

Awareness is Predictably Lossy

Ariel Berlinger

University of Haifa <https://orcid.org/0009-0005-6372-5505>

Ariel Goldstein

Princeton University <https://orcid.org/0000-0003-4017-3363>

Baruch Eitam

`beitam@psy.haifa.ac.il`

University of Haifa <https://orcid.org/0000-0002-2202-3073>

Article

Keywords:

Posted Date: August 18th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10591267/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Additional Declarations: There is **NO** Competing Interest.

Abstract

When something is plainly in view, are we always aware of seeing it and is awareness always beneficial? We tested whether a clearly visible but task-irrelevant feature can fail to become reportable while still guiding behavior. Across four experiments ($N = 584$), participants viewed meaningful two-part objects under prolonged, unmasked exposure, with only one-part task-relevant. When unexpectedly asked to report the irrelevant part, participants performed at chance; yet the same feature reliably guided choices in a perceptual comparison task (55% and 60% correct; chance = 33%). The advantage belonged to the least aware: participants who reported noticing no change outperformed those who did, an inverse relationship confirmed in a large preregistered replication. Awareness is not simply stronger perception, but a selective transformation of what we see.

Main text

The visual world feels rich and vivid, yet decades of research show that observers routinely fail to report large and clearly visible changes in their environment. For example, change blindness (1, 2) and inattentional blindness (3–5) establish the gap between what was visible and what is reportable. Yet, this gap between the felt richness of perceptual experience and the poverty of what can be explicitly reported has proven surprisingly difficult to interpret. On the one hand, sparse accounts hold that there is no gap: conscious experience is itself severely limited, and only information that passes through processing bottlenecks is ever experienced (6, 7). On the other hand, rich accounts interpret the gap as evidence that conscious experience is richer than verbal reports suggest. On this view, failures of report reflect downstream limitations on access and memory rather than the perception itself (8–10). After much debate, neither position has been decisively ruled out, since neither camp could provide empirical evidence that would severely challenge the plausibility of the alternative position.

A simple manipulation suggests a way out of this impasse. When a stimulus is task-irrelevant, its clearly visible features can become effectively unreportable, not because they were unseen, but because relevance gates what reaches explicit report (11–13). Similarly, observers can fail to report an attribute they have just used to guide their own behavior, but is now irrelevant (14). Together, these findings suggest that the relationship between experience and report, rather than being constant, is highly dependent on the (task) relevance of the information, even when visibility, ‘cognitive load,’ and other factors are held constant. In other words, manipulating task relevance provides a way to experimentally dissociate visual perceivability from reportability, in turn creating conditions in which otherwise seen stimuli fail to reach reportability.

One framework that is useful for formulating this distinction is the Multi-Representational Media (MRM) framework (15). MRM proposes that what is ordinarily called consciousness, in fact, comprises two layers: an early high-bandwidth conscious-perceptual representation, and a later awareness that arises when perceptual content is enriched by the activation of relevance-weighted (semantic, propositional, or otherwise abstract) knowledge. On this account, relevance determines what enters awareness, not what is perceptually experienced.

For example, when looking at a cat on a couch, there is first the ‘raw’ high-bandwidth conscious-perceptual representation of the perceivable aspects of the scene, unfiltered and relatively invariant across individuals. In adult humans, this conscious-perceptual representation is rapidly enriched and reduced to the distinct awareness of the cat and couch and a more limited number of elements of their features, which will tend to vary among individuals and situations due to the differences in relevance of the accessible knowledge (i.e., due to the differences in what and how aspects of the experienced sensation are enriched).

Overview

The present experiments use this distinction as a working framework and test whether task relevance can dissociate perceptual guidance from explicit report. Across four experiments, we use meaningful and familiar stimuli, prolonged unmasked presentation, minimal delays between presentation and probing, and a perceptual comparison task to rule out the effects of surprise, task difficulty, and pre-allocation of spatial attention as alternative explanations.

We designed two pairs of experiments. In Experiments 1A and 1B, participants viewed a meaningful visual stimulus, a flower in a pot, while one of the two parts (flower or pot) was made task-relevant through repeated probing (e.g., participants were repeatedly asked to explicitly report either the flower's or pot's color), whereas the other always remained task-irrelevant. A critical surprise probe explicitly asking the participant about the irrelevant part's color allowed us to measure its reportability.

Having established that irrelevance alone is sufficient to render clearly visible features unreportable (as well as ruling out other plausible factors), Experiments 2A and 2B replaced explicit report with a non-intrusive perceptual comparison task. The subtle probing now enables us to ask not whether observers can explicitly report the irrelevant object's color, but whether information about that feature can nonetheless influence perceptual identification.

If task-irrelevant features are not simply lost, then performance on the irrelevant feature should remain above chance despite poor explicit report. A further possibility, suggested by layered accounts such as MRM, is that explicit awareness of the critical change may not always improve performance. Unintuitively, if awareness involves selective, relevance-weighted reconstruction of perceptual content, then awareness that the task has changed could, in some cases, index *less* (irrelevant) perceptual information. Under this interpretation, participants who are least aware (i.e., those who are unaware of the subtle switch in the probe) should show equal or even better performance than those who are aware (i.e., least relevance-based selection).

Because this inverse relationship between performance and awareness is both counterintuitive and difficult to reconcile with any unitary (i.e. consciousness and awareness are one) account of consciousness or stipulated relations between attention and consciousness, Experiment 2B was conducted as a preregistered replication of Experiment 2A and was run until a stringent level of evidence was met.

Materials and Methods

Participants

For Experiments 1A and 1B, we recruited participants via Prolific Academic. Participants were required to be at least 18 years old, to have normal or corrected-to-normal vision, not be color blind, and speak either Arabic, English, or Hebrew (all but one were English speakers). Experiment 1A initially recruited 60 participants, and Experiment 1B initially recruited 50 participants. In both experiments, additional participants were recruited in batches of 10 until the preregistered Bayesian stopping criterion was reached. Participants were compensated at Prolific's standard hourly rate.

Exclusion criteria: requiring more than 10 trials to achieve four consecutive correct responses during training, failing the attention check (see below), or failing to complete the experiment. After exclusions, the final samples

were N = 60 (no exclusions) for Experiment 1A and N = 47 (3 exclusions) for Experiment 1B.

Experiments 2A and 2B recruited English-speaking participants via Prolific Academic and had the same eligibility criteria. Experiment 2B was a preregistered direct replication of Experiment 2A. Data collection began with 70 participants and continued in batches of 20 until Bayesian preregistered stopping criteria were reached. Participants were compensated with an additional bonus of £0.50 for achieving greater than 90% accuracy while meeting the 2-second response window.

Exclusion criteria were: failing the attention check, achieving less than 75% accuracy on the task-relevant feature, or failing to complete the experiment. After exclusions, the final samples were N = 69 (1 exclusion, for completing the experiment twice) for Experiment 2A and N = 408 (22 exclusions: 6 for failing the attention check, 7 with success on the main task being under 75%, and 9 who completed the experiment twice) for Experiment 2B.

Stimuli and Apparatus

Stimuli consisted of high-resolution images of a flower in a pot, with both object parts clearly distinguishable and differently colored. Pot colors were orange RGB(255, 128, 0), purple RGB(237, 56, 237), green RGB(0, 178, 115), or navy RGB(0, 0, 209); flower colors were teal RGB(0, 191, 208), red RGB(255, 0, 0), yellow RGB(255, 230, 0), or green RGB(0, 208, 115). No image displayed identical colors for the pot and flower (see Supplementary Fig. S4 for all stimuli presented).

In Experiments 1A and 2A, images were presented centrally for 1000 ms without masking. Images measured 212 × 300 pixels, with the flower and pot together occupying approximately 28.7% of the image area; the remaining area consisted of a border matching the background color.

In Experiment 1B, the same stimuli were presented at one of eight randomized positions arranged in a circle (0°, 45°, 90°, 135°, 180°, 225°, 270°, 315°), randomly rotating around the left-center axis of the image rather than the screen center. This manipulation introduced spatial uncertainty, requiring participants to actively locate the relevant object on each trial and preventing advance allocation of spatial attention to a fixed spatial region. All experiments were conducted online using PsychoPy (Pavlovía) (28) and through Prolific (29).

Design and Procedure

All experiments employed a three-phase design: training, trials probing the relevant object-portion's color (with a single surprise task trial followed by a probe of the relevant portion's color), and a surprise-test of the irrelevant portion's color.

Experiments 1A and 1B

Training phase

Participants were pseudo-randomly assigned to report the color of either the flower or the pot throughout the experiment. Training continued until participants produced four consecutive correct responses or completed a maximum of 10 trials. Participants who failed to meet this criterion were excluded. Incorrect responses were followed by corrective feedback focusing participants on the relevant object's feature. Training trials were not included in the total trial count.

Trial phase

Participants completed 42 trials in which they selected the color of the task-relevant object from four response options. At predefined intervals (trials 1, 4, 7, 13, 21, 24, 27, 34, and 38), participants also indicated the basis of their response by selecting one of four options: memory, intuition, guess, or other.

Attention check and surprise control

On trial 21, following stimulus presentation but before probing the relevant part's color, participants were surprisingly asked to click (with the mouse) on the number 4 from a set of options (0, 2, 4, 7) before continuing. Participants who failed this check were excluded. This trial also served as the surprise-control condition: performance on the subsequent relevant object-part's color report provided a measure of the effect of an unexpected event on reporting independently of irrelevance. In the present (and previous) design(s), manipulated irrelevance and surprise are inherently confounded at the time of probing. This precludes a straightforward interaction model separating the two factors. We therefore quantified the effect of surprise on reports of the relevant color and used it as a comparison standard.

Surprise-test phase

After trial 42, participants were unexpectedly probed on the color of the previously irrelevant object-part. This was immediately followed by probing the relevant object-part color. This sequence (stimuli > asking about the previously irrelevant object part > probing for both object parts again across four trials (trials 43–46), during which both were now task-relevant, allowing assessment of performance level when participants are explicitly attempting to report both.

Experiments 2A and 2B

The overall structure of these two experiments was identical to Experiments 1A and 1B, with two key modifications. First, instead of a question asking them to report the color they have just seen (of either the relevant or irrelevant object-part), participants completed a three-alternative forced-choice (3AFC) perceptual comparison task. Following each stimulus presentation, three alternative images were displayed differing only in the color of the probed object-part (i.e., the color of the other part-object was correct but uninformative for discriminating between the options), and participants selected the image they had just seen. In most trials, participants were required to respond within a 2-second window; trials exceeding this limit were terminated. On surprise trials, the limit was extended to 10 seconds. Training trials were included in the total trial count.

The 2-second response deadline was introduced following a pilot study (Experiment 2, Pilot; Supplementary Materials) in which performance on the critical trial was negatively correlated with task completion time ($\rho = -0.295$, $p = 0.003$), suggesting that without time pressure, some participants fail to rely on the residual perceptual representation of the stimulus (Sperling, 1960). Note that imposing the same time limit on the explicit probes used in Experiments 1A and 1B does not change the pattern of results (Experiment 2 Method Control; Supplementary Materials).

The critical surprise trial also differed. In this trial, and without any notice of changing the task's requirement, the task-relevant object part's color was now held constant across all three alternatives (and correct, now uninformative in regard to discriminating between the response options), whereas the alternatives differed only

in the color of the previously irrelevant part. This manipulation enabled us to isolate performance for the irrelevant feature without requiring an explicit report (and awareness of one's knowledge or lack thereof) of that stimulus' properties.

Assessment of Awareness

Immediately following the critical trial, participants were (falsely) informed that for approximately half of the participants in the experiment, the final trial had differed from the others. This was done to better control their prior beliefs about the probability of the two options (i.e., pull them toward indifference about the response options if they had no internal signal) as well as to minimize potential suspicion-driven reactivity (e.g., treating the question as a psychologist "trick question"). Participants were then asked to indicate, via a binary response, which of the two groups they believed they were in. Following their response, participants were asked to write a free-text explanation of what they believed had changed in that trial.

Free-text responses were subsequently categorized into five groups by two independent judges blind to the experimental hypotheses:

- Aware that the relevant feature was now constant
- Aware that the irrelevant feature now varied
- Aware of both changes
- Perceived no difference
- Other (does not fit any of the above)

Preregistered Hypotheses and Statistical Analysis

All experiments were preregistered on OSF prior to data collection. Statistical inference used Bayesian methods throughout, with a primary stopping criterion of $BF = 8$ (evidence for the research hypothesis), $BF = 1/8$ (evidence for the null) for all stated hypotheses.

For Experiments 1A and 1B, three preregistered questions were transformed into statistical hypotheses:

1. Does irrelevance reduce performance to chance level, using Bayesian binomial tests against chance;
2. Does surprise alone reduce performance to chance level, assessed using Bayesian binomial tests; and
3. Does the effect of irrelevance exceed the effect of surprise, assessed using Bayesian contingency tests?

For Experiment 2A, three preregistered and one exploratory (#4) question were tested as statistical hypotheses; in Experiment 2B all four questions and corresponding statistical hypotheses were preregistered:

1. Is performance on the irrelevant feature better than chance
2. Is performance following surprise better than chance
3. Does performance on the irrelevant feature differ from performance following the surprise probe
4. Does performance vary as a function of awareness, assessed using Bayesian analyses of the binary awareness judgment and free-text categorizations?

As stated above, free-text responses were coded. This was done independently by two research assistants who were blind to the experimental hypotheses and to whether the participant succeeded on the critical trial. Following individual coding, they compared their responses and resolved cases of conflict. Raw agreement between the two (before comparison) was 87% with a Cohen's kappa of 0.833.

Results

Relevance renders visible features unreportable

In each experiment, participants saw differently colored flowers in pots (one flower in one pot per trial, see Fig. S4) and repeatedly reported the color of one, either the flower or the pot, and were unexpectedly probed, on trial 42 (the 'critical trial'), on the color of the other (Fig. 1C). In Experiment 1A, stimuli were presented at a fixed central location; in Experiment 1B, stimuli appeared at one of eight randomized positions and orientations, introducing spatial uncertainty to rule out advance allocation of spatial attention. All other procedural details were identical (see Supplementary Materials and Methods).

Results across the two preregistered Experiments 1A (N = 60) and 1B (N = 47) show that participants performed near-ceiling on the task-relevant feature during both training and main-task trials, confirming task comprehension, task ease, and perceptual discriminability. On the critical trial, however, performance on the previously task-irrelevant feature dropped to chance levels in both experiments (1A: M = 20%, BF $\approx$ 0.195; 1B: M = 23%, BF $\approx$ 0.158), despite participants having repeatedly viewed the stimuli throughout the task (see Fig. 1A & 1B).

This reduction in performance could not be explained by surprise alone. On trial 21, participants were surprised with an unexpected task before reporting the relevant feature, providing a measure of the effect of surprise on performance (Fig. 1C). Following the unexpected number task, performance on the feature of the task-relevant element indeed worsened but remained well above chance in both experiments (1A: M = 67%, BF $\gg$ 1000; 1B: M = 55%, BF $\approx$ 3143), and Bayesian comparisons strongly favored a difference between the performance after surprise on the relevant object part's probe and the performance on the probe of the irrelevant part (1A: BF $\gg$ 1000; 1B: BF $\approx$ 37). In Experiment 1B, this dissociation persisted despite spatial uncertainty that prevented advance allocation of spatial attention to the relevant region (i.e., a standard spatial-attention explanation is insufficient here). After the previously irrelevant feature was probed and became task-relevant, performance rose to a level that is substantially above chance (1A: M = 85%; 1B: M = 73%).

In both Experiments 1A and 1B, participants also indicated the basis of their recent response (on trials 1, 4, 7, 13, 21, 24, 27, 34, 38, 43, 44, 45 and 46), by selecting one of four options: memory, intuition, guess, or other. For the results, see the supplementary materials (Materials and Methods, subsection Metacognitive assessment).

Unreportable features guide behavior

Experiments 2A (N = 69) and 2B (N = 408) (given a Bayesian stopping criterion of BF $\geq$ 8 or $\leq$ 1/8 for all hypotheses; for the sequential analyses see Supplementary Fig. S3), examined whether stimuli that are irrelevant when seen (and hence, are unreportable) influence behavior. The method was identical to the two previous experiments, save for the following changes: 1) training trials were folded into the number of trials for a total of 42 steps before the critical trial. 2) the probes (different images of the flower in the pot) differed only in the

variability of the color of the task-relevant or task-irrelevant object-part (pot or flower in the pot). For probing the relevant object part (trials 1-42), the response options varied only on the colors of the relevant part (i.e., the irrelevant part's color was constant and correct for all options). Conversely, on the critical and last trial (trial 43), when the irrelevant part's color was probed, response options varied only on the color of the (previously) task-irrelevant part (i.e., the relevant part's color was constant and correct for all options). 3) following the critical trial, participants were immediately asked about the (last) trial, (a) whether they thought it was different from any other trial (binary), and (b) how it differed (free text). Both questions served as measures of awareness (Fig. 2C). 4) no other metacognitive information was collected.

As predicted and in contrast with the first two experiments that used explicit probes, results showed that, performance on the critical trial was well above chance level ($\sim 33.33\%$) in both experiments (2A: $M = 55\%$, $BF = 175$; 2B: $M = 60\%$, $BF \gg 1000$), though, as the following section shows, this aggregate effect masks the influence of a participant's awareness of the change. Also differing from the previous experiments, Bayesian analyses support no difference between performance on the probe of the relevant dimension after surprise and the (surprise) probe of the irrelevant dimension (both $BF < 0.001$). The preregistered replication in Experiment 2B confirmed the robustness of these effects.

Awareness associated with lower success

As stated above, immediately following the critical trial, participants indicated via a binary response probe whether they had noticed any change in the last trial, and then provided a free-text explanation of what they believed had changed. Responses were coded into five categories. Because the "Other" category comprised responses that were not theoretically interpretable or informative, it was excluded from the analysis, leaving four categories for analysis.

Awareness of change: Binary judgment. In both experiments, participants who reported no awareness of a change showed the highest level of performance on the irrelevant feature (2A: $M = 80\%$; 2B: $M = 69\%$). Performance was lower among participants who reported awareness of a change (2A: $M = 36\%$; 2B: $M = 54\%$). Bayesian analyses confirmed that the two groups differed from each other and from chance in both experiments (2A: $BF = 250.78$; 2B: $BF = 9.49$, see supplementary Fig. S3).

Awareness of change: Free-text analysis. A finer-grained categorization of free-text responses replicated and extended this pattern. In both experiments, participants reporting no perceived difference showed the highest level of performance (2A: $M = 88.2\%$, $n = 17$; 2B: $M = 75.3\%$, $n = 73$). In Experiment 2A, participants who accurately reported that both features had changed showed the lowest performance level of all groups ($M = 16.7\%$, $n = 6$), with intermediate performance among those reporting a change in the irrelevant feature ($M = 50.0\%$, $n = 24$) or stability in the relevant feature ($M = 35.7\%$, $n = 14$). In Experiment 2B, this gradient did not fully replicate; the performance of participants who reported that both changes occurred (who were arguably the most aware that they were probed on the irrelevant knowledge) was intermediate rather than the worst of the three groups ($M = 60.2\%$, $n = 88$). Crucially, the key result held: participants classified as having perceived no difference in the last trial compared to all previous ones (i.e., those who lacked any awareness that they were probed about the irrelevant knowledge) consistently outperformed all other categories in both experiments (see Fig. 3). For a more detailed breakdown of the free-text categories, see Supplementary Fig. S7.

Interpretation

These experiments establish three results. First, task-irrelevance alone is sufficient to render clearly visible, meaningful stimuli unreportable. Second, these unreportable features nonetheless guide behavior, leaving a trace equivalent 'in strength' to that of task-relevant features following a surprise task. Third, and most strikingly, awareness of the irrelevant feature is associated with worse, not better, behavioral performance; the less observers reported awareness of the critical manipulation, the better was their performance on the perceptual comparison task involving task irrelevant information.

The rich/sparse debate reconsidered

These findings do not simply favor one side of the rich/sparse debate; they reframe it. Sparse accounts interpret failures of report as evidence that experience itself is limited. Still, the present results show that – under conditions specifically designed to preserve perceptual availability –

unreportable features can guide behavior as effectively as reportable ones. At the very least, this suggests that report failure does not necessarily imply the absence of a usable perceptual representation or, for that matter, the absence of a conscious experience.

Rich accounts, by contrast, typically interpret failures of report as reflecting limitations of access, memory, or decision processes rather than limitations of perception itself. Although the present findings are broadly compatible with this perspective, existing accounts are still insufficient to fully explain them. If failures of report merely reflected downstream limitations, awareness and behavioral performance should be positively related: better access should produce both better reports and better performance. The inverse relationship observed here is difficult to reconcile with that prediction. Together, these findings suggest that the rich/sparse dichotomy is not merely unresolved but may be misconceived; the two positions share the assumption that report is a transparent, if imperfect (16, 17), window into experience. The present results suggest that report and experience can be reliably dissociated. Under some conditions, they may even vary inversely.

On the alternative interpretations

One might interpret these results as a clear-cut demonstration of unconscious processing, a non-reportable perceptual system operating outside awareness (18). Although this was not tested directly, we consider this interpretation implausible. The stimuli were deliberately prolonged, unmasked, meaningful, and presented without competing load, conditions specifically designed (and conventionally employed) to ensure conscious perception, or seeing. Furthermore, performance on the irrelevant object-part's probe was substantial, not merely above chance, and equivalent to performance on features that were explicitly attended and reported. To attribute this to unconscious processing is to commit to the two strong claims that consciousness is absent under conditions optimized for its presence and that it contributes nothing to performance. We conclude that a more plausible interpretation is that the relevant features were consciously perceived but failed to be transformed into reportable awareness: a distinction that is not merely terminological (cf. (7)) but empirically consequential, as the inverse awareness-performance relationship demonstrates. Unconscious processing accounts predict no relationship between awareness and performance, or a positive one (19, 20); therefore, they will struggle to accommodate a reliable negative relationship without additional post hoc assumptions.

A related alternative interpretation holds that the inverse awareness-performance relationship reflects a strategy shift (21, 22) rather than a dissociation between layers of consciousness: participants who became aware of the

change may have abandoned their initial perceptual comparison in favor of deliberate re-evaluation, and it is this strategic interference (rather than any property of awareness itself) that depressed performance. We consider this account incomplete for two reasons. First, it presupposes precisely what needs to be explained: that unaware participants were successfully guided by something. In Experiment 2B alone, 150 participants reported no awareness of any change, yet performed substantially above chance as a group and this robust, large-sample, finding cannot be attributed to deliberate processing. If deliberate re-evaluation is ineffective for irrelevant features, as the aware group's performance confirms, then the above-chance performance of these 150 unaware participants must reflect guidance from a representation formed without deliberate encoding and accessed without deliberate retrieval (18). Thus, the strategy-shift account may explain the aware group's failure but cannot account for the unaware group's success making it less parsimonious.

Second, the strategy-shift account generates a specific empirical prediction: if aware participants were engaging in deliberate re-evaluation, slower response times should predict worse performance specifically in the aware group, as deliberation consumes time without producing accurate responses (23). To examine this possibility, we fitted a logistic regression predicting critical-trial performance from response time, awareness group, and their interaction. Response time predicted performance across the two groups, with slower responses associated with lower accuracy. Crucially, the awareness factor remained predictive of performance on the irrelevant dimension after controlling for response time, and the response time $\times$ awareness interaction was not reliable. A BIC-based model comparison further favored the additive model over the interaction model, providing evidence that the awareness level provided distinct predictive information (see Supplementary: Response time models). Although this analysis cannot definitively rule out strategy-based explanations, it provides no support for the specific prediction that it was deliberative re-evaluation that impaired performance only among participants who reported awareness of the change.

Taken together, these observations suggest that a simple strategy-shift account is insufficient to explain the full pattern of results. In particular, it may explain the reduced performance associated with reported awareness more readily than it explains the robust above-chance performance observed among participants who reported no awareness of the critical manipulation.

A layered account of consciousness

The present pattern of results is naturally accommodated by the MRM framework (Berlinger & Eitam, 2026), which proposes that consciousness comprises two functionally distinct layers: an early perceptual representation that is high-bandwidth and relevance-unfiltered, and a later awareness that arises when perceptual content is enriched by relevance-weighted knowledge activation. On this account, task-irrelevant features are perceptually (consciously) represented as phenomenal experience but fail to activate the knowledge or processes required for reportable awareness, and can therefore guide behavior while remaining unavailable to report. The inverse awareness-performance relationship follows directly: when knowledge activation does occur, it indexes perceptual content that was transformed into awareness. As compared to raw sensory experience, awareness is a predictably lossy representation.

This account stands in tension with strong report-based interpretations of the Global Workspace Theory (GWT), which holds that conscious experience is constituted by global broadcast, the wide distribution of information across cortical networks that makes it available for report and control (24, 25). On such accounts, information that remains task-irrelevant, unreportable, and unaccompanied by awareness would be classified as non-

conscious. The present findings complicate that classification. The irrelevant features did not merely produce weak 'residual' priming; they guided behavior reliably, and their influence was strongest precisely when participants reported no awareness of the change. This pattern is not easily accommodated by a unitary broadcast account in which reportability, awareness, and behavioral availability are expected to co-vary. Integrated Information Theory (IIT) is less directly challenged, since it grounds consciousness in intrinsic causal structure rather than reportability (26). However, the inverse relationship between awareness and performance still calls for an account of how perceptual organization, access, and report can dissociate. In this respect, the present findings motivate a layered view in which perceptual representation and reportable awareness are functionally distinct.

Implications and future directions

At the very least, these findings call for a fundamental reconsideration of how consciousness is operationalized in empirical research. The present results demonstrate that reportability, awareness reports, and behavioral availability can dissociate, and under some conditions may even vary in opposite directions. This observation raises important questions about the extent to which report-based measures are used as proxies for consciousness (27); they may not be just imperfect; they may, in some conditions, be measures of a categorically different phenomenon.

Future work should seek to identify the neural and computational processes underlying the dissociation observed here, determine the conditions under which perceptual content transitions into awareness, and examine whether the inverse awareness-performance relationship generalizes across modalities and stimulus types.

The present results demonstrate that reportability, awareness reports, and behavioral influence cannot be assumed to reflect a single underlying process. Any adequate account must explain how these measures can dissociate, and under some conditions vary in opposite directions. More generally, the findings suggest that the relationship between perception, awareness, report, and behavior is more complex than is often assumed.

Declarations

All participants provided informed consent prior to participation. Participants were presented with a description of the task and its duration and had an inherent right to withdraw at any time, as guaranteed through Prolific platform's policies. Only after reading the description did participants click to begin the experiment.

Acknowledgments: Artificial-intelligence-assisted technologies were used in proofreading of this manuscript, following which it was reevaluated by human proofreaders. The multi-representation media (MRM) model discussed in this article was previously developed and published by two of the authors.

Funding: Israel Science Foundation grant 1714/23 (BE)

Funding: United States-Israel Binational Science Foundation grant 2022/321 (BE)

Author contributions:

Conceptualization: AB, AG, BE

Methodology: AB, BE

Investigation: AB
Formal analysis: AB, BE
Visualization: AB
Funding acquisition: BE
Supervision: AG, BE
Writing – original draft: AB
Writing – review & editing: AB, AG, BE

Diversity, equity, ethics, and inclusion: This research received approval from a local ethics board (ID: 422/22).

Competing interests: The authors declare that they have no competing interests.

Data, code, and materials availability: The hypotheses and methods were preregistered on OSF for Experiments 1A, 1B, 2A, and 2B. All study materials, data, and analysis scripts are publicly available on OSF under Experiment 2B: <https://osf.io/vyjze/>. The preregistrations are available at <https://osf.io/2anp5/>, <https://osf.io/sf9y6/>, <https://osf.io/tgjrd/>, and <https://osf.io/vyjze/>.

References

1. M. S. Jensen, R. Yao, W. N. Street, D. J. Simons, Change blindness and inattention blindness. *WIREs Cognitive Science* **2**, 529–546 (2011).
2. D. J. Simons, D. T. Levin, Change blindness. *Trends in Cognitive Sciences* **1**, 261–267 (1997).
3. A. Mack, I. Rock, Inattention blindness: Perception without attention. (1998).
4. U. Neisser, R. Becklen, Selective looking: Attending to visually specified events. *Cognitive psychology* **7**, 480–494 (1975).
5. D. J. Simons, C. F. Chabris, Gorillas in Our Midst: Sustained Inattention Blindness for Dynamic Events. *Perception* **28**, 1059–1074 (1999).
6. M. A. Cohen, D. C. Dennett, N. Kanwisher, What is the bandwidth of perceptual experience? *Trends in cognitive sciences* **20**, 324–335 (2016).
7. M. A. Cohen, D. C. Dennett, Consciousness cannot be separated from function. *Trends in cognitive sciences* **15**, 358–364 (2011).
8. N. Block, On a confusion about a function of consciousness. *Behavioral and brain sciences* **18**, 227–247 (1995).
9. N. Block, Consciousness, accessibility, and the mesh between psychology and neuroscience. *Behavioral and Brain Sciences* **30**, 481–499 (2007).
10. V. A. Lamme, Towards a true neural stance on consciousness. *Trends in cognitive sciences* **10**, 494–501 (2006).
11. B. Eitam, Y. Yeshurun, K. Hassan, Blinded by irrelevance: Pure irrelevance induced “blindness”. *Journal of Experimental Psychology: Human Perception and Performance* **39**, 611–615 (2013).
12. B. Eitam, R. Shoval, Y. Yeshurun, Seeing without knowing: task relevance dissociates between visual awareness and recognition. *Annals of the New York Academy of Sciences* **1339**, 125–137 (2015).

13. B. Wyble, M. Hess, R. E. O'Donnell, H. Chen, B. Eitam, Learning how to exploit sources of information. *Mem Cogn* **47**, 696–705 (2019).
14. H. Chen, B. Wyble, Amnesia for Object Attributes: Failure to Report Attended Information That Had Just Reached Conscious Awareness. *Psychol Sci* **26**, 203–210 (2015).
15. A. M. Berlinger, B. Eitam, Rich Experience and Sparse Awareness: Sensing More Than We Know. *Elements in Perception* (2026).
16. H. Lau, D. Rosenthal, Empirical support for higher-order theories of conscious awareness. *Trends in cognitive sciences* **15**, 365–373 (2011).
17. M. Michel, D. Beck, N. Block, H. Blumenfeld, R. Brown, D. Carmel, M. Carrasco, M. Chirimuuta, M. Chun, A. Cleeremans, Opportunities and challenges for a maturing science of consciousness. *Nature human behaviour* **3**, 104–107 (2019).
18. S. Dehaene, J.-P. Changeux, Experimental and Theoretical Approaches to Conscious Processing. *Neuron* **70**, 200–227 (2011).
19. J. L. Voss, K. A. Paller, Remembering and knowing: Electrophysiological distinctions at encoding but not retrieval. *Neuroimage* **46**, 280–289 (2009).
20. J. L. Voss, K. A. Paller, An electrophysiological signature of unconscious recognition memory. *Nature neuroscience* **12**, 349–355 (2009).
21. S. Hellmann, M. Zehetleitner, M. Rausch, Simultaneous modeling of choice, confidence, and response time in visual perception. *Psychological review* **130**, 1521 (2023).
22. G. Pennycook, J. A. Cheyne, D. J. Koehler, J. A. Fugelsang, Is the cognitive reflection test a measure of both reflection and intuition? *Behav Res* **48**, 341–348 (2016).
23. S. M. Fleming, E. J. Van Der Putten, N. D. Daw, Neural mediators of changes of mind about perceptual decisions. *Nature neuroscience* **21**, 617–624 (2018).
24. B. J. Baars, *A Cognitive Theory of Consciousness* (Cambridge University Press, 1993;
[https://books.google.com/books?hl=en&lr=&id=7w6lYeJRqyoC&oi=fnd&pg=PR11&dq=Baars,+B.+J.+A+Cognitive+Theory+of+Consciousness+\(1988\).&ots=CYaOc3-4bK&sig=7SLHjYrU0W6Bcdm20AI00kyCLbw](https://books.google.com/books?hl=en&lr=&id=7w6lYeJRqyoC&oi=fnd&pg=PR11&dq=Baars,+B.+J.+A+Cognitive+Theory+of+Consciousness+(1988).&ots=CYaOc3-4bK&sig=7SLHjYrU0W6Bcdm20AI00kyCLbw)).
25. S. Dehaene, J.-P. Changeux, L. Naccache, J. Sackur, C. Sergent, Conscious, preconscious, and subliminal processing: a testable taxonomy. *Trends in cognitive sciences* **10**, 204–211 (2006).
26. L. Albantakis, L. Barbosa, G. Findlay, M. Grasso, A. M. Haun, W. Marshall, W. G. Mayner, A. Zaeemzadeh, M. Boly, B. E. Juel, Integrated information theory (IIT) 4.0: formulating the properties of phenomenal existence in physical terms. *PLoS computational biology* **19**, e1011465 (2023).
27. F. Stockart, M. Schreiber, P. Amerio, D. Carmel, A. Cleeremans, L. Y. Deouell, Z. Dienes, P. Elosegi, S. Gayet, A. Goldstein, Studying unconscious processing: Contention and consensus. *Behavioral and Brain Sciences*, 1–77 (2025).
28. J. W. Peirce, PsychoPy—psychophysics software in Python. *Journal of neuroscience methods* **162**, 8–13 (2007).
29. S. Palan, C. Schitter, Prolific. ac—A subject pool for online experiments. *Journal of behavioral and experimental finance* **17**, 22–27 (2018).

30. G. Sperling, The information available in brief visual presentations. *Psychological monographs: General and applied* **74**, 1 (1960).
31. Z. Dienes, R. Scott, Measuring unconscious knowledge: distinguishing structural knowledge and judgment knowledge. *Psychological Research* **69**, 338–351 (2005).

Figures

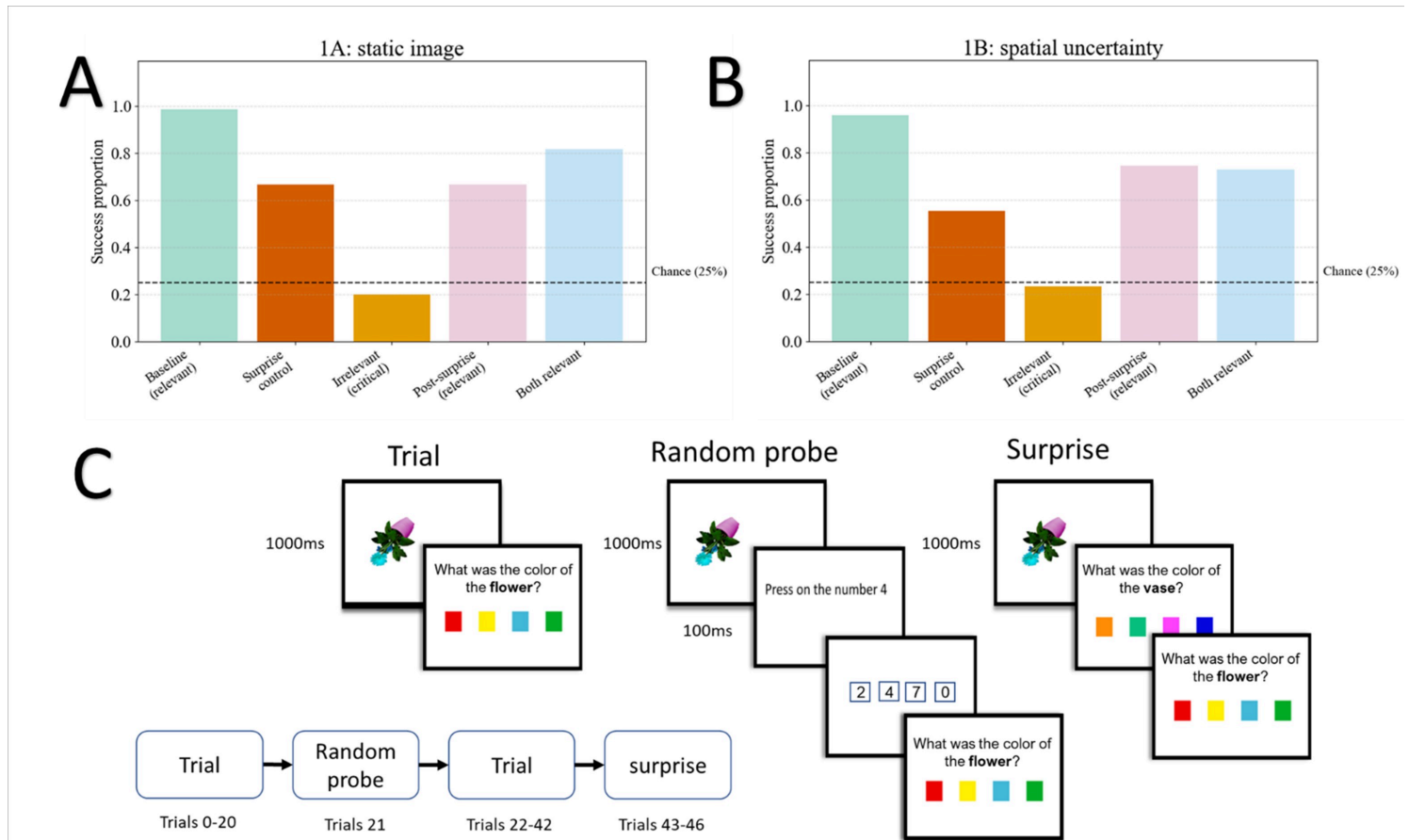


Figure 1

Irrelevance induced the failure of report. (A) Experiment 1A (fixed presentation; $N = 60$) and (B) Experiment 1B (under spatial uncertainty; $N = 47$). Accuracy is shown for each condition: baseline performance on the task-relevant feature, performance on the task-relevant feature following a surprise-inducing probe, the critical probe of the previously irrelevant feature, and subsequent trials in which both features became relevant. In both experiments, performance remained high for relevant features and was reduced by surprise, and dropped further to chance levels when the same, clearly visible feature was task-irrelevant. Performance recovered once the feature became task-relevant again. The dashed line indicates chance performance (25%); error bars represent 95% confidence intervals. (C) Schematic of the experimental procedure. In Experiment 1A, stimuli were presented at a fixed central location; in Experiment 1B, stimuli appeared at one of eight randomized positions and orientations, introducing spatial uncertainty. All other procedural details were identical. Participants repeatedly reported the color of a task-relevant object-part's feature (the pot or the flower in the pot), selecting from four color options presented as squares below the stimulus. The flowchart illustrates the sequential block structure with arrows indicating direction: 20 trials in which participants reported the relevant color, followed by

a surprise control trial in which participants completed an unexpected number task ('press the 4 key') before again reporting on the relevant object part's color, followed by 20 further trials and finally, the critical surprise trial in which participants were unexpectedly probed about a previously irrelevant object-part's feature, followed by three trials in which both features became task-relevant.

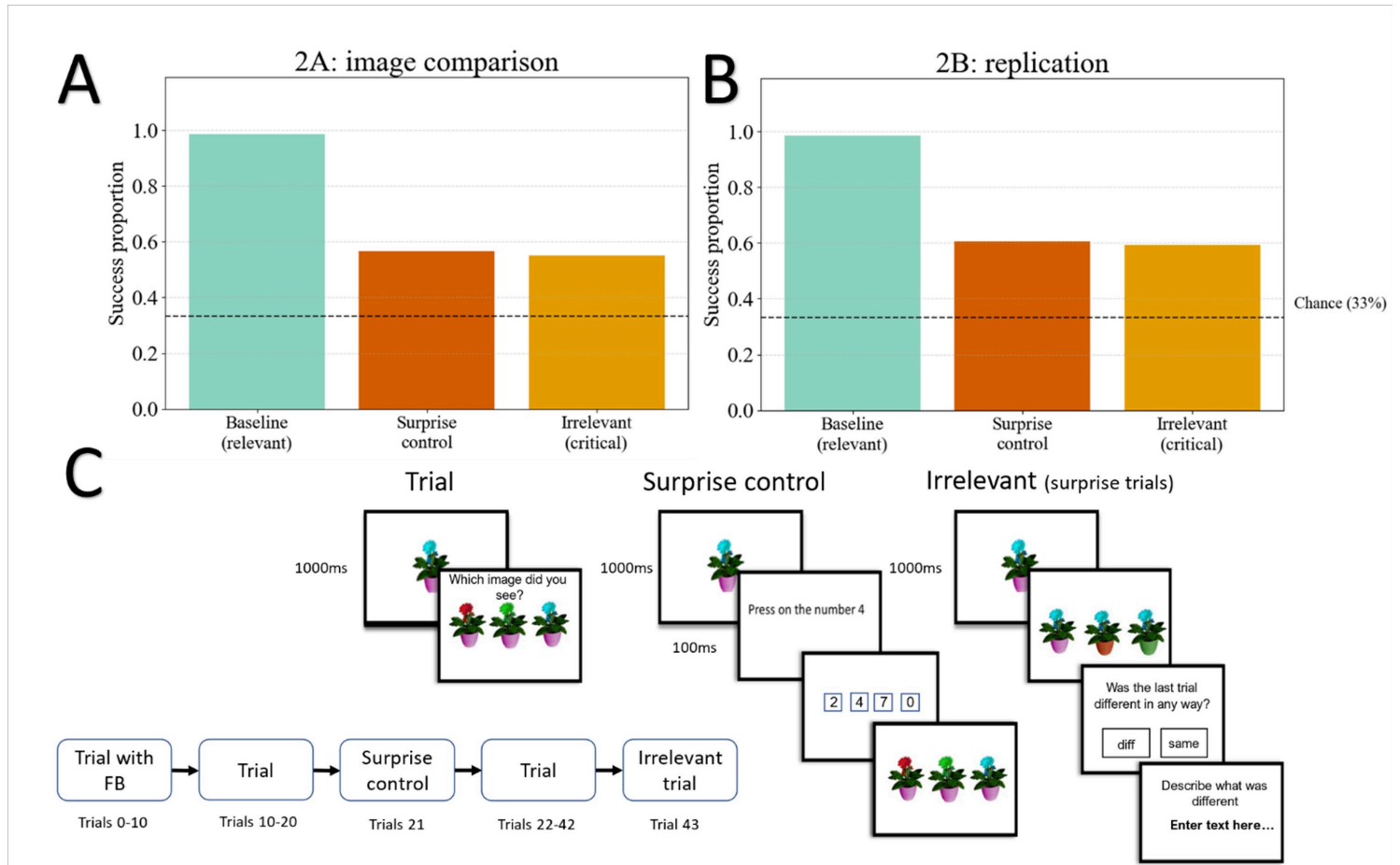


Figure 2

Task-irrelevant features can guide perceptual identification. (A) Experiment 2A (image comparison; N = 69) and (B) Experiment 2B (preregistered replication; N = 408). Mean accuracy in a forced-choice identification task is shown for baseline trials (relevant feature identification), a surprise control condition (same discrimination performed after a surprise instruction to click on the digit 4 on the screen), and the critical trial in which success in 1-back discrimination depended in full on correctly identifying a previously task-irrelevant feature. The critical trial revealed that performance was well above chance. Furthermore, performance that depended on a feature of a previously irrelevant object-part did not differ from the surprise control in which the relevant object-part's feature was probed. The dashed line indicates chance performance (~33%). (C) Schematic of the experimental procedure. The flowchart shows the sequential block structure: trials 0-10 consisted of training trials with corrective feedback; trials 10-20 were standard identification trials; trial 21 was the surprise control trial; trials 22-42 were further standard trials; and trial 43 was the critical trial probing the irrelevant object part's feature. On each standard trial, participants viewed a stimulus for 1000ms and selected which of three alternative images, differing only in the color of the task-relevant object-part's feature, they had just seen. In the surprise control trial, an unexpected number task preceded the identification response. On the critical trial, response alternatives differed only in the color of the previously irrelevant object-part's feature. Following the critical trial, participants

indicated via binary judgment whether the trial had differed from previous trials and provided a free-text description of any perceived change.

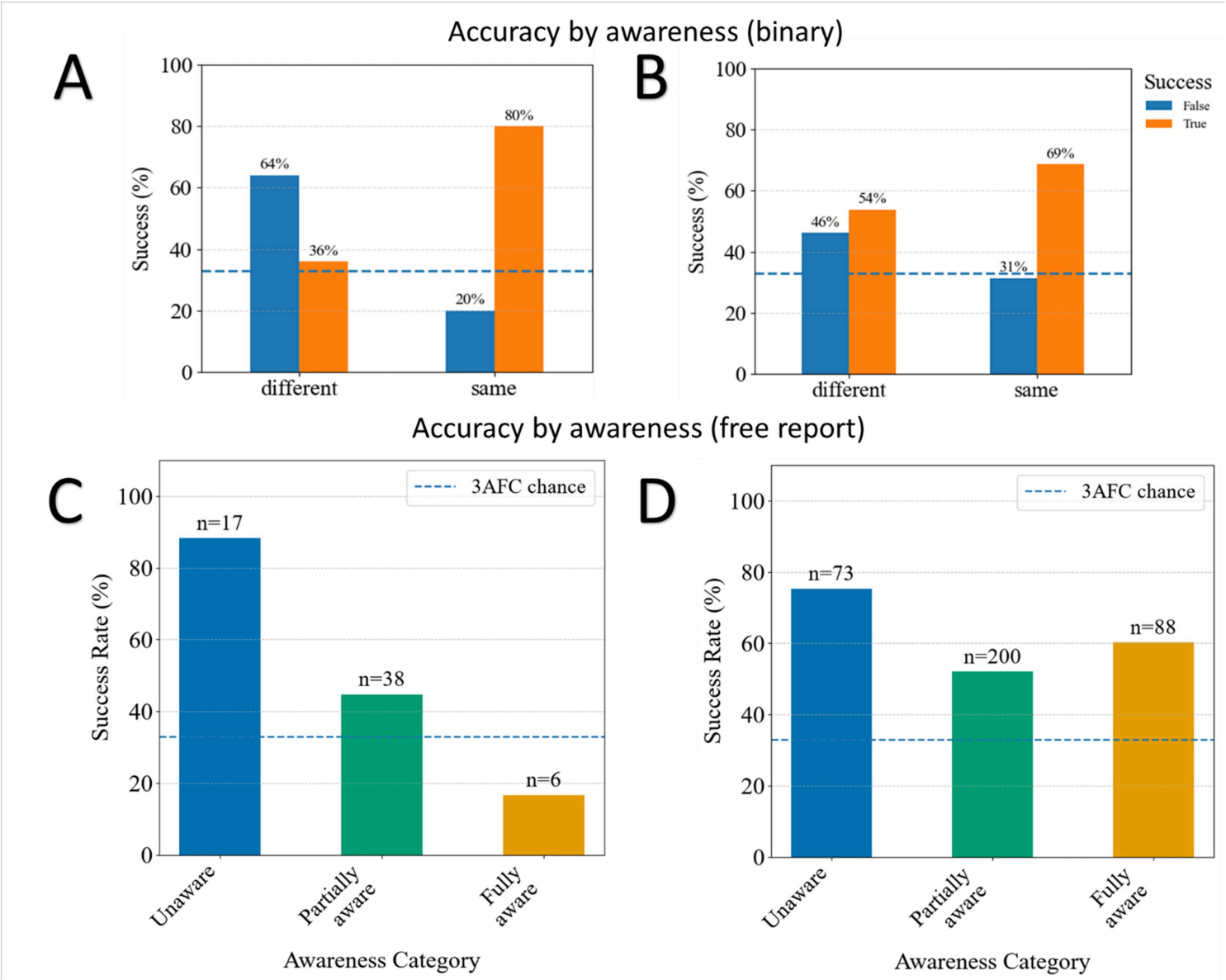


Figure 3

Awareness in the trial probing irrelevant information is associated with reduced performance on the previously irrelevant information. (A–B) Accuracy on the critical trial as a function of binary metacognitive report (“same” vs. “different”) in Experiment 2A (A) and Experiment 2B (B). On this trial, response alternatives differed only in the feature of the previously irrelevant object part. In both experiments, participants who reported no awareness of a change (“same”) showed higher accuracy than those who reported that the trial was different. (C–D) Accuracy as a function of participants’ free-text responses in Experiment 2A (C) and Experiment 2B (D), collapsed into three awareness categories: Unaware / no perceived difference (blue), Partially aware / aware that either the irrelevant feature varied or the relevant feature remained constant (green), and fully aware / aware of both changes (orange). Dashed lines indicate chance performance (33%).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Supplementary.docx](#)