AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Experimental Psychology: General

ISSN: 0096-3445

https://doi.org/10.1037/xge0000622

No Evidence for Unconscious Initiation and Following of Arithmetic Rules: A Replication Study

Amir Tal[1, 2] and Liad Mudrik[1, 2, 3]
[1] School of Psychological Sciences, Tel-Aviv University
[2] Sagol School for Neuroscience, Tel-Aviv University
[3] Brain, Mind, and Consciousness Program, Canadian Institute for Advanced Research, Toronto, Canada

The field of consciousness studies has yielded various—sometimes contradicting—accounts regarding the function of consciousness, ranging from denying it has such function to claiming that any high-level cognitive function requires consciousness. Empirical findings supporting both accounts were reported, yet some of them have been recently revisited based on failures to replicate. Here, we aimed at replicating a remarkable finding reported by Ric and Muller (2012); participants were able to follow an unseen instruction, integrate it with a subsequently presented pair of unseen digits, and accordingly either add the digits (resulting in a priming effect), or simply represent them. This finding thus demonstrates unconscious task-switching, temporal integration (involving mental chaining), and arithmetic operation. Finding such high-level processes in the absence of awareness is of pivotal importance to our understanding of consciousness, as it challenges prominent theories in the field (e.g., the global neuronal workspace). Accordingly, in light of the self-correction wave in psychological science in general and in the field of consciousness studies in particular, we report here a preregistered replication aimed at testing the reproducibility of this finding, while also better controlling for subjects' awareness of both the instruction and the digits. Across two highly powered experiments, our results failed to replicate the original effect. We, therefore, conclude that the current evidence does not support the claim that arithmetic operations (specifically, addition) can be flexibly initiated without awareness, in line with the current arguments for a more limited scope of unconscious processing.

Public Significance Statement

A major open question in psychology is why we developed consciousness. Why did evolution grant us the ability to experience some processes, and is there any functional benefit to having such conscious experiences? One way to study the latter question focuses on characterizing what cannot be done without consciousness. Here, we re-visit a finding from 2012 showing that people can do math unconsciously in a flexible manner (i.e., following invisible instructions). In two highly powered experiments, we failed to find the same effect. We conclude that consciousness may indeed be necessary for tasks that require integrating different kinds of information in a complex and flexible manner, in line with several theories of consciousness.

Keywords: replication, unconscious processing, arithmetic, visual masking, integration

Supplemental materials: https://doi.org/10.1037/xge0000622.supp

The scope of unconscious processing has been a matter of ongoing debate (e.g., Hassin, 2013; Kouider & Dehaene, 2007; Moors et al., 2017; Mudrik & Deouell, 2022; Shanks, 2017): What types of processes can take place in the absence of awareness, and how deep do these processes run? The possible gap between our subjective experience and the mechanisms of information processing that shape our behavior has fascinated generations of scientists, starting from the early suggestions of subliminal processes by Johann Herbart and Hermann Von Helmholtz (for review, see Kouider & Dehaene, 2007) or the seminal work of Peirce and Jastrow (1884), who were

Laura E. Thomas served as action editor.

Liad Mudrik https://orcid.org/0000-0003-3564-6445

Liad Mudrik is a Canadian Institute for Advanced Research tanenbaum fellow in the Brain, Mind, and Consciousness program.

Amir Tal played a lead role in data curation, formal analysis, investigation, project administration, software, validation, visualization, and writing–original draft, a supporting role in conceptualization, and an equal role in methodology and writing–review and editing. Liad Mudrik played a lead role in conceptualization, funding acquisition, resources, and supervision, a supporting role in writing–original draft, and an equal role in methodology and writing–review and editing.

Correspondence concerning this article should be addressed to Liad Mudrik, School of Psychological Sciences, Tel-Aviv University, Ramat Aviv, P.O. Box 39040, Tel Aviv 69978, Israel. Email: mudrikli@tauex.tau.ac.il

the first to dissociate between participants' subjective report and their objective performance. Ever since, the study of unconscious processing seems to have been oscillating with respect to high-level effects in the absence of awareness; findings reporting such effects are followed by criticism, typically leading to methodological improvements in the way unconscious processing is measured and isolated from conscious processing (e.g., the development and improvement of the contrastive method, Merikle & Reingold, 1998, following criticism by Eriksen, 1960, and others), and the acceptance of some—but not all—effects as valid. Then, new evidence for even higher-level processing emerge, and the cycle repeats (see again Kouider & Dehaene, 2007).

Currently, the field seems to be in the midst of such a wave of criticism and skepticism about high-level processing without awareness. The underlying assumptions and definitions used in the study of unconscious processes are being reexamined (Peters et al., 2017; see also Moors & Hesselmann, 2018; Stein et al., 2024, for more specific methodological points). Alongside articles that still demonstrate high-level or integrative effects (e.g., D. Biderman et al., 2020; Chong et al., 2014; Fahrenfort et al., 2017; van Gaal et al., 2014), other articles report null findings (Faivre et al., 2014; Faivre & Koch, 2014; Kang et al., 2011; Tal, Sar-Shalom, et al., 2024; Yang et al., 2017); note, however, that absence of evidence is not evidence for absence, though some of these articles support their claim with Bayesian analysis (e.g., Heyman & Moors, 2014), which mitigates this concern, to some extent. Moreover, several key findings that originally reported high-level integration of information in the absence of awareness have been revisited; for example, findings of object-scene integration (Mudrik et al., 2011; Mudrik & Koch, 2013) were not replicated in newer, higher-powered studies (N. Biderman & Mudrik, 2018; Moors et al., 2016). An influential article which found that both sentence reading and performing arithmetic operations can be done unconsciously (Sklar et al., 2012; see also Karpinski et al., 2019) was also challenged on several grounds (Moors & Hesselmann, 2018; Rabagliati et al., 2018; Shanks, 2017; see also Moors & Hesselmann, 2019, in reply to Karpinski et al., 2019).

Yet, the ability to perform arithmetic operations was also demonstrated in other works. Alongside studies that mostly probed the unconscious retrieval of simple arithmetic facts (García-Orza et al., 2009), one study reported performing addition on two subliminally presented numbers, upon instruction (Ric & Muller, 2012). Critically, the results of this article go well beyond demonstrating unconscious application of arithmetic rules, since participants were also reported to flexibly *initiate* these rules. And so, priming from the stimulus "2 ǁ 3" to the target "5" was only found when participants were presented with the instruction "add" and not when they were asked to "represent" or "relativize". Such priming of arithmetic operations by a preceding instruction has been demonstrated in previous works (Roussel et al., 2002; Sohn & Carlson, 1998); however, to the best of our knowledge, always under conscious perception of the stimuli. Here, quite strikingly, this differential pattern was found *even when both the number and the instruction were subliminally presented* (Experiments 2 and 3; note, however, that proper assessment of prime visibility, achieved by taking objective measures of awareness on the same participants who completed the main task, was only performed in Experiment 3. Thus, we focus here on Experiment 3 as it is the only one providing compelling evidence for unconscious arithmetic and integration). This is especially surprising, since to obtain such a result, several conditions should be met: participants need to: (a) unconsciously read and understand the instructions; (b) show flexibility by following them, hereby changing processing strategies between trials; (c) maintain the instruction in working memory throughout the trial; (d) integrate between the instruction and the prime numbers; and (e) perform the actual addition operation on those numbers. The finding that these processes could all take place in the absence of awareness is remarkable, especially because it stands in sharp contradiction to some prominent theories of consciousness.

For example, according to the global neuronal workspace (GNW) theory (Dehaene & Changeux, 2011; Dehaene & Naccache, 2001), consciousness is mediated by a subset of specialized cortical pyramidal cells with long-range axons and numerous reciprocal connections. These neurons are held to form a neuronal workspace, integrating processing modules that are otherwise encapsulated. The theory clearly states that unconsciously processed stimuli would not exert any top-down control on behavior (Dehaene & Naccache, 2001), of the type observed in the study by Ric and Muller (2012). Interestingly, while other studies also demonstrated task-switching effects driven by subliminal cues (for reviews, see Ansorge et al., 2014; van Gaal & Lamme, 2012), these were typically limited to symbolic stimuli, and either involved a previous conscious exposure to the cues (Reuss et al., 2012), or a visible instruction cue that was preceded by a congruent or an incongruent prime (De Pisapia et al., 2012; H. C. Lau & Passingham, 2007; Mattler, 2003, 2007; Zhou & Davis, 2012). Thus, to our knowledge the study by Ric and Muller (2012) constitutes the only demonstration of actual verbal instructions that are subliminally presented to the subjects, with no prior conscious exposure, and are potent enough to affect their behavior.

The GNW theory clearly differentiates between encapsulated and modular unconscious processes and global, integrative conscious ones. This seems to imply that integration between the verbal instruction and the pair of numbers should not be performed in the absence of consciousness. Indeed, and in contrast to the findings of Ric and Muller (2012), Sackur and Dehaene (2009) found that an elementary arithmetic operation can be performed without awareness, but it cannot be chained to another mental operation (notably, however, there participants were asked to add [or subtract] 2 to a digit, and then compare the result with 5, while here mental chaining involves settling on the correct task and then adding or representing the numbers). In a more recent article, Dehaene et al. (2017) further claimed that sequential performance of several tasks requires conscious perception. Importantly, the claim that integration and consciousness are tightly related and should co-occur is not unique to GNW, but seems to underlie various accounts of conscious versus unconscious processing (for review, see Mudrik et al., 2014).

Another intriguing aspect of the original results is that they were obtained using a task in which the target stimulus was separated from the prime numbers pair by a 1,200 ms mask, to allow enough time for the addition process to take place (commonly assumed to take at least 800 ms; Deschuyteneer & Vandierendonck, 2005; Geary & Wiley, 1991). Yet, the endurance of the effect over such a long period of time challenges another widely accepted claim in the field, namely that unconscious effects are typically short-lived, in the magnitude of no more than few hundred milliseconds (Greenwald et al., 1996; Kouider & Dehaene, 2007). The results of the study by Ric and Muller (2012), therefore, are of pivotal importance to our understanding of the depth and strength of unconscious processing, and the relations between consciousness and integration, alongside the more specific question of performing arithmetic operations without awareness.

Here, we accordingly wished to reexamine the findings of the Ric and Muller's (2012) study; first, given the general importance of replications as means to self-correct and improve our scientific praxis (e.g., Koole & Lakens, 2012; Nosek et al., 2012). This seems even more crucial in the field of unconscious processing, where the effects are typically relatively weak (Draine & Greenwald, 1998). Second, in light of the abovementioned criticism against other studies that demonstrated unconscious high-level integration (e.g., Biderman & Mudrik, 2017; Moors et al., 2017; Shanks, 2017; Tal, Sar-Shalom, et al., 2024), such a replication attempt seems especially warranted.

A further motivation for this study is to better control for subjects' awareness of the masked stimuli, in order to make sure that the effect—if found—indeed represents unconscious processing per se. In the original study, participants' awareness was assessed using a discrimination task at the end of the study, where they were asked to guess which pair of digits was presented in the previous trial out of all possible pairs using the digits 1–5 (so that chance level is 4%). This is problematic for several reasons. First, such a free-recall task is much more difficult than a forced-choice recognition task in which the pair is actually presented together with another pair. This makes it less sensitive and accordingly increases the chances of underestimating subjects' awareness level (Snodgrass, 2004). It further minimizes the chances to detect partial awareness (Kouider & Dupoux, 2004); in a recognition task, participants could perform above chance even if they only saw one of the digits, while in the free-recall task used by Ric and Muller (2012), seeing one of the digits does not guarantee a correct response. Second, this objective measure does not allow for an online, trial-by-trial assessment of participants' perceptual experience. Thus, the results might potentially be contaminated by trials in which participants actually managed to see some of the stimuli. A better approach would be to combine both objective and subjective measures, as is often done in the field (Reingold & Merikle, 1988; Seth et al., 2008). Third, this discrimination task only probed awareness of the digits and not awareness of the instruction.

Accordingly, in this registered report, we conducted a replication of the study by Ric and Muller (2012) under a preregistered and peer-reviewed protocol.[1] Our design followed the original procedure, yet with an additional subjective measure at the end of each trial, which allowed us to assess the depth of suppression and examine the possible relations between participants' awareness and the effect. In addition, we conducted two posttest objective tests to overcome the abovementioned limitations of the original study. Finally, to increase our chances to obtain the effect, we tripled the number of trials, hereby increasing the signal-to-noise ratio, and added a calibration session to determine individual perceptual thresholds for each subject, which allowed us to present the strongest stimuli possible while maintaining invisibility.

Notably, the changes we made in experimental protocol might have altered the prospects of obtaining an effect (e.g., due to the calibration and the subjective measures, subjects had knowledge about the masked primes, which may have affected their performance by evoking conscious strategies that were not present in the original study). Thus, we conducted an additional, more direct replication of the original study (Experiment 2) with a design identical to the original (yet with the tripled number of trials, to ensure a strong enough signal-to-noise ratio).

Transparency and Openness

This is a registered report. All methods have been peer-reviewed before data collection had begun. The experiment was preregistered in the Open Science Framework website, including sample size and subject exclusion criteria. Below, we report how we determined our sample size, all data exclusions, all data inclusion/exclusion criteria, whether inclusion/exclusion criteria were established prior to data analysis, all manipulations, and all measures in the study. All experimental materials, the data, and analyses codes can be found in the APA repository hosted by the Center for Open Science at https://osf.io/btm7v (Tal & Mudrik, 2024).

Experiment 1

Method

Participants

One hundred twenty-six people participated in the study, in line with our predefined sample size (85 identified as female, 41 as male; age: $M = 24.42$ years, $SD = 3.33$; 113 right handed). A sample size of 126 participants was determined using the G*Power software (Faul et al., 2009), with a planned power of 95% and α level of 0.05 (one group, two measurements, since we are interested in the interaction, calculated as the difference between the differences, epsilon = 1, correlation among repeated measures = 0.21, as found in the original data), so to be able to detect an effect that is greater or equal to $\eta_p^2 = 0.04$ (between a weak and moderate effect) for the critical interaction in Experiment 3 described in Ric and Muller (2012). No data analysis was conducted until predefined sample size has been collected, besides making sure that all collected participants do not meet the exclusion criteria.

However, to reach this sample size, more than double the number of participants had to be run ($N = 269$; 173 identified as female, 95 as male, one as other; age: $M = 24.75$ years, $SD = 3.51$; 240 right handed), yielding an extremely high exclusion rate (53.2%). Note that such a high exclusion rate raises the concern of regression to the mean leading to a false positive result (Shanks, 2017). However, since no effect was found, this concern does not apply here (yet see a preregistered analysis to account for this issue in the Supplemental Materials).

Participants were excluded due to different criteria (since one criterion refers to the number of trials, participant exclusion was performed after trial exclusion; see Results section below): Seventy-four participants (27.5%) performed above chance level in at least one of the two posttests, indicating visibility of masked primes. Two participants (0.7%) did not appear to understand the experimental

[1] The authors thank the editors, Richard Morrey and Nelson Cowan, the reviewers—Pieter Moors, Fronçois Ric, and David Shanks—for making excellent suggestions that helped us substantially improve the design. Further, the authors thank again Fronçois Ric as well as Dominique Muller for sharing their data, experimental codes, and materials. Last, the authors thank Yoni Amir for his help in composing the English fluency test used for Experiment 2 and research assistants Sean Chady, Tzlil Lijishal, Shai Fischer, Karin Uritsky, Karmi Patt Shamir, and Itamar Duek for their help with data collection.

instructions, based on their postexperiment debriefing. Two participants (0.7%) reported seeing the unconscious stimuli and/or understanding the manipulation of the study. Twenty-one participants (7.8%) performed below 75% accuracy in the main task, and hence were considered not to understand/follow the experimental instructions. Eighty-seven participants (32.3%) had less than 14 usable trials in at least one of the experimental conditions (less than the original trial amount in Ric & Muller, 2012), and hence the signal-to-noise ratio from these participants was considered too low to be included.

Importantly, only the first criterion was defined in the registered report, as it has to do with unconscious processing (see Participant Exclusion section). The other three criteria have been added during data collection, upon encountering data of problematic quality. To make sure these exclusions have not compromised the validity of reported results, we also analyzed the data keeping only exclusion Criterion 1 (i.e., with 195 subjects), and the main findings remained the same (see Supplemental Materials).

All participants were native English speakers (i.e., had English as their mother tongue), and had normal or corrected-to-normal vision. Participants received monetary compensation or Tel Aviv University course credit for their participation. To confirm that English was indeed their mother tongue, participants were screened prior to the experiment based on a nonstandardized questionnaire where they were asked to rate their reading fluency in English (poor/mediocre/excellent) and report what was the first language they spoke and what was the first language they read. Only participants who reported excellent reading skills and defined English as their primary language (i.e., English was the first language they spoke and the first language they read) were included. All participants completed a consent form prior to taking the study. Gender was assessed with the question "Please specify your gender (M/F/O)," options indicating the responses "male," "female," and "other," respectively.

Apparatus

Stimuli were presented on a LCD monitor (23" ASUS SyncMaster) with 1920 × 1080 resolution and 100 Hz refresh rate. Stimulus presentation was done using Matlab and Psychtoolbox 3 (Brainard, 1997; Pelli, 1997). Participants sat in a dimly lit room, by themselves, 60 cm away from the monitor.

Procedure and Stimuli

In this conceptual replication, the procedure followed that of Ric and Muller (2012; Experiment 3), except for the following changes: (a) using the English words "add" and "feel" instead of the French words "additioner" and "representer"; (b) having a trial-by-trial subjective measure; (c) adding two calibration sessions aimed at estimating the individual perceptual threshold of each participant for each stimulus; (d) having two posttest sessions aimed at estimating participants' objective performance on the masked primes, (e) tripling the number of trials, to increase the signal-to-noise ratio and make sure the obtained effects are reliable (in the original study, there were only 14 trials in each experimental cell).

Accordingly, the experimental trials were divided according to a 2 × 2 design, with instruction (add/feel) and target-prime relations (prime digits equal sum/prime digits do not equal sum). In addition, nonexperimental trials were identical to the experimental ones, but the target was a letter and not a digit. These trials were not analyzed. Trial order was pseudorandomized with the constraint that the correct response was not identical in more than three consecutive trials.

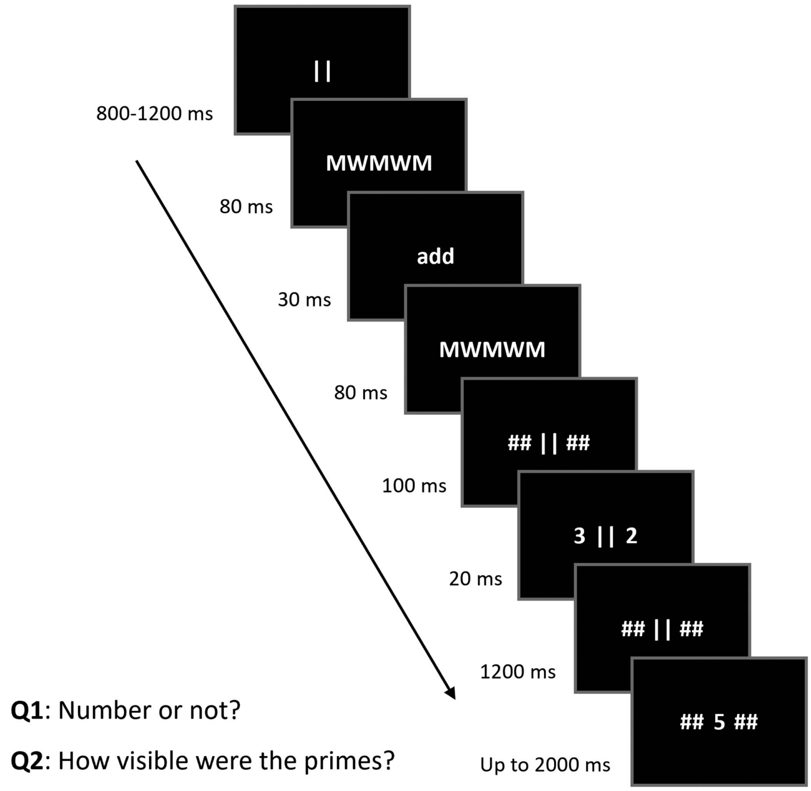
In all sessions of the experiment, trials had the same fixed sequence of stimuli (presented as white on top of a black background). In each trial, a fixation point (800–1,200 ms) appeared, followed by a mask (MWMWMWMW) for 80 ms. Then, the prime instruction appeared (*add* or *feel*) for 30 ms, followed by another mask for 80 ms (see Figure 1, Panel A). This was followed by a second mask ("## | | ##"), presented for 100 ms, and the two digits prime (henceforth, "flankers"), presented for 20 ms. The flankers were positioned to be perceived at the same time (angular distance of 4.5°). They were followed by a mask ("## | | ##") for 1,200 ms, and then, finally, the target stimulus appeared. The target was presented at the center of the screen, between the masks, replacing the fixation sign (e.g., "## 4 ##"); in half the trials, the target was a number (i.e., 3, 4, 5, 6), and in the other half—a letter (i.e., the letters A, B, C, and D). When the target was a number, it was either equal to the sum of the two, or not (for full trial sequence, see Figure 1). Importantly, there was no overlap in content between the target digit and the flankers which preceded it. And, in trials where the target was different from the sum of the flankers, the targets was equally distributed above and below the sum value. In all but the practice session, trials were self-paced.

At the beginning of each session, the experimenter instructed participants about the trial sequence and the upcoming task. They emphasized the importance of reporting subjective awareness as accurately as possible, and paying close attention to the stimuli throughout the experiment, despite its repetitive and potentially boring character. Below, we describe the different sessions of the experiment.

Practice Session. The experiment began with four example trials shown to the participants in order to familiarize them with the stimuli. These sequences were advanced by the experimenter, requiring only passive viewing from the participants.

Calibration Sessions. The aim of the calibration sessions was to determine individual thresholds for the primes and masks contrasts (defined by image transparency, using Matlab's α function). This allowed us to present the stimuli at their strongest version (in terms of contrast), while keeping them invisible, so to maximize the chances to obtain an effect under unconscious processing. Accordingly, we maximized the chances of obtaining an effect as we focused on the strongest signal we could obtain for each participant, while assuring that the stimuli remained largely invisible throughout the experiment. One session focused on the word prime, and the other—on the numbers prime (order counterbalanced between participants). Each session included 100 trials, in which the prime of interest was either presented upright or inverted (randomly intermixed, with the constraint that prime orientation was never the same in four consecutive trials). A staircase procedure (Levitt, 1971) was used, with the initial masks contrast set to 0.85 and the initial prime contrast set to 0.7. Masks and prime contrasts changed based on participants' performance on the orientation judgment task, in which they were asked to determine if the prime was presented upright or inverted (Biderman & Mudrik, 2017). When participants were correct in their response, masks contrast were ramped up by 0.05; when they were incorrect, they were ramped down by 0.05. If masks contrast reached 1, prime contrast was reduced using the same criterion and the same contrast units (0.05), with a low boundary of 0.1. At the end of the calibration session, masks contrast was set to the second highest level reached throughout the last 40 trials of the

Figure 1
Experimental Sequence of the Experiment



Note. Each trial started with a fixation, which was followed by a sandwich-masked instruction prime ("add" or "feel"). The instruction prime was in turn followed by an additional sandwich-masked prime, in the form of a pair of digits (flankers). Then, a target appeared, either a letter or a digit. The digit was either the sum of the two flankers, or not. In the main experimental session, participants were asked to first determine if the target was a number or not, and then to rate how visible were the primes on the Perceptual Awareness Scale. Q1 = first question per trial; Q2 = second question per trial.

session—or, if masks contrasts reached 1, it was set to 1—and prime contrast was set to the second lowest level reached throughout the last 40 trials of the session.

Main Session. The main session included three blocks of 112 experimental trials each. At the end of each trial, participants were prompted to determine as fast as possible whether the target was a number or not, using the right and left arrow keys, counterbalanced between participants (note that in the original article, the keys A and P were used. Yet here, since subjects also report prime visibility, we asked them to use one hand for the primary target task, and the other for the visibility report). The target remained on screen for 2 s, unless participants responded earlier, after which it was replaced by the fixation stimulus until key press. Then, participants were asked to rate to what extent they were able to see either the instruction or the flankers on a scale of 1–4 (Perceptual Awareness Scale or PAS; Ramsøy & Overgaard, 2004), where 1 was "I didn't see anything," 2 was "I had a vague perception of something," 3 was "I saw a clear part of either the numbers or the word," and 4 was "I saw either the numbers or the word clearly."

Posttest Discrimination Sessions. After the main session, participants underwent two posttest discrimination sessions (N = 140 trials each). In each trial, the same experimental stimulation took place, yet at the end of the trial participants did not answer a question about the target, but instead were presented with a question about the prime—in one posttest session, about the flankers, and in the other—about the instruction (order counterbalanced between subjects). In the flankers posttest, two pairs of digits were presented—the actual pair that appeared as prime in the trial and a "foil" pair (location randomly determined). The foil pair was made of two digits that were not members of the actual prime, randomly selected from the other available digits between 1 and 5. Notably, this approach sets a strict objective measure for awareness, because consciously perceiving only one of the flankers would still yield above-chance performance (since there is no digit repetition between the actual pair and the foil pair), even though it is not sufficient for priming (which requires the integration of both flankers). Thus, our approach assured that the results could not be explained by partial awareness of one of the digits in the pair (Kouider & Dupoux, 2004). In the instruction posttest, the two instructions appeared (location randomly determined). In both posttests, the correct answer did not repeat in the same location for more than three consecutive trials.

In both sessions, participants were asked to determine which one of the two stimuli was presented during the trial, and—in case they did not perceive them—guess. Because targets appeared until response during the main session, this two-alternative choice was presented following the average time it took to respond to the target in the main session (per each participant according to their average RT), in order to present the target for the same amount of time as during the main session. The subjective measure rating question was then presented, similar to the main session. Note that we chose to use a posttest objective measure rather than a trial-by-trial one. Though this might introduce fatigue effects on the one hand (yielding lower performance) and practice effects (yielding higher performance) on the other, we thought that this is a better choice. Mainly, this was done to avoid a mismatch between participants' response about the target (digit vs. letter) and the prime (pair recognition), which might introduce confusion and reduce the chances of obtaining an effect (Schmidt & Vorberg, 2006). Notably, previous studies have shown that when participants report having the lowest level of stimulus visibility using the PAS, their performance on a trial-by-trial objective measure is at chance (Peremen & Lamy, 2014; see also Biderman & Mudrik, 2017; Mudrik & Koch, 2013; Ramsøy & Overgaard, 2004), so our usage of a trial-by-trial subjective measure combined with two posttest objective measures should have provided sufficient means to better assess participants' awareness of the primes.

Yet another concern refers to participants' frustration and different mindset, compared with the main experiment. While there, they were able to follow the task, here they were mostly guessing, and not seeing the stimuli on which they were probed. This may lead to them ceasing to pay attention to the task, yielding near-chance performance which does not necessarily reflect their level of awareness (Finkbeiner, 2011; Pratte & Rouder, 2009). To overcome this issue, the posttest included catch trials (25% of the trials; $N = 35$) in which the contrast of the primes was set to four levels above the individual threshold set for that participant (i.e., +0.2 opaqueness, see above) and the contrast of the prime's mask to four levels below the individual threshold. These trials, which were not included in the analysis, served as "motivation boosters", making the task less difficult, and accordingly—getting subjects less frustrated and more engaged in it. Note that since participants did not know which trial includes a visible prime and which trial includes an invisible one, they could not change their strategy accordingly. To avoid carryover effects ("awareness priming"; Lamy et al., 2017; Lin & Murray, 2014) from visible trials that might affect prime judgment in the upcoming invisible trial, prime content was not repeated between visible and invisible trials.

Analysis

Trial Exclusion. Similar to Ric and Muller (2012), we excluded incorrect trials, and trials in which participants' response time (RT) was shorter than a minimal cutoff threshold of 300 ms or longer than a maximum one of 1,000 ms. In addition to these thresholds, used in the original study, lower boundaries of 250 ms, 350 ms and 400 ms were tested, in combination with upper boundaries of 1,500 ms and 2000 ms, along with individualized thresholds of 1.5 *SD* below and above mean participant RT, to minimize bias caused by the cleaning method (Ratcliff, 1993; Rougier et al., 2018). Furthermore, we inverse-transformed the data, as was also done in the original analysis scheme. Critically, we only used trials in which participants rated the visibility as "1" (that is "I didn't see anything"), with an exception when individuals reported more Visibility 2 trials then Visibility 1 (see below).

Participant Exclusion. Exclusion was based on performance in posttest sessions (following Hesselmann et al., 2016). First, individual performance in Visibility 1 trials during posttest was submitted to a binomial test (chance level: 50%; α: .05). Participants performing above chance were excluded from analysis. Next, participants reporting more Visibility 2 than Visibility 1 trials in both posttest sessions were treated separately. A binomial test, like before, was performed on Visibility 1 combined with Visibility 2 trials of these participants, and participants performing above chance according to this test, in either of the posttests, were excluded from analysis. Participants who were at chance in both sessions when including both Visibility 1 and Visibility 2 trials remained in analysis, and their Visibility 2 trials were combined with Visibility 1 trials in all forthcoming analyses, including those of the main session. The rationale for this approach is that consistent report of Visibility 2 accompanied by chance performance might represent a response bias rather than an actual perceptual difference (i.e., the subject might be too conservative in their ratings), warranting response normalization. This approach increases the power (both in terms of subjects and trials) and minimizes subject exclusion, while enforcing strict within-participant performance and response bias tests (see again Hesselmann et al., 2016). However, given recent criticism against participant exclusion in studies of unconscious processing (Shanks, 2017), analyses in this study were performed and reported both for included subjects only, and for the complete sample size, estimate if the effects might stem from selection processes and regression to the mean.

Confirmatory Analyses

Priming Effect. A two-way repeated-measures analysis of variance analysis with prime instruction (add vs. represent) and target (sum of prime digit vs. not sum) was conducted on both reaction times and accuracy. Note that we decided not to use a regression analysis, as was done in the original study, given criticism against this approach (Greenwald & De Houwer, 2017; Sand & Nilsson, 2016). Importantly, we obtained the original data from the authors and were able to find effects there using analysis of variance, $F(1, 31) = 5.87$, $p = .021$, $\eta_p^2 = 0.16$. We accordingly expected to find an interaction for reaction times, which would have been followed by a planned comparison for the effect of target within each prime instruction level. We did not expect to obtain an effect on accuracy, as no such effect was found in the original study.

In addition to the above Null Hypotheses Statistical Testing (NHST), we conducted Bayesian analyses in order to estimate the strength of the evidence in favor or against an effect. Bayes factors (BF) were obtained from the statistical software JASP, by comparing a full model that includes all main effects and interactions with a reduced model in which the effect of interest (i.e., the interaction) is not included. Default settings were used (r scale fixed effects = 0.5, r scale random effects = 1, r scale covariates = 0.354, model prior = uniform). We adopted the convention that a BF less than 0.1 implies strong evidence for the lack of an effect (i.e., H_0 is at least 10 times more likely than H_1 given the data), a BF between 0.1 and 0.33 provides moderate evidence for the lack of an effect, a BF between 0.33 and 3 suggests insensitivity of the data (anecdotal evidence for the lack or presence of an effect, for $0.33 < BF < 1$ or $1 < BF < 3$, respectively), a BF between 3 and 10 denotes moderate evidence for the presence of an effect (i.e., H_1), a BF between 10 and 100 implies strong evidence, and a BF greater than 100 suggests extreme evidence for the presence of an effect (Lee & Wagenmakers, 2013). In our preregistration we planned to use the BayesFactor package (Version 0.9.11-1; Morey & Rouder, 2015) for the R software environment (R Core Team, 2015) for Bayesian analyses. We eventually opted for using JASP instead for convenience, as it was also used to calculate effect sizes (η_p^2) per NHST analyses.

Analysis of Block 1 Trials Only. The current experiment tripled the number of trials which were included in the original experiment. To make sure this did not somehow reduce the chances of obtaining an effect, we also ran the same analyses detailed in Section A above on data obtained in the first block only, which has the exact number of trials as in the original study. Yet, given the low number of trials in each experimental cell in this analysis, we expected the analysis of all trials to yield more reliable results.

Assessment of Prime Visibility. We report the distribution of participants′ visibility ratings both in the main session and in the posttests (without conducting any analysis on them, as we have no hypotheses/research questions with respect to the ratings). Subsequent analyses then focused on Visibility 1 trials only (with the exception described in Participant exclusion with respect to correcting for response bias). We also examined if participants" performance in each of the posttests was different from chance, both on accuracy and on *d*′. *d*′ was calculated using the Palamedes toolbox (Prins & Kingdom, 2009). In extreme cases, hits and false alarm rates of 0 were replaced with 0.5/*n*, and those of 1 were replaced with (*n* − 0.5)/*n*, where *n* is the number of signal or noise trials, respectively (Stanislaw & Todorov, 1999). Like all other analyses, performance was compared to chance using both traditional NHST (*t*-tests) and Bayesian testing (BF *t*-test with default settings, see above). This is especially critical here, as the goal is to demonstrate chance performance (which shows that participants were indeed unaware of the stimuli). For that end, Bayesian statistics are helpful as they allow assessing the strength of evidence in favor of H_0, as opposed to the NHST approach (Rouder et al., 2009), and they were shown to be more sensitive to instances where the effect actually depended on awareness (Sand & Nilsson, 2016).

Results

Trial Exclusion

This step was conducted on the entire data set (prior to participant exclusion; see Method section above). The following exclusion rates accordingly pertain to the entire data set, including the excluded participants. First, 6.10% of trials were excluded for being incorrect. This suggests that participants were overall on task and followed instructions. In terms of speed, 3.71% of responses were quicker than 300 ms and 19.12% were slower than 1,000 ms. Finally, in 0.02% of trials, participants clicked both response buttons simultaneously. Taken together, 25.88% of trials were excluded from the analysis due to at least one of these three preregistered criteria. Double responses led to the exclusion of 0.35% of the instruction posttest trials and 0.23% of the digits posttest trials.

Assessment of Prime Visibility

Most of the trials were given the rating of 1 (*I didn't see anything*) in the perceptual awareness scale (PAS; Ramsøy & Overgaard, 2004) during test (M = 82.0%, SD = 17.5%), and during both the instruction posttest (M = 74.7%, SD = 26.1%) and digits posttest (M = 75.3%, SD = 28.3%; Figure 1).

Following our correction for response bias, more trials were considered invisible than those rated PAS = 1 (PAS ≤ 2 for a few participants; see *Subject exclusion*): M = 86.0%, SD = 12.2% in test, M = 80.6%, SD = 20.5% in instruction posttest, M = 81.1%, SD = 23.0% in digits posttest. In these trials, performance in the objective task during both posttests did not differ from chance level in the included sample: instruction posttest, M = 50.3%, SD = 5.6%, $t(125) = 0.57$, $p = .568$, $BF_{10} = 0.116$, and digits posttest, M = 50.0%, SD = 5.8%, $t(125) = 0.01$, $p = .992$, $BF_{10} = 0.173$; Figure 2 left panel. Groupwise *d*′ sensitivity did not differ from zero as well: instruction posttest, $M = 0.02$, $SD = 0.30$, $t(125) = 0.91$, $p = .367$, $BF_{10} = 0.148$, and digits posttest, $M = 0.02$, $SD = 0.30$, $t(125) = 0.70$, $p = .483$, $BF_{10} = 0.126$; Figure 2 right panel. Together, these results confirm that the masking manipulation of the study worked well for the included sample.

Priming Effect

A repeated-measures two-way analysis of variance with trial instruction ("add" or "feel") and target type (equal or different from the sum of the prime digits) as factors was conducted over invisible trials. No priming effect was found, hereby failing to replicate the original finding.

In reaction time ($M = 30.05$, $SD = 6.51$ trials per experimental condition; Figure 3 left panel), no main effects were found for trial instruction, $F(1, 125) = 1.541$, $p = .217$, $\eta_p^2 = 0.012$, nor for target type, $F(1, 125) = 0.100$, $p = .752$, $\eta_p^2 < 0.001$. Critically, and different from the original study, no interaction was found between the two conditions, $F(1, 125) = 0.192$, $p = .662$, $\eta_p^2 = 0.002$. These null results were confirmed also by Bayesian analysis, with evidence moderately favoring the null hypothesis ($BF_{10} = 0.139$). A similar null result was also found for participants' response accuracy ($M = 30.71$, $SD = 6.55$ trials per experimental condition; Figure 3 right panel): No main effects were found for trial instruction, $F(1, 125) = 0.297$, $p = .587$, $\eta_p^2 = 0.002$, nor for target type, $F(1, 125) = 1.394$, $p = .240$, $\eta_p^2 = 0.011$, and neither was an interaction effect, $F(1, 125) = 0.757$, $p = .386$, $\eta_p^2 = 0.006$, BF_{10} = 0.285; notably though, no effect of accuracy was found in the original work as well.

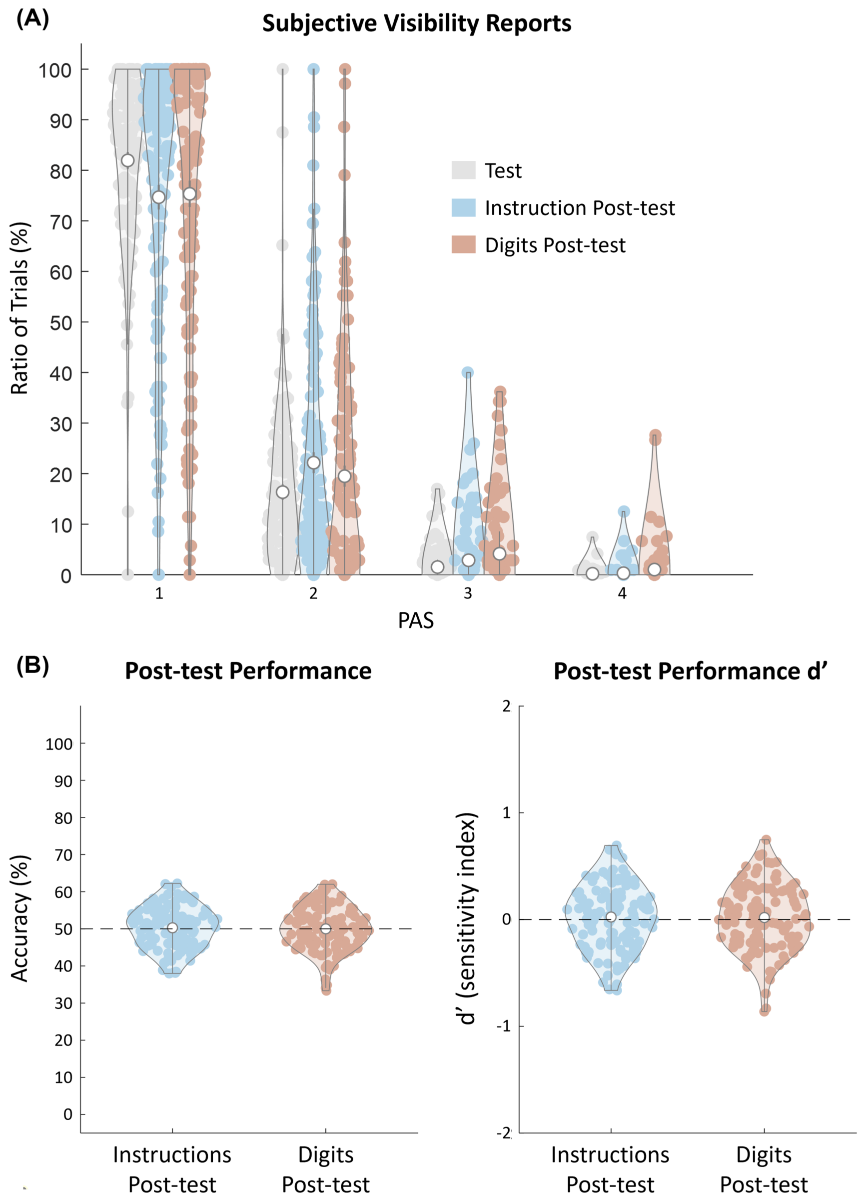
In line with our preregistered plan, we carried out several additional analyses to examine the robustness of these null findings (detailed in Supplemental Materials). Specifically, priming effects were examined under different participant exclusion criteria: applying only the preregistered participant exclusion criterion ($N = 195$) and applying all but that exclusion criterion ($N = 179$). They were also examined under different trial exclusion criteria: using different absolute and individualized RT cutoff values and including only the first 112 trials of the experiment. All results were in line with the results of the main analysis, yielding no evidence for priming caused by the interaction between instructions and target type.

Conclusion

The findings of Experiment 1 provide compelling evidence toward a null result. The original effect, showing that initiation and execution of arithmetic addition can occur when both the instruction

Figure 2

Visibility Ratings and Objectiive Performance for Masked Stimuli During Experiment 1

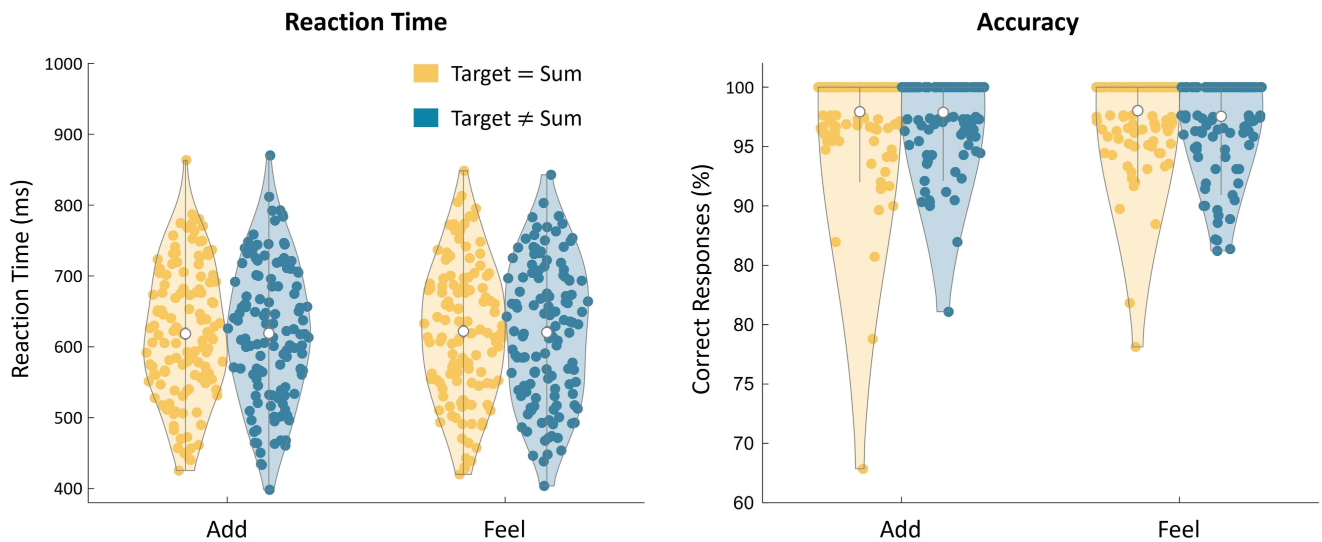


Note. (A) Subjective visibility (PAS ratings) during test and both posttests indicate that participants did not experience seeing the stimuli in most trials. (B) Accuracy (left panel) and *d′* scores (right panel) in invisible trials of both posttests indicate objective performance at chance level. Colored dots represent individual subjects, white circles represent the mean, and whiskers standard error of the mean. These conventions will be used in all figures. PAS = Perceptual Awareness Scale. See the online article for the color version of this figure.

to add and the digits to add are not consciously visible, was not replicated in our design, which was substantially better powered, and applying more stringent awareness measures.

However, this null result might stem from the differences between our paradigm and the original experiment. Most importantly, the design of Experiment 1 included an online measure of awareness. This may have caused participants to divide their resources between two separate tasks: the main task of judging whether a character is a letter or a digit, and a secondary task of judging the visibility of preceding primes, thereby resulting in a failure to reproduce the original effect. Indeed, such a cost for online measures have been previously reported in the literature (Fischer et al., 2011; Kiefer et al., 2023). To address this concern, we conducted a second experiment in which no online visibility measures were taken, and no calibration session was conducted. Thus, the experiment was a direct replication of the design in Ric and Muller (2012), yet

Figure 3
Lack of Unconscious Priming in Experiment 1



Note. Recognizing targets as digits was not quicker (left panel) nor more accurate (right panel) when they were preceded by two masked digits that sum to it and a masked instruction to "add." See the online article for the color version of this figure.

containing 336 trials instead of the original 112, to increase the power. The main experimental session was again followed by two posttest sessions to assess visibility at the end of the experiment.

Experiment 2

As the first experiment did not yield evidence for unconscious arithmetic operations, an additional experiment was carried out. This experiment was a more direct replication of the original study, to make sure the failure to replicate did not stem from the difference between the new design and the original one.

Method

Participants

One hundred twenty-six participants comprised our predefined sample for Experiment 2 (81 identified as female, 43 as male, and two as other; age: $M = 26.09$ years, $SD = 3.97$; 106 right handed). Here, as there was no criteria for objective performance (see Experiment 2 *Apparatus, Procedure, and Stimuli*), exclusion rates were much lower (18.2%): Twenty-eight participants reported seeing or understanding the manipulation of the experiment and were accordingly excluded. Overall, therefore, 154 participants were run (96 identified as female, 56 as male, and two as other; age: $M = 25.94$ years, $SD = 4.01$; 133 right handed).

Apparatus, Procedure, and Stimuli

Experiment 2 shared all materials and methods of Experiment 1 (including the tripled number of trials), yet no subjective measures was taken during the main experimental session, and no calibration session was administered, akin to the original study. Thus, contrast of primes and masks was set to maximal ($\alpha = 1$) for all participants, and participant exclusion was based on retrospective subjective report of seeing any of the masked stimuli, following the original design of Ric and Muller (2012). Importantly, these subjective reports were taken after the main session and before the posttest sessions, to comply with the original design.

An additional change, which deviated from the preregistered protocol yet was approved by the journal editor, pertained to participant recruitment. Due to the difficulty we encountered in Experiment 1 to recruit a large volume of native English speakers at Tel Aviv University, the inclusion criterion was relaxed from native speakers to highly fluent speakers. This allowed us to include participants whose first language was not necessarily English, as long as they were highly fluent in it. To test for fluency, we used a battery of questions from the English section of a psychometric test by the Israeli National Institute for Testing and Evaluation. This test is used in Israel to assess the fit of prospective students to higher education (similar to the American Scholastic Aptitude Test). In Israel, students who answer 85% of the English section questions correctly are considered fluent and are thus exempt from taking English courses during their degrees. Therefore, prospective participants answered 13 questions we selected (eight questions of sentence combinations and five questions about an unseen text), and were only invited to the study if they answered at least 11 of them correctly.

Analyses

Analyses were identical to the ones performed on Experiment 1 data.

Results

Trial Exclusion

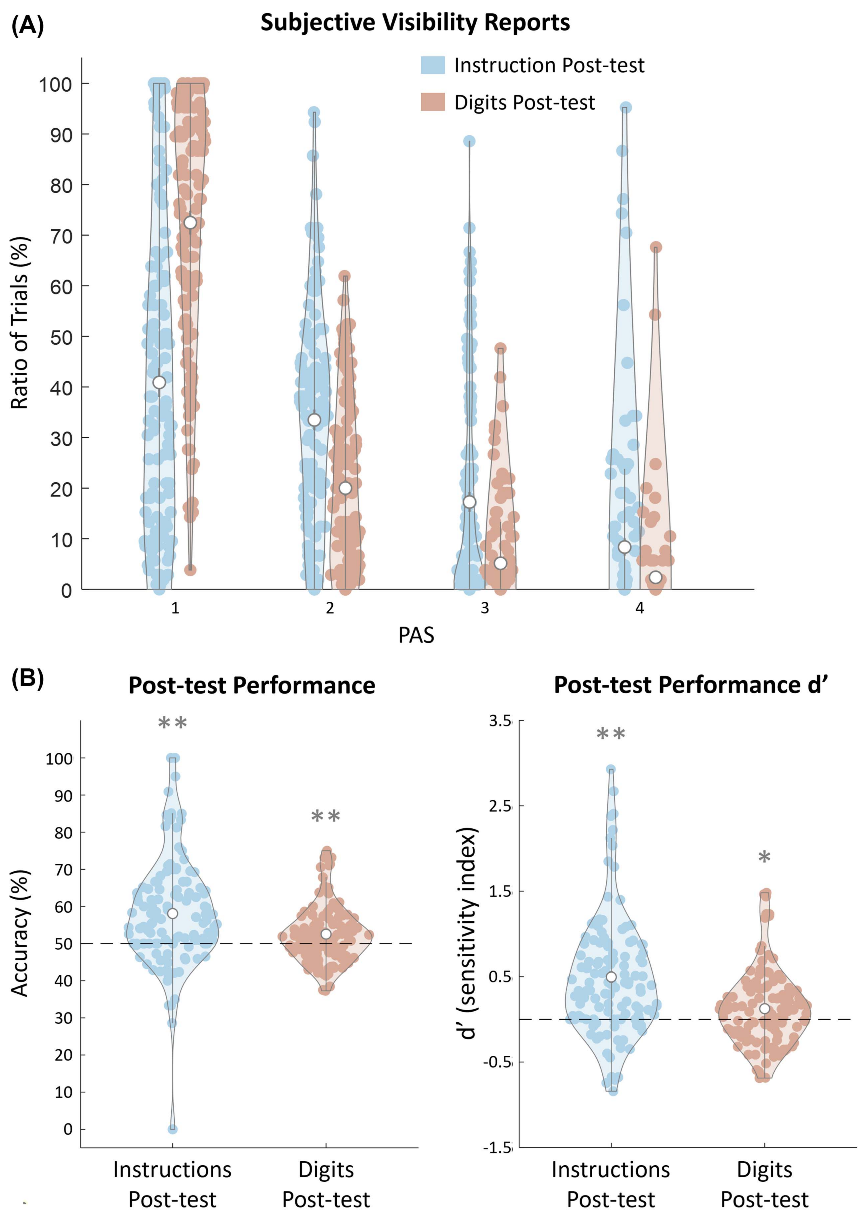
As in Experiment 1, accuracy was high, leading to the exclusion of only 2.40% of trials due to incorrect responses. In addition, 0.09% of trials were shorter than 300 ms and 3.81% longer than 1,000 ms, and were subsequently removed from analysis. Together, 6.14% of test trials were excluded from analysis. Double responses led to negligible exclusion of test and posttest trials (under 0.01%).

Assessment of Prime Visibility

In Experiment 2, invisibility (PAS = 1) remained the most prevalent response both in the instruction posttest (M = 40.9%, SD = 30.8%) and in the digits posttest (M = 72.5%, SD = 25.3%; Figure 4A). Here, as opposed to Experiment 1, performance under PAS = 1 trials in the objective task exceeded chance, in both the instruction posttest, M = 58.1%, SD = 15.3%, $t(124) = 5.91$, $p < .001$, BF_{10} = 357973.672; one subject did not have PAS = 1 trials in this posttest and hence could not contribute to the analysis, and in the digits posttest, M = 52.6%, SD = 7.6%, $t(125) = 3.78$, $p < .001$, BF_{10} = 74.794; Figure 4B. Groupwise d' sensitivity was

Figure 4

Visibility of Masked Stimuli During Experiment 2 Posttests



Note. (A) Subjective reports in the digits posttest indicated that masked information was mostly not seen (PAS = 1). (B) Objective performance (left panel) and signal detection sensitivity (right panel) in trials rated as invisible were above chance in both posttests. PAS = Perceptual Awareness Scale. See the online article for the color version of this figure.

* $p < .01$. ** $p < .001$.

accordingly greater than zero in both cases: instruction posttest, $M = 0.50$, $SD = 0.70$, $t(120) = 7.83$, $p < .001$, $BF_{10} = 3.517 \times 109$, and digits posttest, $M = 0.12$, $SD = 0.44$, $t(124) = 3.19$, $p = .002$, $BF_{10} = 11.972$; six participants' PAS = 1, trials did not include at least one type of posttest trial, and hence could not contribute to a d' analysis; Figure 4C. Together, these results demonstrate that exclusion based solely on postexperiment debriefing, as used in Ric and Muller (2012), may include trials in which participants were aware of the stimuli, and participants whose objective performance—even under trials they have rated as invisible—still exceeds chance (indicating either subjective-but-not-objective invisibility, akin to blindsight in healthy participants, H. C. Lau & Passingham, 2006; Meeres & Graves, 1990, or simply a too strict criterion in rating their visibility, such that even when they did have some conscious experience, they still gave a "didn't see" rating).

Priming Effect

Although Experiment 2 was a direct replication of the original study, and even though objective performance was above chance (potentially suggesting some residual awareness), no evidence of priming was found in reaction time. Neither the main effects, trial instruction: $F(1, 125) = 1.535$, $p = .218$, $\eta_p^2 = 0.012$; target type: $F(1, 125) = 0.498$, $p = .482$, $\eta_p^2 = 0.004$, nor the interaction, $F(1, 125) = 0.057$, $p = .812$, $\eta_p^2 < 0.001$; $M = 39.38$, $SD = 2.85$ trials per experimental condition; Figure 5 left panel, were significant. These null results were again confirmed using Bayesian analysis, with moderate evidence favoring the null hypothesis ($BF_{10} = 0.153$). A similar null result was also found in participants' response accuracy, main effect of trial instruction: $F(1, 125) = 0.008$, $p = .929$, $\eta_p^2 < 0.001$; main effect of target type: $F(1, 125) = 1.771$, $p = .186$, $\eta_p^2 = 0.014$; interaction effect: $F(1, 125) = 0.659$, $p = .418$, $\eta_p^2 = 0.005$, $BF_{10} = 0.207$; $M = 40.38$, $SD = 2.43$ trials per experimental condition; Figure 5 right panel.

As in Experiment 1, we also carried out these analyses on the entire sample collected for the study, and using different trial exclusion criteria, and none yielded an interaction effect in RT nor in accuracy (see Supplemental Materials).

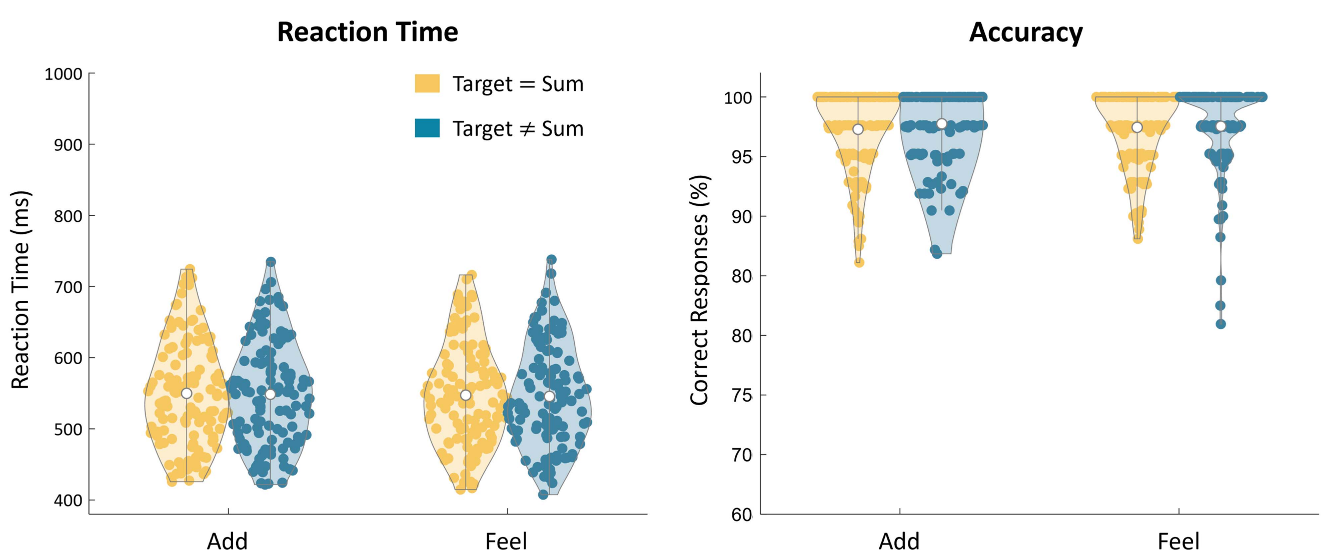
Conclusion

Albeit staying faithful to the design of Ric and Muller (2012), Experiment 2 failed to replicate the results reported there. As in Experiment 1, we found no evidence for instruction-based arithmetic without awareness. Our conclusion that this cannot be carried out unconsciously is thus strengthened, as the null finding cannot be ascribed to differences with the original paradigm.

Aggregated Analysis

Given our failure to replicate the original results, we wanted to further test if additional power afforded by a larger sample would have allowed us to detect an effect. Thus, we conducted an additional exploratory analysis, where priming was examined in data aggregated across the main samples of both experiments, resulting in a sample size of $N = 252$. Here again, evidence for priming was not found in main effects of instruction, RT: $F(1, 251) = 0.032$, $p = .858$, $\eta_p^2 < 0.001$; accuracy: $F(1, 251) = 0.186$, $p = .667$, $\eta_p^2 < 0.001$, in main effects of target type, RT: $F(1, 251) = 0.478$, $p = .490$, $\eta_p^2 = 0.002$; accuracy: $F(1, 251) = 0.003$, $p = .959$, $\eta_p^2 < 0.001$, and, most importantly, in their interactions, RT: $F(1, 251) = 0.244$, $p = .622$, $\eta_p^2 < 0.001$, $BF_{10} = 0.123$; accuracy: $F(1, 251) = 1.421$, $p = .234$, $\eta_p^2 = 0.006$, $BF_{10} = 0.234$. Notably, this sample size corresponds to a power of 99.9% of detecting our effect of interest.

Figure 5
Lack of Unconscious Priming in Experiment 2



Note. Reaction time (left panel) and accuracy (right panel) were not found to be affected by the experimental conditions nor their interaction. See the online article for the color version of this figure.

Comparison Between Experiments 1 and 2

Online measures of awareness were removed from the design of Experiment 2 due to the concern that they might hamper priming effects (that may have otherwise existed in Experiment 1). This relates to an ongoing methodological debate in the field regarding the costs and benefits of online measures of visual awareness in studies of unconscious processing (Avneon & Lamy, 2018; Fischer et al., 2011). We thus leveraged the two large and robust data sets of these experiments to conduct an additional exploratory analysis. We compared participants' performance in the two experiments, to better assess the costs of using online measures, as opposed to testing for consciousness in dedicated posttest sessions.

As can be expected, and in line with the previous studies (Avneon & Lamy, 2018; Hesselmann et al., 2018; Kiefer et al., 2023), RT in Experiment 1 (which included the additional visibility judgment task per trial) was slower than in Experiment 2, which comprised a single task only; $t(250) = 7.57$, $p < .001$; Figure 6 left panel. Also, within-subject standard deviation of RT was greater in Experiment 1 than in Experiment 2, $t(250) = 7.76$, $p < .001$; Figure 6 middle panel, implying that the addition of visibility measures yielded noisier (i.e., more variable) data. Follow-up analyses revealed a decrease over time in RT (calculated over six epochs of 56 trials each), both in average RT, $F(3.4, 852.2) = 46.021$, $p < .001$, $\eta_p^2 = 0.155$, and in RT variability, $F(4.5, 1129.4) = 33.581$, $p < .001$, $\eta_p^2 = 0.118$, and that these decreases were greater in Experiment 1 than in Experiment 2, interaction between epoch and experiment; RT average: $F(3.4, 852.2) = 3.751$, $p = .008$, $\eta_p^2 = 0.015$, $BF_{10} = 2.882$; RT *SD*: $F(4.5, 1129.4) = 6.783$, $p < .001$, $\eta_p^2 = 0.026$, $BF_{10} = 2601.219$. Mauchly's test of sphericity yielded $p < .05$ in these analyses, and so Greenhouse–Geisser corrected values are reported. This suggests that the slower and more variable RT associated with online visibility measures may be reduced by training during the task (and so reflect task difficulty; see General Discussion section). Accuracy, on the contrary, was equivalent across experiments, $t(250) = 0.73$, $p = .465$; Figure 6 right panel, indicating that it was not compromised by the added online measure of awareness.

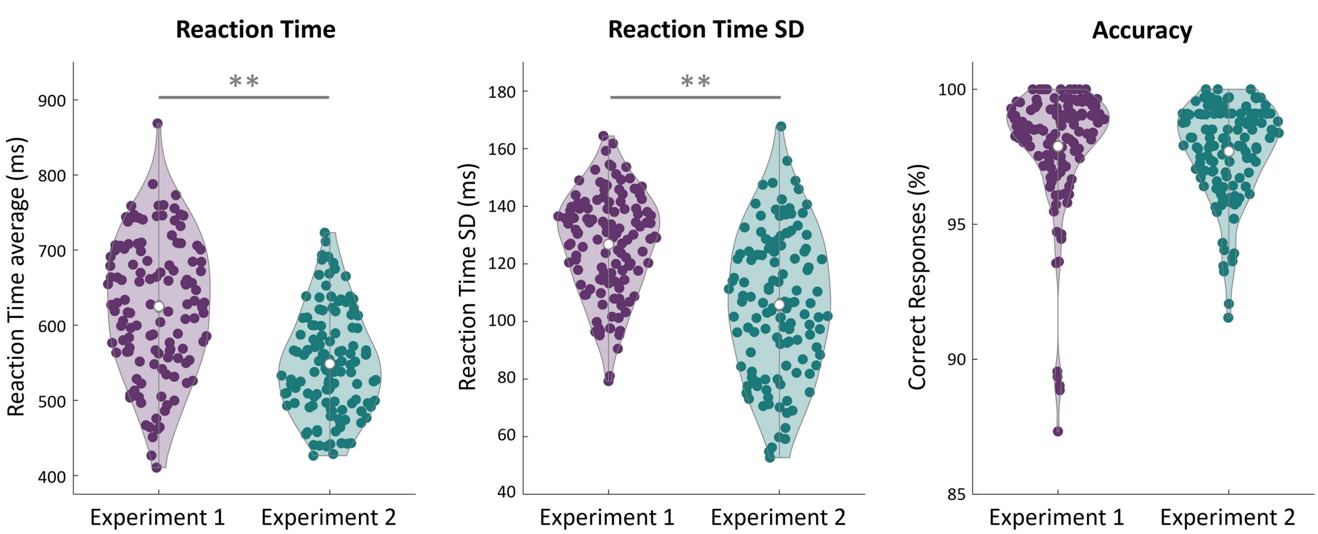
General Discussion

The present study is a preregistered attempt to replicate the findings reported in Ric and Muller (2012), in which arithmetic addition of two invisible digits was flexibility initiated by invisible instructions. Across two experiments, our findings conclusively point to the contrary. In contrast to the original report, masked instructions and digits did not affect participants' responses to a subsequent target. We therefore conclude that either performing arithmetic operations unconsciously, or following invisible instructions (or both) might not be possible, at least under visual masking.

We believe that this failure to replicate is genuine (Simons, 2014), as we went to extra length to provide the original result with the best chances to be reproduced (Button et al., 2013; Nosek et al., 2022): The two experiments reported here were based on the design and the data of the original study. The sample size used in both experiments was calculated to have a 95% probability of replicating the original finding. In addition, we have taken two further measures to increase the power of our design and the signal-to-noise ratio. First, we used three times the number of trials of the original study. Second, in Experiment 1, a calibration session was administered to find presentation settings that would maximize the signal of the masked stimuli while still keeping them below visibility threshold (Biderman & Mudrik, 2017; Koivisto & Grassini, 2018; Tal, Schechtman, et al., 2024). As we failed to find an effect despite these measures, we also conducted Experiment 2, as a more direct replication of the original study (Zwaan et al., 2018). Finally, the results of both experiments

Figure 6

The Influence of Trial-by-Trial Awareness Measurements on Performance



Note. RT was slower in Experiment 1, in which each trial included also a report of visibility (left panel). Slower speeds were accompanied by larger within-subject variability of speed (middle panel). Accuracy, however, was not different in the two experiments, hence did not seem to be affected by the online measurement of awareness (right panel). RT = response time. See the online article for the color version of this figure.

** $p < .001$.

were validated through additional analyses, using different trial and participant exclusion criteria (including examining only the same number of trials used in the original study, to make sure extending the experiment did not eliminate the effect), to verify that they are robust to the selection of specific data-cleaning parameters and to mitigate concerns of post hoc selection effects. Taken together, all these analyses further strengthen our conclusion that the original result might be a false positive one.

The current work, accordingly, joins other recent failures to replicate earlier works suggesting high-level processing without awareness (e.g., N. Biderman & Mudrik, 2018; Moors et al., 2016; Rabagliati et al., 2018). Thus, it supports a more skeptical approach about the scope of unconscious processes (Newell & Shanks, 2014), at least when these are tested using manipulations that render stimuli invisible. More specifically, it strengthens claims that consciousness might be needed for high-level integration of semantic information (Mudrik et al., 2014; Zher-Wen & Yu, 2023), especially over large windows of integration (Hirschhorn et al., 2021; Van Opstal & Rooyakkers, 2022). In the current design, three types of integration were required for an effect to be found: first, between the two digits, for the arithmetic operation to be performed. Second, between the instruc-
tions and the digits, requiring (a) mental chaining of operations; (b) executive functioning, as the instructions should have evoked flexibly adding/representing the digits; and (c) integration over time (as they were 180 ms apart). And third, between the prime digits and the target, which were separated by 1,200 ms. For all three, conflicting findings have been reported. Some argued for unconscious application of arithmetic operations (García-Orza et al., 2009; Sklar et al., 2012), while others suggested otherwise (Moors & Hesselmann, 2018, 2019; Rabagliati et al., 2018). Similarly, some found evidence for executive functions occurring without awareness (Hughes et al., 2009; Van Gaal et al., 2010; van Gaal & Lamme, 2012), while others showed that flexible behavior requires conscious processing (Ben-Haim et al., 2021; Cheesman & Merikle, 1986). Mental chaining of operations was also shown to require conscious processing (Sackur & Dehaene, 2009). Finally, integration over time also yielded conflicting results, with classical claims about the fast decline of the unconscious signal (Greenwald, Draine, & Abrams, 1996), and others arguing for long-lasting unconscious integration (Reber & Henke, 2012). For all three debates, our results align with those claiming for limited integration without consciousness, though more research is needed to determine which of the three types of integration failed in this study.

In any case, our findings accord with theories of consciousness like the GNW (Dehaene & Naccache, 2001), suggesting that conscious processing might be functionally distinct (see also Lamme, 2020; Lamme & Roelfsema, 2000). This seems to challenge other theories stating that consciousness is not functionally defined (e.g., integrated information theory, Albantakis et al., 2022; or higher-order thought theories, H. Lau & Rosenthal, 2011), though again, further research is needed to reach firm conclusions on this issue, given the wealth of contradicting findings on unconscious processing and its scope (for reviews, see Hassin, 2013; Kouider & Dehaene, 2007; Mudrik & Deouell, 2022; van Gaal & Lamme, 2012).

More generally, this work is part of an ongoing effort in psychological sciences to reexamine old findings using more refined experimental and statistical tools. This effort emphasizes the importance of replication in psychological sciences (Koole & Lakens, 2012; Nosek et al., 2012; Open Science Collaboration, 2015), as means to increase robustness and weed out false positive findings. Beyond the importance of replications, comparing the current work with the original one provides an interesting opportunity to reflect on the changes in norms instilled by the open science (Open Science Collaboration, 2015; Shrout & Rodgers, 2018) and good scientific practices (e.g., Crüwell et al., 2019; Cumming, 2014) movements in the last decade or so: determining sample sizes a-priori based on power analysis (Cohen, 2013; Erdfelder et al., 1996) or preregistering experimental methods (Nosek et al., 2019), including as registered reports (Chambers & Tzavella, 2022) that were quite rare in psychological research, have now become common practice (Nosek et al., 2022). Similarly, using Bayesian statistics to estimate the conclusiveness of the evidence, especially for null results, has become much more accepted (Kruschke, 2014; van de Schoot et al., 2021; Wagenmakers et al., 2018). This trend is highly relevant the field of unconscious processing, where null results are of special importance for determining that participants are not above chance in the objective measure of awareness (Shanks et al., 2021). Our study combines these approaches, thereby demonstrating how they can substantially affect experimental designs and outcomes.

Interestingly, the failure of higher-powered designs to replicate old findings goes against the once popular belief that effects found under small samples are even more reliable than those found under large ones, as significance is assumed to be more easily obtained when the sample size is larger. Albeit unintuitive (Kahneman & Tversky, 1972), small samples are in fact more likely to produce extreme results than large samples are, and so both positive and null effects found under small samples are less reliable (Button et al., 2013). Accordingly, a cornerstone of scientific replication is to conduct them under sufficient and well-planned statistical power (Button et al., 2013).

Irrespective of the failure to replicate the original result, the present study allowed as to compare the potential costs and benefits of online measures, as opposed to probing consciousness in a separate block (Avneon & Lamy, 2018; Fischer et al., 2011; Kiefer et al., 2023). Here, we had two highly powered experiments with a similar design, one of which included an online measure of visibility (Experiment 1) and another that did not (Experiment 2). Comparing participants' behavior in these two setups, we first found no effect of online measures on the main priming effect of interest. However, this comparison is not very meaningful given that priming was not found to begin with. More interestingly though, and in line with the past studies (Avneon & Lamy, 2018; Hesselmann et al., 2018; Kiefer et al., 2023), we found that responses in Experiment 1 were slower than in Experiment 2, possibly because of the additional resources required when performing multiple tasks within each trial. Going beyond previous studies, we observed that the decrease in RT under dual task settings was accompanied by an increase in RT variance, and that both effects diminished over time (i.e., RT became quicker and more consistent over the progression of the experiment). This suggests that the costs associated with online visibility judgments could be mitigated with practice, as is found in general dual task settings (Strobach & Torsten, 2017).

In sum, the current work demonstrated the irreproducibility of a previously reported unconscious processing effect, under more stringent experimental and statistical methods than originally used. In contrast to the original findings, we report no evidence that subliminal stimuli can lead to the unconscious initiation and carrying out of arithmetic operations. In addition to the scientific importance of this finding, we further shed light on the effect of probing consciousness on a trial-by-trial basis and show that they may be mitigated with training. Further research might help develop protocols that would allow researchers to enjoy the benefits of the trial-by-trial probing of awareness, while minimizing their costs with practice.

Constraints on Generality

This study was conducted in Israel and included participants who were either native English speakers or fluent in English. Therefore, it has probably contained mostly bilingual individuals. One could accordingly claim that the results might not necessarily generalize into the general population, most of which is monolingual. Nevertheless, we suggest that the results of this study do not depend on linguistic skills and would generalize to any language.

Context of the Research

A main line of research of our group focuses on the role of consciousness in integration of information. The relations between these two phenomena—consciousness and integration—have been widely studied and are a matter of ongoing debate, dating back to philosophical writings about consciousness and its nature (for review of theoretical claims and empirical findings, see Mudrik et al., 2014). We previously examined the scope and limits of unconscious integration with respect to object-scene integration (Biderman & Mudrik, 2017; Mudrik et al., 2011; Mudrik & Koch, 2013; Tal, Sar-Shalom, et al., 2024), multisensory integration (Faivre et al., 2014), and other types of integration as well (e.g., between categorical or lexical symbols; D. Biderman et al., 2020; for review, see Hirschhorn et al., 2021). The question is of pivotal importance to our understanding of consciousness, given that some leading theories in the field (e.g., the global neuronal workspace; Dehaene & Naccache, 2001) predict no high-level integration in the absence of awareness. The study by Ric and Muller (2012) substantially challenged such theories, because it strikingly demonstrates both temporal integration (involving mental chaining) and arithmetic operation (integrating two numbers) in the absence of awareness. The original findings are of further interest because they also entail unconscious task switching, claimed by some to be limited to conscious processing. Thus, and given previous replication failures in the field (including of our own work; see again Biderman & Mudrik, 2017), we thought it is both timely and essential to examine the reproducibility of this effect.

References

Albantakis, L., Barbosa, L., Findlay, G., Grasso, M., Haun, A. M., Marshall, W., Mayner, W. G. P., Zaeemzadeh, A., Boly, M., & Juel, B. E. (2022). *Integrated information theory (IIT) 4.0: Formulating the properties of phenomenal existence in physical terms*. arXiv. https://doi.org/10.48550/arXiv.2212.14787

Ansorge, U., Kunde, W., & Kiefer, M. (2014). Unconscious vision and executive control: How unconscious processing and conscious action control interact. *Consciousness and Cognition: An International Journal*, *27*, 268–287. https://doi.org/10.1016/j.concog.2014.05.009

Avneon, M., & Lamy, D. (2018). Reexamining unconscious response priming: A liminal-prime paradigm. *Consciousness and Cognition: An International Journal*, *59*, 87–103. https://doi.org/10.1016/j.concog.2017.12.006

Barbot, A., & Kouider, S. (2012). Longer is not better: Nonconscious overstimulation reverses priming influences under interocular suppression. *Attention, Perception, & Psychophysics*, *74*(1), 174–184. https://doi.org/10.3758/s13414-011-0226-3

Ben-Haim, M. S., Dal Monte, O., Fagan, N. A., Dunham, Y., Hassin, R. R., Chang, S. W. C., & Santos, L. R. (2021). Disentangling perceptual awareness from nonconscious processing in rhesus monkeys (*Macaca mulatta*). *PNAS Proceedings of the National Academy of Sciences of the United States of America*, *118*(15), Article e2017543118. https://doi.org/10.1073/pnas.2017543118

Biderman, D., Shir, Y., & Mudrik, L. (2020). B or 13? Unconscious top-down contextual effects at the categorical but not the lexical level. *Psychological Science*, *31*(6), 663–677. https://doi.org/10.1177/0956797620915887

Biderman, N., & Mudrik, L. (2018). Evidence for implicit—But not unconscious—Processing of object-scene relations. *Psychological Science*, *29*(2), 266–277. https://doi.org/10.1177/0956797617735745

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, *10*(4), 433–436. https://doi.org/10.1163/156856897X00357

Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., & Munafò, M. R. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, *14*(5), 365–376. https://doi.org/10.1038/nrn3475

Chambers, C. D., & Tzavella, L. (2022). The past, present and future of registered reports. *Nature Human Behaviour*, *6*(1), 29–42. https://doi.org/10.1038/s41562-021-01193-7

Cheesman, J., & Merikle, P. M. (1986). Distinguishing conscious from unconscious perceptual processes. *Canadian Journal of Psychology/Revue canadienne de psychologie*, *40*(4), 343–367. https://doi.org/10.1037/h0080103

Chong, T. T.-J., Husain, M., & Rosenthal, C. R. (2014). Recognizing the unconscious. *Current Biology*, *24*(21), R1033–R1035. https://doi.org/10.1016/j.cub.2014.09.035

Cohen, J. (2013). *Statistical power analysis for the behavioral sciences*. Routledge. https://doi.org/10.4324/9780203771587

Crüwell, S., Stefan, A. M., & Evans, N. J. (2019). Robust standards in cognitive science. *Computational Brain & Behavior*, *2*(3–4), 255–265. https://doi.org/10.1007/s42113-019-00049-8

Cumming, G. (2014). The new statistics: Why and how. *Psychological Science*, *25*(1), 7–29. https://doi.org/10.1177/0956797613504966

De Pisapia, N., Turatto, M., Lin, P., Jovicich, J., & Caramazza, A. (2012). Unconscious priming instructions modulate activity in default and executive networks of the human brain. *Cerebral Cortex*, *22*(3), 639–649. https://doi.org/10.1093/cercor/bhr146

Dehaene, S., & Changeux, J. P. (2011). Experimental and theoretical approaches to conscious processing. *Neuron*, *70*(2), 200–227. https://doi.org/10.1016/j.neuron.2011.03.018

Dehaene, S., Lau, H., & Kouider, S. (2017). What is consciousness, and could machines have it? *Science*, *358*(6362), 486–492. https://doi.org/10.1126/science.aan8871

Dehaene, S., & Naccache, L. (2001). Towards a cognitive neuroscience of consciousness: Basic evidence and a workspace framework. *Cognition*, *79*(1–2), 1–37. https://doi.org/10.1016/S0010-0277(00)00123-2

Deschuyteneer, M., & Vandierendonck, A. (2005). Are "input monitoring" and "response selection" involved in solving simple mental arithmetical sums? *European Journal of Cognitive Psychology*, *17*(3), 347–370. https://doi.org/10.1080/09541440440000032

Draine, S. C., & Greenwald, A. G. (1998). Replicable unconscious semantic priming. *Journal of Experimental Psychology: General*, *127*(3), 286–303. https://doi.org/10.1037/0096-3445.127.3.286

Eimer, M., & Schlaghecken, F. (2003). Response facilitation and inhibition in subliminal priming. *Biological Psychology*, *64*(1–2), 7–26. https://doi.org/10.1016/S0301-0511(03)00100-5

Erdfelder, E., Faul, F., & Buchner, A. (1996). GPOWER: A general power analysis program. *Behavior Research Methods, Instruments & Computers*, *28*(1), 1–11. https://doi.org/10.3758/BF03203630

Eriksen, C. W. (1960). Discrimination and learning without awareness: A methodological survey and evaluation. *Psychological Review*, *67*(5), 279–300. https://doi.org/10.1037/h0041622

Fahrenfort, J. J., van Leeuwen, J., Olivers, C. N. L., & Hogendoorn, H. (2017). Perceptual integration without conscious access. *PNAS Proceedings of the National Academy of Sciences of the United States of America*, *114*(14), 3744–3749. https://doi.org/10.1073/pnas.1617268114

Faivre, N., & Koch, C. (2014). Inferring the direction of implied motion depends on visual awareness. *Journal of Vision*, *14*(4), Article 4. https://doi.org/10.1167/14.4.4

Faivre, N., Mudrik, L., Schwartz, N., & Koch, C. (2014). Multisensory integration in complete unawareness: Evidence from audiovisual congruency priming. *Psychological Science*, *25*(11), 2006–2016. https://doi.org/10.1177/0956797614547916

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*(4), 1149–1160. https://doi.org/10.3758/BRM.41.4.1149

Finkbeiner, M. (2011). Subliminal priming with nearly perfect performance in the prime-classification task. *Attention, Perception, & Psychophysics*, *73*(4), 1255–1265. https://doi.org/10.3758/s13414-011-0088-8

Fischer, R., Kiesel, A., Kunde, W., & Schubert, T. (2011). Selective impairment of masked priming in dual-task performance. *The Quarterly Journal of Experimental Psychology*, *64*(3), 572–595. https://doi.org/10.1080/17470218.2010.505984

García-Orza, J., Damas-López, J., Matas, A., & Rod ríguez, J. M. (2009). "2 x 3" primes naming "6": Evidence from masked priming. *Attention, Perception, & Psychophysics*, *71*(3), 471–480. https://doi.org/10.3758/APP.71.3.471

Geary, D. C., & Wiley, J. G. (1991). Cognitive addition: Strategy choice and speed-of-processing differences in young and elderly adults. *Psychology and Aging*, *6*(3), 474–483. https://doi.org/10.1037/0882-7974.6.3.474

Greenwald, A. G., & De Houwer, J. (2017). Unconscious conditioning: Demonstration of existence and difference from conscious conditioning. *Journal of Experimental Psychology: General*, *146*(12), 1705–1721. https://doi.org/10.1037/xge0000371

Greenwald, A. G., Draine, S. C., & Abrams, R. L. (1996). Three cognitive markers of unconscious semantic activation. *Science*, *273*(5282), 1699–1702. https://doi.org/10.1126/science.273.5282.1699

Hassin, R. R. (2013). Yes it can: On the functional abilities of the human unconscious. *Perspectives on Psychological Science*, *8*(2), 195–207. https://doi.org/10.1177/1745691612460684

Hesselmann, G., Darcy, N., Ludwig, K., & Sterzer, P. (2016). Priming in a shape task but not in a category task under continuous flash suppression. *Journal of Vision*, *16*(3), Article 17. https://doi.org/10.1167/16.3.17

Hesselmann, G., Darcy, N., Rothkirch, M., & Sterzer, P. (2018). Investigating masked priming along the "vision-for-perception" and "vision-for-action" dimensions of unconscious processing. *Journal of Experimental Psychology: General*, *147*(11), 1641–1659. https://doi.org/10.1037/xge0000420

Heyman, T., & Moors, P. (2014). Frequent words do not break continuous flash suppression differently from infrequent or nonexistent words: Implications for semantic processing of words in the absence of awareness. *PLOS ONE*, *9*(8), Article e104719. https://doi.org/10.1371/journal.pone.0104719

Hirschhorn, R., Kahane, O., Gur-Arie, I., Faivre, N., & Mudrik, L. (2021). Windows of integration hypothesis revisited. *Frontiers in Human Neuroscience*, *14*, Article 617187. https://doi.org/10.3389/fnhum.2020.617187

Hughes, G., Velmans, M., & De Fockert, J. (2009). Unconscious priming of a no-go response. *Psychophysiology*, *46*(6), 1258–1269. https://doi.org/10.1111/j.1469-8986.2009.00873.x

Kahneman, D., & Tversky, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology*, *3*(3), 430–454. https://doi.org/10.1016/0010-0285(72)90016-3

Kang, M. S., Blake, R., & Woodman, G. F. (2011). Semantic analysis does not occur in the absence of awareness induced by interocular suppression. *The Journal of Neuroscience*, *31*(38), 13535–13545. https://doi.org/10.1523/JNEUROSCI.1691-11.2011

Karpinski, A., Briggs, J. C., & Yale, M. (2019). A direct replication: Unconscious arithmetic processing. *European Journal of Social Psychology*, *49*(3), 637–644. https://doi.org/10.1002/ejsp.2390

Kiefer, M., Harpaintner, M., Rohr, M., & Wentura, D. (2023). Assessing subjective prime awareness on a trial-by-trial basis interferes with masked semantic priming effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *49*(2), 269–283. https://doi.org/10.1037/xlm0001228

Koivisto, M., & Grassini, S. (2018). Unconscious response priming during continuous flash suppression. *PLOS ONE*, *13*(2), Article e0192201. https://doi.org/10.1371/journal.pone.0192201

Koole, S. L., & Lakens, D. (2012). Rewarding replications: A sure and simple way to improve psychological science. *Perspectives on Psychological Science*, *7*(6), 608–614. https://doi.org/10.1177/1745691612462586

Kouider, S., & Dehaene, S. (2007). Levels of processing during non-conscious perception: A critical review of visual masking. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *362*(1481), 857–875. https://doi.org/10.1098/rstb.2007.2093

Kouider, S., & Dupoux, E. (2004). Partial awareness creates the "illusion" of subliminal semantic priming. *Psychological Science*, *15*(2), 75–81. https://doi.org/10.1111/j.0963-7214.2004.01502001.x

Kruschke, J. (2014). *Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan*. Academic Press.

Lamme, V. A. F. (2020). Visual functions generating conscious seeing. *Frontiers in Psychology*, *11*, Article 83. https://doi.org/10.3389/fpsyg.2020.00083

Lamme, V. A. F., & Roelfsema, P. R. (2000). The distinct modes of vision offered by feedforward and recurrent processing. *Trends in Neurosciences*, *23*(11), 571–579. https://doi.org/10.1016/S0166-2236(00)01657-X

Lamy, D., Carmel, T., & Peremen, Z. (2017). Prior conscious experience enhances conscious perception but does not affect response priming☆. *Cognition*, *160*, 62–81. https://doi.org/10.1016/j.cognition.2016.12.009

Lau, H., & Rosenthal, D. (2011). Empirical support for higher-order theories of conscious awareness. *Trends in Cognitive Sciences*, *15*(8), 365–373. https://doi.org/10.1016/j.tics.2011.05.009

Lau, H. C., & Passingham, R. E. (2006). Relative blindsight in normal observers and the neural correlate of visual consciousness. *PNAS Proceedings of the National Academy of Sciences of the United States of America*, *103*(49), 18763–18768. https://doi.org/10.1073/pnas.0607716103

Lau, H. C., & Passingham, R. E. (2007). Unconscious activation of the cognitive control system in the human prefrontal cortex. *The Journal of Neuroscience*, *27*(21), 5805–5811. https://doi.org/10.1523/JNEUROSCI.4335-06.2007

Lee, M. D., & Wagenmakers, E. J. (2013). *Bayesian data analysis for cognitive science: A practical course*. Cambridge University Press.

Levitt, H. (1971). Transformed up-down methods in psychoacoustics. *Journal of the Acoustical Society of America*, *49*, 467–477. https://doi.org/10.1121/1.1912375

Lin, Z., & Murray, S. O. (2014). Priming of awareness or how not to measure visual awareness. *Journal of Vision*, *14*(1), Article 27. https://jov.arvojournals.org/article.aspx?articleid=2295565

Mattler, U. (2003). Priming of mental operations by masked stimuli. *Perception & Psychophysics*, *65*(2), 167–187. https://doi.org/10.3758/BF03194793

Mattler, U. (2007). Inverse target- and cue-priming effects of masked stimuli. *Journal of Experimental Psychology: Human Perception and Performance*, *33*(1), 83–102. https://doi.org/10.1037/0096-1523.33.1.83

Meeres, S. L., & Graves, R. E. (1990). Localization of unseen visual stimuli by humans with normal vision. *Neuropsychologia*, *28*(12), 1231–1237. https://doi.org/10.1016/0028-3932(90)90039-Q

Merikle, P. M., & Reingold, E. M. (1998). On demonstrating unconscious perception: Comment on Draine and Greenwald (1998). *Journal of Experimental Psychology: General*, *127*(3), 304–310. https://doi.org/10.1037/0096-3445.127.3.304

Moors, P., Boelens, D., van Overwalle, J., & Wagemans, J. (2016). Scene integration without awareness: No conclusive evidence for processing scene congruency during continuous flash suppression. *Psychological Science*, *27*(7), 945–956. https://doi.org/10.1177/0956797616642525

Moors, P., & Hesselmann, G. (2018). A critical reexamination of doing arithmetic nonconsciously. *Psychonomic Bulletin & Review*, *25*(1), 472–481. https://doi.org/10.3758/s13423-017-1292-x

Moors, P., & Hesselmann, G. (2019). Unconscious arithmetic: Assessing the robustness of the results reported by Karpinski, Briggs, and Yale (2018). *Consciousness and Cognition: An International Journal*, *68*, 97–106. https://doi.org/10.1016/j.concog.2019.01.003

Moors, P., Hesselmann, G., Wagemans, J., & van Ee, R. (2017). Continuous flash suppression: Stimulus fractionation rather than integration. *Trends in Cognitive Sciences*, *21*(10), 719–721. https://doi.org/10.1016/j.tics.2017.06.005

Morey, R. D., & Rouder, J. N. (2015). *BayesFactor: Computation of Bayes factors for common designs* (R package Version 0.9.10-2) [Computer software].

Mudrik, L., Breska, A., Lamy, D., & Deouell, L. Y. (2011). Integration without awareness: Expanding the limits of unconscious processing. *Psychological Science*, *22*(6), 764–770. https://doi.org/10.1177/0956797611408736

Mudrik, L., & Deouell, L. Y. (2022). Neuroscientific evidence for processing without awareness. *Annual Review of Neuroscience*, *45*(1), 403–423. https://doi.org/10.1146/annurev-neuro-110920-033151

Mudrik, L., Faivre, N., & Koch, C. (2014). Information integration without awareness. *Trends in Cognitive Sciences*, *18*(9), 488–496. https://doi.org/10.1016/j.tics.2014.04.009

Mudrik, L., & Koch, C. (2013). Differential processing of invisible congruent and incongruent scenes: A case for unconscious integration. *Journal of Vision*, *13*(13), Article 24. https://doi.org/10.1167/13.13.24

Newell, B. R., & Shanks, D. R. (2014). Unconscious influences on decision making: A critical review. *Behavioral and Brain Sciences*, *37*(1), 1–19. https://doi.org/10.1017/S0140525X12003214

Nosek, B. A., Beck, E. D., Campbell, L., Flake, J. K., Hardwicke, T. E., Mellor, D. T., van't Veer, A. E., & Vazire, S. (2019). Preregistration is hard, and worthwhile. *Trends in Cognitive Sciences*, *23*(10), 815–818. https://doi.org/10.1016/j.tics.2019.07.009

Nosek, B. A., Hardwicke, T. E., Moshontz, H., Allard, A., Corker, K. S., Dreber, A., Fidler, F., Hilgard, J., Kline Struhl, M., Nuijten, M. B., Rohrer, J. M., Romero, F., Scheel, A. M., Scherer, L. D., Schönbrodt, F. D., & Vazire, S. (2022). Replicability, robustness, and reproducibility in psychological science. *Annual Review of Psychology*, *73*(1), 719–748. https://doi.org/10.1146/annurev-psych-020821-114157

Nosek, B. A., Spies, J. R., & Motyl, M. (2012). Scientific utopia: II. Restructuring incentives and practices to promote truth over publishability. *Perspectives on Psychological Science*, *7*(6), 615–631. https://doi.org/10.1177/1745691612459058

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251), Article aac4716. https://doi.org/10.1126/science.aac4716

Peirce, C. S., & Jastrow, J. (1884). On small differences in sensation. *National Academy of Sciences*, *3*, 75–83.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*(4), 437–442. https://doi.org/10.1163/156856897X00366

Peremen, Z., & Lamy, D. (2014). Do conscious perception and unconscious processing rely on independent mechanisms? A meta-contrast study. *Consciousness and Cognition: An International Journal*, *24*, 22–32. https://doi.org/10.1016/j.concog.2013.12.006

Peters, M. A. K., Kentridge, R. W., Phillips, I., & Block, N. (2017). Does unconscious perception really exist? Continuing the ASSC20 debate. *Neuroscience of Consciousness*, *2017*(1), Article nix015. https://doi.org/10.1093/nc/nix015

Pratte, M. S., & Rouder, J. N. (2009). A task-difficulty artifact in subliminal priming. *Attention, Perception, & Psychophysics*, *71*(6), 1276–1283. https://doi.org/10.3758/APP.71.6.1276

Prins, N., & Kingdom, F. A. A. (2009). *Palamedes: Matlab routines for analyzing psychophysical data* [Computer software]. https://www.palamedestoolbox.org

R Core Team. (2015). *R: A language and environment for statistical computing* [Computer software]. https://www.R-project.org

Rabagliati, H., Robertson, A., & Carmel, D. (2018). The importance of awareness for understanding language. *Journal of Experimental Psychology: General*, *147*(2), 190–208. https://doi.org/10.1037/xge0000348

Ramsøy, T. Z., & Overgaard, M. (2004). Introspection and subliminal perception. *Phenomenology and the Cognitive Sciences*, *3*(1), 1–23. https://doi.org/10.1023/B:PHEN.0000041900.30172.e8

Ratcliff, R. (1993). Methods for dealing with reaction time outliers. *Psychological Bulletin*, *114*(3), 510–532. https://doi.org/10.1037/0033-2909.114.3.510

Reber, T. P., & Henke, K. (2012). Integrating unseen events over time. *Consciousness and Cognition: An International Journal*, *21*(2), 953–960. https://doi.org/10.1016/j.concog.2012.02.013

Reingold, E. M., & Merikle, P. M. (1988). Using direct and indirect measures to study perception without awareness. *Perception & Psychophysics*, *44*(6), 563–575. https://doi.org/10.3758/BF03207490

Reuss, H., Kiesel, A., Kunde, W., & Wühr, P. (2012). A cue from the unconscious—Masked symbols prompt spatial anticipation. *Frontiers in Psychology*, *3*, Article 397. https://doi.org/10.3389/fpsyg.2012.00397

Ric, F., & Muller, D. (2012). Unconscious addition: When we unconsciously initiate and follow arithmetic rules. *Journal of Experimental Psychology: General*, *141*(2), 222–226. https://doi.org/10.1037/a0024608

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian *t* tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review*, *16*(2), 225–237. https://doi.org/10.3758/PBR.16.2.225

Rougier, M., Muller, D., Ric, F., Alexopoulos, T., Batailler, C., Smeding, A., & Aube, B. (2018). A new look at sensorimotor aspects in approach/avoidance tendencies: The role of visual whole-body movement information. *Journal of Experimental Social Psychology*, *76*, 42–53. https://doi.org/10.1016/j.jesp.2017.12.004

Roussel, J. L., Fayol, M., & Barrouillet, P. (2002). Procedural vs. direct retrieval strategies in arithmetic: A comparison between additive and multiplicative problem solving. *European Journal of Cognitive Psychology*, *14*(1), 61–104. https://doi.org/10.1080/09541440042000115

Sackur, J., & Dehaene, S. (2009). The cognitive architecture for chaining of two mental operations. *Cognition*, *111*(2), 187–211. https://doi.org/10.1016/j.cognition.2009.01.010

Sand, A., & Nilsson, M. E. (2016). Subliminal or not? Comparing null-hypothesis and Bayesian methods for testing subliminal priming. *Consciousness and Cognition: An International Journal*, *44*, 29–40. https://doi.org/10.1016/j.concog.2016.06.012

Schmidt, T., & Vorberg, D. (2006). Criteria for unconscious cognition: Three types of dissociation. *Perception & Psychophysics*, *68*(3), 489–504. https://doi.org/10.3758/BF03193692

Seth, A. K., Dienes, Z., Cleeremans, A., Overgaard, M., & Pessoa, L. (2008). Measuring consciousness: Relating behavioural and neurophysiological approaches. *Trends in Cognitive Sciences*, *12*(8), 314–321. https://doi.org/10.1016/j.tics.2008.04.008

Shanks, D. R. (2017). Regressive research: The pitfalls of post hoc data selection in the study of unconscious mental processes. *Psychonomic Bulletin & Review*, *24*(3), 752–775. https://doi.org/10.3758/s13423-016-1170-y

Shanks, D. R., Malejka, S., & Vadillo, M. A. (2021). The challenge of inferring unconscious mental processes. *Experimental Psychology*, *68*(3), 113–129. https://doi.org/10.1027/1618-3169/a000517

Shrout, P. E., & Rodgers, J. L. (2018). Psychology, science, and knowledge construction: Broadening perspectives from the replication crisis. *Annual Review of Psychology*, *69*(1), 487–510. https://doi.org/10.1146/annurev-psych-122216-011845

Simons, D. J. (2014). The value of direct replication. *Perspectives on Psychological Science*, *9*(1), 76–80. https://doi.org/10.1177/1745691613514755

Sklar, A. Y., Levy, N., Goldstein, A., Mandel, R., Maril, A., & Hassin, R. R. (2012). Reading and doing arithmetic nonconsciously. *PNAS Proceedings of the National Academy of Sciences of the United States of America*, *109*(48), 19614–19619. https://doi.org/10.1073/pnas.1211645109

Snodgrass, M. (2004). The dissociation paradigm and its discontents: How can unconscious perception or memory be inferred? *Consciousness and Cognition: An International Journal*, *13*(1), 107–116. https://doi.org/10.1016/j.concog.2003.11.001

Sohn, M.-H., & Carlson, R. A. (1998). Procedural framework for simple arithmetic skills. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*(4), 1052–1067. https://doi.org/10.1037/0278-7393.24.4.1052

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments & Computers*, *31*(1), 137–149. https://doi.org/10.3758/BF03207704

Stein, T., van Gaal, S., & Fahrenfort, J. J. (2024). How (not) to demonstrate unconscious priming: Overcoming issues with post-hoc data selection, low power, and frequentist statistics. *Consciousness and Cognition: An International Journal*, *119*, Article 103669. https://doi.org/10.1016/j.concog.2024.103669

Strobach, T., & Torsten, S. (2017). Mechanisms of practice-related reductions of dual-task interference with simple tasks: Data and theory. *Advances in Cognitive Psychology*, *13*(1), 28–41. https://doi.org/10.5709/acp-0204-7

Tal, A., & Mudrik, L. (2024). *Unconscious initiation and following of arithmetic rules? A replication study*. https://osf.io/btm7v/

Tal, A., Sar-Shalom, M., Krawitz, T., Biderman, D., & Mudrik, L. (2024). Awareness is needed for contextual effects in ambiguous object recognition. *Cortex: A Journal Devoted to the Study of the Nervous System and Behavior*, *173*, 49–60. https://doi.org/10.1016/j.cortex.2024.01.003

Tal, A., Schechtman, E., Caughran, B., Paller, K. A., & Davachi, L. (2024). The reach of reactivation: Effects of consciously triggered versus unconsciously triggered reactivation of associative memory. *PNAS Proceedings of the National Academy of Sciences of the United States of America*, *121*(10), Article e2313604121. https://doi.org/10.1073/pnas.2313604121

van de Schoot, R., Depaoli, S., King, R., Kramer, B., Märtens, K., Tadesse, M. G., Vannucci, M., Gelman, A., Veen, D., Willemsen, J., & Yau, C. (2021). Bayesian statistics and modelling. *Nature Reviews Methods Primers*, *1*(1), Article 1. https://doi.org/10.1038/s43586-020-00001-2

van Gaal, S., & Lamme, V. A. F. (2012). Unconscious high-level information processing: Implication for neurobiological theories of consciousness. *The Neuroscientist*, *18*(3), 287–301. https://doi.org/10.1177/1073858411404079

van Gaal, S., Naccache, L., Meuwese, J. D., van Loon, A. M., Leighton, A. H., Cohen, L., & Dehaene, S. (2014). Can the meaning of multiple words be integrated unconsciously? *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *369*(1641), Article 20130212. https://doi.org/10.1098/rstb.2013.0212

van Gaal, S., Ridderinkhof, K. R., Scholte, H. S., & Lamme, V. A. F. (2010). Unconscious activation of the prefrontal no-go network. *The Journal of Neuroscience*, *30*(11), 4143–4150. https://doi.org/10.1523/JNEUROSCI.2992-09.2010

Van Opstal, F., & Rooyakkers, M. (2022). Unconscious information integration: A replication and the role of spatial window in masking experiments. *Cognition*, *225*, Article 105113. https://doi.org/10.1016/j.cognition.2022.105113

Wagenmakers, E. J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., Love, J., Selker, R., Gronau, Q. F., Šmíra, M., Epskamp, S., Matzke, D., Rouder, J. N., & Morey, R. D. (2018). Bayesian inference for psychology. Part I: Theoretical advantages and practical ramifications. *Psychonomic Bulletin & Review*, *25*(1), 35–57. https://doi.org/10.3758/s13423-017-1343-3

Yang, Y.-H., Tien, Y.-H., Yang, P.-L., & Yeh, S.-L. (2017). Role of consciousness in temporal integration of semantic information. *Cognitive, Affective & Behavioral Neuroscience*, *17*(5), 954–972. https://doi.org/10.3758/s13415-017-0525-9

Zher-Wen, & Yu, R. (2023). Unconscious integration: Current evidence for integrative processing under subliminal conditions. *British Journal of Psychology*, *114*(2), 430–456. https://doi.org/10.1111/bjop.12631

Zhou, F. A., & Davis, G. (2012). Unconscious priming of task sets: The role of spatial attention. *Attention, Perception, & Psychophysics*, *74*(1), 105–114. https://doi.org/10.3758/s13414-011-0221-8

Zwaan, R. A., Etz, A., Lucas, R. E., & Donnellan, M. B. (2018). Making replication mainstream | behavioral and brain sciences | Cambridge core. *Behavioral and Brain Sciences*, *41*, Article e120. https://doi.org/10.1017/S0140525X17001972

Received January 14, 2018
Revision received June 7, 2024
Accepted June 17, 2024 ■