circle and diamond) and the two others to the right index finger (square and ellipsoid), or vice versa. These target shapes were preceded by either congruent or incongruent primes that were rendered invisible using a forward and a backward line pattern mask. Variations in brain activity were measured using fMRI (indirect task). Both congruent (5.7%) and incongruent trials (6%) were excluded because of being either incorrect or missing. After scanning, participants' awareness of the prime was assessed inside the scanner using a discrimination task (direct task). Note that part of the claim of this study relies on functional connectivity analyses that could not be included in the reanalysis.

Standard Reasoning: "The d' measure of prime visibility (Green and Swets 1966) was 0.03 on average (SD = 0.39) and did not significantly deviate from zero ($t(30) = 0.43$, $P = 0.670$)." (p.2475). "[...] analyses revealed lower brain activity for congruent (CON) than incongruent (INC) trials in ventrolateral and dorsomedial frontal and inferior parietal brain regions as well as in the basal ganglia. The present study thereby confirms and extends earlier findings by demonstrating an involvement of the visuomotor network in subliminal visuomotor processing [...]" (p.2476).

Table S14: Reported results from Ulrich & Kiefer (2016)

	Indirect measure			Direct measure			Sensitivity Comparison	
	<i>N</i>	<i>K</i>	$d' \pm SE$	<i>N</i>	<i>K</i>	$d' \pm SE$	$d'_{diff} \pm SE$	CI 95%
RIPL	31	120	0.19 ± 0.04	31	48	0.03 ± 0.07	0.16 ± 0.08	[-0.00, 0.33]
RMSFG	31	120	0.19 ± 0.04	31	48	0.03 ± 0.07	0.16 ± 0.08	[-0.00, 0.32]
RIFG	31	120	0.18 ± 0.04	31	48	0.03 ± 0.07	0.15 ± 0.08	[-0.01, 0.31]
LIPL	31	120	0.18 ± 0.04	31	48	0.03 ± 0.07	0.15 ± 0.08	[-0.01, 0.31]
RCN	31	120	0.17 ± 0.04	31	48	0.03 ± 0.07	0.14 ± 0.08	[-0.02, 0.30]
LIFG	31	120	0.17 ± 0.04	31	48	0.03 ± 0.07	0.14 ± 0.08	[-0.02, 0.30]

Note. RIPL: right inferior parietal lobule; RMSFG: right medial superior frontal gyrus; RIFG: right inferior frontal gyrus; LIPL: left inferior parietal lobule; RCN: right caudate nucleus; LIFG: left inferior frontal gyrus.

van Gaal et al. (2010). Participants performed a go/no-go task with no-go signals rendered invisible using metacontrast masking. Neural activity was measured using fMRI while participants were performing the go/no-go task (indirect measure). To assess their awareness of the no-go signal, participants were asked to perform a discrimination task while still lying in the scanner and a d' -prime value was computed (direct measure). Four participants who performed better than chance at the discrimination task were excluded from further analysis (regression to the mean).

Standard Reasoning: "The combination of these factors effectively rendered the participants incapable of perceiving the square/diamond, as evidenced by chance-level performance on a two-choice discrimination task administered after the experiment [...]" (p. 4144). "To examine the acti-

vation related to the unconscious initiation of inhibitory control, we contrasted responded, strongly masked no-go trials with responded, strongly masked go trials.”(p. 4146). “In a go/no-go paradigm, we masked no-go signals to the point that they could no longer be detected to investigate the depth of processing of strongly masked (unconscious) no-go signals in the human brain. Strongly masked no-go signals were observed to activate brain regions central to networks that have been associated with conscious response inhibition, namely the IFC and the pre-SMA.” (p. 4147).

Table S15: Reported results from van Gaal et al. (2010)

	Indirect measure			Direct measure			Sensitivity Comparison	
	<i>N</i>	<i>K</i>	$d' \pm SE$	<i>N</i>	<i>K</i>	$d' \pm SE$	$d'_{diff} \pm SE$	CI 95%
RIFC	20	240	0.17 ± 0.05	20	48	0.12 ± 0.08	0.05 ± 0.09	[-0.14, 0.25]
LIFC	20	240	0.15 ± 0.04	20	48	0.12 ± 0.08	0.03 ± 0.09	[-0.16, 0.22]
pre-SMA	20	240	0.14 ± 0.04	20	48	0.12 ± 0.08	0.02 ± 0.09	[-0.17, 0.21]

Note. RIFC: right inferior frontal cortex; LIFC: left inferior frontal cortex; Pre-SMA: pre-supplementary motor area.

Yang et al. (2012). In this study, masked fearful and neutral face stimuli were presented to participants during an encoding phase. During the retrieving phase, half the faces had the same valence as during the encoding and half not. After the main experiment, participants were asked to perform a discrimination task during a post-awareness assessment, outside of the scanner (direct measure). BOLD variations in the amygdala and other brain regions were measured during the encoding and retrieving phases (indirect measure). The specificity of this study compared to the others included in the reanalysis is that two groups of participants were constituted based on awareness scores (regression to the mean). We focused the reanalysis on data collected from the unaware group.

Standard Reasoning: “[...] the unaware participants (N=13) had a chance level in detecting the faces ($p > 0.1$). [...] for unaware participants, fearful faces (vs. neutral) produced stronger activation in the right amygdala (20, 5, 8, $t(12) = 6.39$) and the right pulvinar (6, 11, 6, $t(12) = 4.96$) [...] the amygdala activation was different for unaware and aware participants during encoding, but similar for the two groups during retrieval.” (p. 5-6).

Table S16: Reported results from Yang et al. (2012)

	Indirect measure			Direct measure			Sensitivity Comparison	
	<i>N</i>	<i>K</i>	$d' \pm SE$	<i>N</i>	<i>K</i>	$d' \pm SE$	$d'_{diff} \pm SE$	CI 95%
R Amygdala	13	240	0.25 ± 0.07	13	40	0.09 ± 0.11	0.16 ± 0.13	[-0.13, 0.45]
R Pulvinar	13	240	0.19 ± 0.06	13	40	0.09 ± 0.11	0.10 ± 0.13	[-0.17, 0.38]

Note. R Amygdala: right amygdala; R Pulvinar: right pulvinar.