Attention, awareness, and the right temporoparietal junction

Andrew I. Wilterson[a], Samuel A. Nastase[b], Branden J. Bio[a], Arvid Guterstam[a,c,d], and Michael S. A. Graziano[a,b,1]

[a]Department of Psychology, Princeton University, Princeton, NJ 08544; [b]Princeton Neuroscience Institute, Princeton University, Princeton, NJ 08544; [c]Department of Clinical Neuroscience, Karolinska Institutet, 171 77 Solna, Stockholm, Sweden; and [d]Department of Neurology, Karolinska University Hospital, 171 77 Solna, Stockholm, Sweden

Edited by Michael E. Goldberg, Columbia University, New York, NY, and approved May 7, 2021 (received for review December 30, 2020)

The attention schema theory posits a specific relationship between subjective awareness and attention, in which awareness is the control model that the brain uses to aid in the endogenous control of attention. In previous experiments, we developed a behavioral paradigm in human subjects to manipulate awareness and attention. The paradigm involved a visual cue that could be used to guide attention to a target stimulus. In task 1, subjects were aware of the cue, but not aware that it provided information about the target. The cue measurably drew exogenous attention to itself. In addition, implicitly, the subjects' endogenous attention mechanism used the cue to help shift attention to the target. In task 2, subjects were no longer aware of the cue. The cue still measurably drew exogenous attention to itself, yet without awareness of the cue, the subjects' endogenous control mechanism was no longer able to use the cue to control attention. Thus, the control of attention depended on awareness. Here, we tested the two tasks while scanning brain activity in human volunteers. We predicted that the right temporoparietal junction (TPJ) would be active in relation to the process in which awareness helps control attention. This prediction was confirmed. The right TPJ was active in relation to the effect of the cue on attention in task 1; it was not measurably active in task 2. The difference was significant. In our interpretation, the right TPJ is involved in an interaction in which awareness permits the control of attention.

attention | awareness | temporoparietal junction | predictive modeling | fMRI

The Attention Schema Theory

The attention schema theory (AST) relates attention to subjective awareness (1–4). In the theory, although attention and awareness are not the same and can vary independently, they do have a specific cognitive relationship to each other. Awareness permits the control of attention. In this introduction, we first briefly summarize the theory and then describe a specific set of experiments that tested some of its predictions.

Psychology and neuroscience distinguish between many categories and subcategories of attention: arousal versus selective attention, exogenous and endogenous attention, spatial and feature attention, and so on. AST concerns the kind of attention that requires a complex control of allocation, namely, selective attention to particular locations, features, or other items. Selective attention, classically, can be drawn to an item exogenously (such as by a sudden movement that automatically attracts attention) or can be directed endogenously (such as when a person searches a scene, directing attention from item to item). Often, of course, attention shifts as a result of both processes. AST proposes that the brain constructs an internal model of selective attention, a constantly updating set of information that describes and monitors the current state of attention, predicts how that state may transition into future states, and predicts consequences of attention on behavior, decision-making, and memory (1–4). This attention schema is both descriptive and predictive and is the brain's simplified self-account of attention. The model is used to help control attention. In a similar manner, the motor system controls the arm with the help of an internal model of the arm, a part of what is sometimes called the body schema (5–7).

If such a model of attention exists, would people be able to gain cognitive access to it, just as we can gain at least some cognitive access to the body schema? And if we do so, what would that internal information lead us to verbally claim about ourselves? According to AST, when we claim to have subjective awareness, the claim stems from the descriptive information present within the attention schema. We are, in effect, describing our own attention, filtered through the brain's slightly schematized way of representing attention. Logically, all claims that we make about ourselves must be dependent on information in the brain. Thus, our claim to have subjective awareness must derive from an internal information set. AST proposes a specific information set as the source for that claim: an internal, schematic model of attention. Many alternative theories of consciousness follow a different approach, proposing that a process in the brain causes a nonphysical feeling to emerge. In our approach, instead, information in the brain causes the system to formulate conclusions about itself. Our approach avoids nonscientific magical or nonphysical feelings (1–4).

One of the lines of reasoning to lead to AST involves the known close relationship between awareness and attention. The two are in some ways similar phenomena and often, though not always, covary (4). Yet attention is a physical, mechanistic process of selective data enhancement in the brain that can be objectively measured, whereas awareness is an intangible property that we only know about personally and that we attest to. According to AST, the reason for this relationship between awareness and attention is that our introspective understanding of awareness, and our belief that we have it, is based on an information set in the brain that depicts attention. The reason why we typically understand awareness to be a nonphysical, ethereal essence is because the brain's model of attention is impoverished, schematic, and lacking in any details

Significance

We all intuitively know what it is to be subjectively aware of something. For decades, psychologists have also studied the brain's mechanisms of directing attention. But awareness and attention are not the same and can be separated. How do they relate? Here, we show that a part of the brain, the right temporoparietal junction, may be involved in a specific relationship in which awareness of an item permits the control of attention on that item.

Author contributions: A.I.W. and M.S.A.G. designed research; A.I.W. and B.J.B. performed research; A.I.W., S.A.N., A.G., and M.S.A.G. analyzed data; and A.I.W. and M.S.A.G. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence may be addressed. Email: graziano@princeton.edu.

Published June 14, 2021.

about the physical mechanisms of attention. The reason why awareness usually tracks attention is that it serves as a model of attention, much as the brain's model of the arm typically tracks the physical arm and dissociates only when the model makes an error. In cases when awareness and attention are dissociated (such as when a person attends to a stimulus but is unaware of it), in this interpretation, the model is in error and has become misaligned from actual attention. The practical consequence of that misalignment, according to AST, is that, with an inaccurate control model, the endogenous control of attention should be impaired. AST is therefore a testable, falsifiable theory. Its central contention is that awareness is the control model for attention. If awareness of item X is compromised, and the endogenous control of attention with respect to item X is normal, then AST is falsified. When awareness is compromised, many functional capabilities are known to remain—but does the endogenous control of attention remain or disappear?

To test the relationship between attention and awareness, we recently conducted a series of behavioral experiments (8). We used a visual attention paradigm in which exogenous attention, endogenous attention, and awareness could be manipulated separately in human subjects. In the task, a visual cue statistically predicted the location of a subsequent visual target. In one version of the task (here called task 1), the cue was colored red, and the subjects were aware of the cue. In that version, subjects' attention was initially exogenously drawn to the cue, and the subjects were successfully able to endogenously shift attention from the cue to the nearby location where the target was anticipated to appear.

In a second version of the task (here called task 2), the cue was colored black, which caused it to be masked, such that subjects were unaware of it. In that condition, although the cue had an impact on subjects, drawing measurable exogenous attention to itself (8), the subjects were severely impaired at endogenously shifting attention from the cue to the predicted location. Crucially, in neither task 1 nor task 2, did subjects explicitly know that their attention was being manipulated or that the cue predicted the location of the target. No explicit strategy or reasoning was involved in either task. Rather, awareness of the cue enabled an implicit ability to control and shift attention from the cue to a nearby location, and a lack of awareness of the cue prevented that implicit ability.

The findings were interpreted as supporting AST. In the theory, awareness of the cue served as the brain's internal model of attention on the cue. Absent that model—absent any information about the attention that had been exogenously drawn to the cue—the controller could no longer shift that focus of attention in a controlled manner. The task would be as difficult as shifting the arm in a controlled manner if the body schema had failed to inform the system where the arm currently is.

Tasks 1 and 2, therefore, put AST to a stringent test. In the presence of awareness of the cue, people can endogenously control the attention that is on the cue, shifting it appropriately (task 1). In the absence of awareness of the cue, people can no longer endogenously control or adjust their attention on the cue in the same manner (task 2).

Other behavioral experiments also suggest that without an awareness of stimulus X, although attention can be exogenously drawn to X, the endogenous control of that attention on X is impaired. One study showed that, without an awareness of stimulus X, one's attention may be briefly, exogenously drawn to X but cannot be strategically sustained on X, even when X is task relevant; in that case, without awareness, attention dissipated and became less than it should have been for the task (9). Another study showed that without an awareness of stimulus X, when exogenous attention is drawn to X, one cannot strategically suppress that attention, even when X is a distractor that is best ignored; in that case, without awareness, attention was higher than it should have been for the task (10). Together, these previous behavioral experiments support the central proposal of AST that awareness is a control model for attention. The model provides information about what attention is doing and what it is likely to do next, such that attention on an item can be endogenously controlled—sustained, adjusted up or down, or shifted away, as needed—and without awareness of the item, while other processes often remain and attention itself can remain, the endogenous control of attention is impaired.

One of the challenges in consciousness research is that being conscious of something does not have many experimentally demonstrable benefits. Without subjective awareness of an object, people can still react to it. Under some circumstances, they can point to it, avoid walking into it, and have their decisions and words influenced by it. Given the range of abilities outside awareness, does awareness have any definite purpose? Why can we not behave like complex, intelligent agents without any subjective awareness? We suggest that, without awareness of an item, the control of attention with respect to that item collapses. When the control of attention collapses, our behavior collapses. Creating a cognitive plan and executing it—like getting the jelly from the fridge, the knife from the drawer, and spreading the jelly on a slice of bread—requires a controlled, sequential movement of attention from one item to the next. Intelligent agency depends on the strategic control of attention, and AST gives a precise account of how the control of attention in turn depends on awareness.

Identifying Brain Networks Related to the Attention Schema

The behavioral experiments therefore suggest that AST makes accurate predictions about the relationship between awareness and attention. Awareness permits the well-regulated control of attention. Can evidence of this process be found in specific networks in the brain? The purpose of the present two experiments was to try to specify activity in the brain associated with constructing predictive models for the purpose of controlling attention. The experiments should not be taken as an overarching test of AST but as a way to test one corner of the theory.

In experiment 1, volunteers performed a version of task 1 in a functional MRI (fMRI) scanner. To explain the logic of the experiment, some of the details must be outlined here (see *Methods* for more task details). On each trial, a salient red cue appeared, followed 500 ms later by a target stimulus. Participants responded to the target as quickly as possible. On most trials (85%), the target appeared 3.5° to one side of the cue (termed "predicted trials"). On fewer trials (15%), the target appeared 3.5° to the other side of the cue ("nonpredicted trials"). Whether the predicted side was to the right or left of the cue was counterbalanced between subjects. In our previous study (8) and the present one, participants reacted more quickly to targets on the predicted side. By implication, the following events occurred: Attention was automatically, exogenously drawn to the cue; subjects endogenously shifted attention to the predicted side of the cue in anticipation of the upcoming target; on predicted trials, the target then appeared at the predicted location, and subjects could react to the target rapidly; and on nonpredicted trials, the target appeared at the nonpredicted location, and a corrective attention shift was required, causing an increase in response latency.

In our previous behavioral studies using this task (8), and the present study, we found that subjects were unaware of the predictive relationship between the cue and the target. They did not realize that they were endogenously shifting attention to a predicted location. Yet this implicit learning was robust and occurred quickly, within fewer than 50 trials. The attention control mechanism evidently rapidly learned to predict where to move attention. In this paradigm, on every trial that the target appeared at the nonpredicted location, the attention control mechanism was faced with a violation of its own prediction. We argue that whatever mechanism performs the task of learning and updating that predictive model, when it is faced with a violation of the model, it will

likely react. Based on the error signal, it may reevaluate or incrementally update the model. Thus, brain networks involved in constructing the predictive model should be more active on nonpredicted trials than on predicted trials. The crucial comparison in experiment 1 is therefore a subtraction between nonpredicted and predicted trials. Brain areas that show greater activity in nonpredicted than in predicted trials are candidates for constructing a predictive model that is used to guide attention, because they react to an error indicating that the predictive model has failed and needs to be updated.

In experiment 2, subjects performed a version of task 2 in the scanner. In task 2, the visual cue was black instead of red, causing it to be masked, such that subjects were no longer visually aware of it (8, 11). Prior experiments showed that the cue, though outside awareness, could still draw exogenous attention to itself (8). It affected the subjects. However, without awareness of the cue, reaction times were similar whether the target appeared to the predicted or nonpredicted side. By implication, without awareness of the cue, the attention control system could no longer construct a predictive model to guide attention. We argued, therefore, that brain areas involved in constructing a predictive model, and using that model to control attention, should satisfy two constraints: They should be sensitive to the difference between predicted and nonpredicted trials in task 1, and they should be insensitive to that difference in task 2.

We previously suggested (1, 2) that the temporoparietal junction (TPJ), especially the right TPJ, may play a central role in constructing an attention schema and using it to aid in the control of attention. This proposal was based on the suggestion that a part of the brain known to be involved in constructing models of other people's attention (during social cognition) may also be involved in constructing models of one's own attention (1, 12–17). Moreover, damage to the right TPJ can lead to severe forms of hemispatial neglect, a clinical disruption of attention and awareness (18, 19), and many fMRI studies indicate a role of the right TPJ in attention (20–27). The right TPJ is likely to be only one part of a larger network involved in these processes, but we focused our predictions on it because of the greater degree of information available about it.

A straightforward set of predictions emerges. In experiment 1, in which awareness of the cue allows the brain to predictively control attention, the right TPJ should become active in association with violations of the predictive model (thus more activity in nonpredicted trials than in predicted trials). In task 2, in which awareness of the cue is removed and as a result the brain no longer uses it to predictively control attention, the right TPJ should no longer show that specific activity difference between predicted and nonpredicted trials. Such a finding would suggest that the right TPJ is involved in making and updating predictions used in the control of attention. Other interpretations of activity in the right TPJ during similar attention tasks have been proposed (20–28). In the *Discussion*, we consider how the details of the present two tasks constrain the interpretations.

Methods

Subjects. All subjects provided informed consent and all procedures were approved by the Princeton Institutional Review Board. For experiment 1 (in which subjects performed task 1), 21 subjects were tested (18 to 26 y old, 14 women, normal or corrected-to-normal vision, and all right handed). One was excluded for poor performance (< 80% of trials correct). One was excluded because of excessive movement during the scan. One was excluded for explicitly noticing an aspect of the task that was intended to be implicit (see *Results for Experiment 1*). Thus, 18 subjects were included in the final analysis for experiment 1. Experiment 2 (in which subjects performed task 2) included 20 new subjects, untested in experiment 1 (18 to 52 y old, 13 women, normal or corrected-to-normal vision, and all right handed). One was excluded for falling asleep. Two were excluded for excessive movement during the scan. Thus, 17 subjects were included in the final analysis for experiment 2.

Tasks. Stimuli were projected with the Hyperion MRI Digital Projection System (Psychology Software Tools) at the end of the scanner bore. Each subject lay face up on the scanner bed with foam surrounding the head to reduce head movements and earplugs to reduce noise. All stimuli were developed and presented using the MATLAB psychophysics toolbox (29, 30). Subjects used a button box held in the right hand for behavioral responses.

Task 1 (performed in experiment 1) differed from task 2 (performed in experiment 2) in that subjects were subjectively aware of the cue in task 1 and not in task 2. To achieve this difference, the color of the cue was red in task 1 (thus visible despite the mask) and black in task 2 (thus successfully masked). Moreover, in task 1 during the initial instructions, as the stimuli were explained to the subjects, the cue and all other visible stimuli were explicitly pointed out, whereas in task 2, which was run on a separate set of subjects, during the initial instructions, all stimuli except the (perceptually invisible) cue were explicitly pointed out. Thus, the subjective invisibility of the cue in task 2 was ensured. In other respects, the behavioral tasks were the same in the two experiments.

Before running any trials, subjects were instructed outside the scanner on the task and given 20 practice trials. Practice was repeated if subjects scored under 80% accuracy on the first attempt.

Fig. 1 shows task 1. The display screen was initially a neutral gray. First, a fixation point (a 0.7° black circle) was shown at the center. Subjects were instructed to fixate on the point and to maintain fixation throughout the trial. After 1,200 ms, the cue appeared at a peripheral location. The cue was an annulus (inner diameter 2.75° and outer diameter 3.0°). The cue could be in any of 10 possible locations around the screen. The gray circles in Fig. 1, panel 2 illustrate possible locations of the cue (spaced 3.5° apart from each other laterally and 7.0° vertically). In task 1, the annulus was bright red as shown in Fig. 1. In task 2, the annulus was black.

After 35 ms, the cue disappeared and a visual mask in the form of an array of black annuli was presented, each the same size and shape as the cue, and arranged in a 7×7 grid at 3.5° spacing, excluding only the central position (Fig. 1, panel 3). In task 1, the initial cue stimulus, being red, was visible to subjects, despite the subsequent array of black annuli. In task 2, the initial cue stimulus, being black, was perceptually invisible, backward masked by the subsequent array of black annuli. The effectiveness of the masking was confirmed in the present experiment, as described in *Results for Experiment 1*, and has also been confirmed in previous studies (8, 11).

After another 465 ms (500 ms after the onset of the cue), the target stimulus was added to the display. The target stimulus could appear at one of two locations in relation to the cue: either one grid position (3.5°) to the left of the where the cue had been presented or one grid position to the right. In Fig. 1, the target is shown one grid location to the right of the cue. The target stimulus consisted of a thin white line visible against the neutral gray background, centered in the black annulus. The line was angled 10° from vertical, tilted either to the left or right. In Fig. 1, the target is shown tilted toward the right. For each subject, one direction was chosen as the predicted, or more frequently presented, direction. For example, if the predicted direction was to the right, then the target appeared to the right of the cue on 85% of trials (called predicted trials) and to the left of the cue on 15% of trials (called nonpredicted trials). Whether the predicted direction was to the right or left was counterbalanced across subjects.

After 80 ms, the target stimulus disappeared. After another 200 ms, all stimuli disappeared, including the black annuli and the fixation point. The screen then remained a blank, neutral gray until the subject's response was given or until the response window timed out after another 720 ms, as detailed next.

Subjects were instructed to respond as quickly as possible after the onset of the target by pressing one key if the target was tilted to the left and a different key if it was tilted to the right. Subjects were allowed a response window of 1,000 ms (80 ms of target stimulus presentation, 200 ms while the black annuli remained on screen, and 720 ms of blank screen). The limited response window was intended to encourage a speeded response. After the response window, a variable, intertrial interval (1,000 to 3,000 ms), was presented during which the screen was blank. Then the next trial began with the presentation of the fixation point.

Asking Subjects about Awareness of Task Events. During the instruction period, before the experiment and outside the scanner in experiment 1 (in which task 1 was performed), each visible stimulus was explicitly pointed out to subjects, including the red (clearly visible) cue, to ensure that subjects were indeed aware of the cue as intended. They were never told that the target was more likely to appear on one predicted side of the cue or that the cue was relevant to the task in any way. In experiment 2 (in which task 2 was performed), during the instructions, the black (masked and invisible) cue was not pointed out, although all other stimuli were. Pointing out the cue in task 1,

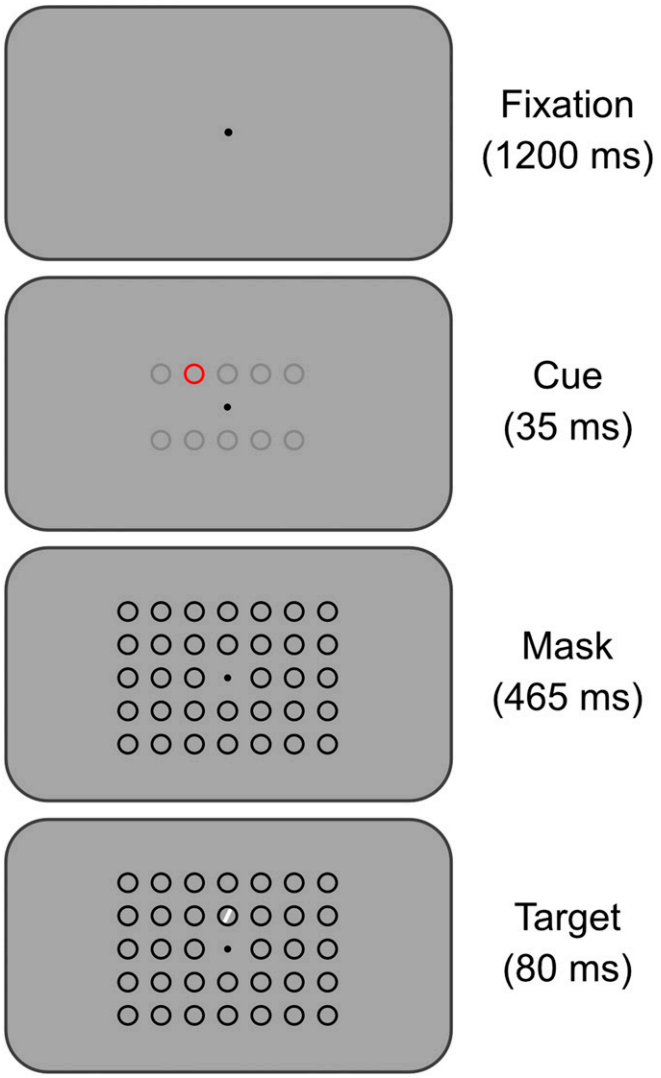

Fig. 1. Paradigm for task 1. Subjects in experiment 1 performed task 1. Subjects in experiment 2 performed task 2. The tasks were similar, except for the color of the cue (red in task 1 and black in task 2). After the fixation point appeared, the cue appeared in 1 of 10 possible locations. Gray circles in panel 2 show possible locations for the cue and were not visible to the subject. A mask of black distractor circles then appeared in a 5×7 grid. The target then appeared. The target could be in one of only two possible locations relative to the cue. It either appeared one grid location to the right of where the cue had been (as shown) or one grid location to the left of where the cue had been. One of these positions relative to the cue was defined as the predicted side and occurred on 85% of trials; the other was defined as the nonpredicted side and occurred on 15% of trials. Whether the predicted side was to the right or left of the cue was counterbalanced between subjects. Subjects discriminated the slant of the target in a reaction-time task.

and not pointing out the cue in task 2, was part of the manipulation to ensure that subjects were aware of it in task 1 and not in task 2.

Furthermore, different subjects were tested in the two experiments to ensure that the subjects performing task 2 were naïve to the cue and had no explicit knowledge that a cue was being presented. Subjects who know that a cue may be presented on some trials are primed to look for one and may be more likely to become aware of the masked cue. Moreover, interleaving aware and unaware trials presents a difficulty in testing whether the masking was successful. If subjects are asked at the end of each trial whether they saw the cue, then the cue becomes the target of a decision and a response, altering the attention and awareness that attaches to the cue (9). In contrast, by separating subjects into two groups, one group presented with the visible cue (task 1) and the other group presented with the masked cue (task 2), we could better ensure that subjects in the first group were aware of the cue and that subjects in the second group remained unaware of it throughout testing.

In experiment 1, after completing the task 1, subjects were given a verbal posttest questionnaire. They were asked, "Did you consistently see the red circular cue during each trial?" Subjects were also asked, "Did you notice any pattern or relationship between the cue and the target stimulus?" Finally, if they did not offer the correct relationship, they were asked, "Did you notice that the target tended to appear more often to one side of the cue, and do you know which that most frequent side was?" Their yes or no answers were recorded. In experiment 2, after completing task 2, subjects were asked, "Did you notice anything about the experiment that was not explained in the instructions?" They were then asked, "Did you see a small black circle appearing before the larger set of back circles?" They were then shown a reduced-speed example of a trial, in which the black cue was clearly visible, and asked, "Did you see anything that looked like this small black circle, appearing before the larger set of circles, during any of the trials?"

Experimental Design and Analysis of Behavioral Data. The task included the following randomly, interleaved trial types. The cue could be located at any of 10 possible grid locations. The target could be one grid location to the left of the cue or one grid location to the right. The target could be tilted toward the left or right. This $10 \times 2 \times 2$ design resulted in 40 trial types. Although all trial types were presented, for purposes of analysis, they were collapsed into two main conditions: The target could be presented either to the left or right of the cue. Trials were not presented in equal proportions. The target was presented to the predicted side of the cue on 85% of trials and to the nonpredicted side on 15% of trials. In all other respects, trial types were counterbalanced. Whether the predicted direction was to the right or left was counterbalanced across subjects. Each subject performed 600 trials, in 10 scanning runs of 60 trials each.

For each subject, we calculated a mean difference of reaction times: ΔRT = (mean reaction time in nonpredicted trials − mean reaction time in predicted trials). Thus, each subject received a single ΔRT score representing the attention effect. A positive score indicated that attention was directed to the predicted side of the cue. The reaction-time analysis included data only from trials in which subjects responded correctly to the target. The pattern of results was not meaningfully changed when all trials were included. A *t* test (two tailed) was performed to determine whether the mean ΔRT among subjects was significantly greater than zero.

Accuracy data (percent correct) is also provided in the *Results for Experiment 1* and *Results for Experiment 2*. However, accuracy was close to ceiling (> 98% average among subjects in task 1 and > 97% in task 2) and therefore proved to be an insensitive measure compared with reaction time, consistent with results from our previous experiments using the same paradigm (8).

MRI Data Collection. Functional imaging data were collected using a Siemens Prisma 3T scanner equipped with a 64-channel head coil. Gradient echo T2*-weighted echo planar images (EPI) with blood oxygen level–dependent (BOLD) contrast were used as an index of brain activity (31). Functional image volumes were composed of 42 near-axial slices with a thickness of 3 mm (with no interslice gap), which ensured that the entire brain excluding the cerebellum, was within the field of view in all subjects (64×64 matrix, 3×3 mm in-plane resolution, TE = 30 ms, and flip angle = 70°). Simultaneous multislice (SMS) acceleration was used (SMS factor = 2). One complete volume was collected every 1,500 ms (TR = 1,500 ms). A total of 2,150 functional volumes were collected for each participant in the main experiment and divided into 10 runs (215 volumes per run). The first five volumes of each run were discarded to account for non–steady-state magnetization. A high-resolution structural image was acquired for each participant at the beginning of the experiment (3D MPRAGE sequence, voxel size = 1 mm isotropic, FOV = 256 mm, 176 slices, TR = 2,300 ms, TE = 2.98 ms, TI = 900 ms, flip angle = 9°, and iPAT GRAPPA = 2). At the end of each scanning session, matching spin echo EPI pairs (anterior to posterior and posterior to anterior) were acquired for blip-up/blip-down field map correction. Subjects completed 10 functional scanning runs, each of 322.5 s (5 min and 22.5 s) duration (consisting of 60 trials). Brief breaks were given between functional scans, although subjects were encouraged to remain as still as possible even during the breaks. Total scanning time for each subject was about 60 min.

MRI Preprocessing. Results included in this manuscript come from preprocessing performed using fMRIPrep version 1.2.3 (32) (RRID: SCR_016216), a Nipype (33) (RRID: SCR_002502) based tool. Each T1-weighted (T1w) volume was corrected for intensity nonuniformity using N4BiasFieldCorrection version 2.2.0 (34) and

skull stripped using antsBrainExtraction.sh version 2.2.0 (using the OASIS template) (35) (RRID: SCR_004757). Spatial normalization to the ICBM 152 Nonlinear Asymmetrical template version 2009c (36) (RRID: SCR_008796) was performed through nonlinear registration with the antsRegistration tool of Advanced Normalization Tools (ANTs) version 2.2.0 using brain-extracted versions of both T1w volume and template. Brain tissue segmentation of cerebrospinal fluid, white matter, and gray matter was performed on the brain-extracted T1w using fast (37) (Functional Magnetic Resonance Imaging of the Brain Software Library [FSL] version 5.0.9, RRID: SCR_002823).

Functional data were slice time corrected using 3dTshift from Analysis of Functional Neuroimages (AFNI) version 16.2.07 (38) (RRID: SCR_005927) and motion corrected using mcflirt (39) (FSL version 5.0.9). This slice time and motion correction was followed by coregistration to the corresponding T1w using boundary-based registration (40) with six degrees of freedom using flirt (FSL). Motion-correcting transformations, BOLD-to-T1w transformation, and T1w-to-template (MNI) warp were concatenated and applied in a single step using antsApplyTransforms (ANTs version 2.2.0) using Lanczos interpolation.

Physiological noise regressors were extracted by applying component-based noise correction (CompCor) (41). Principal components were estimated for the anatomical CompCor variant (aCompCor). A mask to exclude signal with cortical origin was obtained by eroding the brain mask, ensuring it only contained subcortical structures. Six components were calculated within the intersection of the subcortical mask and the union of the cerebrospinal fluid and the white matter masks calculated in T1w space, after their projection to the native space of each functional run. Frame-wise displacement (42) was calculated for each functional run using the implementation of Nipype.

Many internal operations of FMRIPREP use Nilearn (43) (RRID: SCR_001362), principally within the BOLD-processing workflow.

MRI Statistical Analysis. Statistical analyses were performed exclusively in fsaverage6 space (44), which treats the cortex as a two-dimensional sheet. To prepare for individual subjects processing, the first five TRs of each scanning run were discarded in order to allow for signal stabilization. Then, for each subject, runs were submitted to AFNI's 3dDeconvolve (38), in order to construct an event-related general linear model (GLM) at each voxel for each subject. In this model, we included the six components of aCompCor as confound regressors. The events of interest were presentations of the target on trials, in which the target appeared in the location predicted by the cue, and presentations of the target on trials when the target appeared in the nonpredicted location relative to the cue. To analyze the response to the target onset (the crucial event in the trial), the hemodynamic response was modeled as a gamma function triggered at target onset.

After creating this initial, event-related GLM, we used the residual time series as a template for smoothing on the preprocessed functional data. Smoothing was achieved using AFNI's SurfSmooth function. The target smoothness for the output was 6 mm full width at half maximum using the HEAT 7 method (45).

Next, the smoothed datasets were normalized to percent signal change to allow for easier interpretation after data analysis. Finally, these smoothed and normalized datasets were again submitted to 3dDeconvolve to generate an event-based GLM. Note that two GLMs were performed: The first, described in the previous paragraph, was used to generate residuals to perform smoothing; the second, described in the current paragraph, was used for the final statistical analysis (45).

For each subject, we created a contrast for brain activity, subtracting activity on predicted trials from activity on nonpredicted trials. This nonpredicted–predicted contrast was taken for each subject then submitted to AFNI's 3dttest++, for comparison against null, on a group level. The resulting data were then thresholded to a family-wise, error-corrected significance level of $P \leq 0.05$ using cluster correction (within-subjects t test, two tailed) in the following manner. Data were thresholded using an uncorrected threshold of $P = 0.05$. Then, clusters under a critical minimum size were excluded from analysis. As is standard for the present type of surface-based analysis, the minimum cluster size was calculated independently for each hemisphere using AFNI's slow_surf_clustsim, a modern simulation-based method for determining cluster thresholds on surface data. For experiment 1, the resulting critical cluster sizes for each hemisphere were 341 mm^2 for the left hemisphere and 337 mm^2 for the right hemisphere. For experiment 2, the critical cluster sizes for each hemisphere were 344 mm^2 for the left hemisphere and 358 mm^2 for the right hemisphere.

Results for Experiment 1

Posttest Questions. In experiment 1, in which subjects performed task 1, subjects were visually aware of the cue and the target but were never told that the cue predicted the target location. The reason why the paradigm was designed with many possible cue locations distributed across the display screen is that, in prior experiments (8), we found that, with many cue locations, most subjects were unable to notice the statistical relationship between cue and target if not explicitly told about it. The visual impression was of a complicated, flashing, and unpredictable stimulus sequence, in which the specific relationship between cue and target was not obvious. The fact that the cue-target contingencies were statistical and not absolute may have also helped to obscure them. In the posttest questions, all subjects reported that during the experiment trials they had clearly seen the red cue. When asked whether they had noticed any pattern or relationship between the cue and the target, although most noted correctly that the two stimuli were typically near each other, all except one said that they had noticed no other pattern, or they suggested patterns that were not in any way related to the actual pattern (e.g., guessing that the target sometimes appeared above or below the cue). Only one realized that the target was more likely to appear to one side of the cue, as opposed to the other side. When the other subjects were asked explicitly whether they had noticed that the target was more likely to appear to one side of the cue, all indicated they had not noticed and, when prompted, did not know which the more frequent side might be. The one subject who noticed the cue-target relationship was removed from analysis. Of the remaining 18 subjects, therefore, any preferential focusing of attention on the location predicted by the cue was likely to be the result of an implicit process.

Task Performance. Fig. 2*A* shows the mean reaction times for predicted and nonpredicted trials (mean RT for predicted trials = 469.98 ms, SEM = 11.47; mean RT for nonpredicted trials = 482.64 ms, SEM = 12.61; mean accuracy for predicted trials = 98.40%, SEM = 0.34; and mean accuracy for nonpredicted trials = 98.18%, SEM = 0.40). Latency was shorter for predicted trials than for nonpredicted trials, suggesting that attention was shifted to the predicted side of the cue in anticipation of the target. The large error bars in Fig. 2*A* derive from between-subjects variability. To assess the significance of the difference, a within-subjects statistical comparison is needed. We computed a within-subjects difference score, ΔRT, to measure the attention effect (see *Methods* for details). Fig. 2*B* shows that attention was directed to the location predicted by the cue, since the mean ΔRT among subjects was significantly greater than 0 (mean ΔRT = 12.66 ms, SEM = 2.70, two-tailed t test, df = 17, t = 4.70,

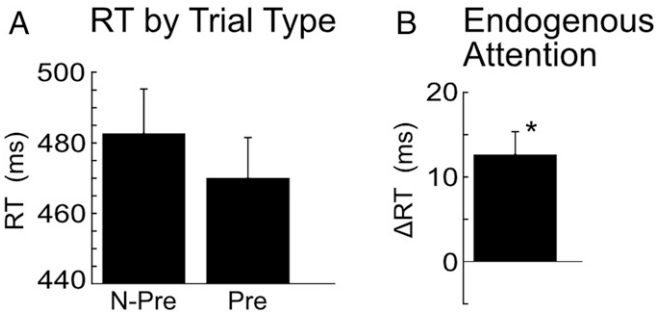

Fig. 2. Behavioral results from experiment 1, in which subjects were aware of the cue but unaware of task contingencies. Data from 18 subjects. Error bars show SE among subjects. (*A*) The *y*-axis shows average reaction time for each target location relative to the cue. Targets were 85% likely to appear to the predicted side of the cue (Pre) and 15% likely to appear to the nonpredicted side of the cue (N-Pre). (*B*) The *y*-axis shows endogenous attention effect: ΔRT = [mean reaction time in nonpredicted trials] – [mean reaction time in predicted trials].

and $P < 0.001$). These results confirm our prior behavioral study (8).

MRI Results. We used a surface-based analysis to find clusters of cortical activity that were significantly greater during nonpredicted trials than during predicted trials (see *Methods*). Fig. 3 shows the results on the inflated surface. Three significant clusters were found in each hemisphere, in roughly mirror-symmetric locations (see Table 1 for coordinates). In the right hemisphere, one cluster was located in the TPJ, mainly in the supramarginal gyrus. A second cluster was located mainly in the precuneus, on the medial aspect of the hemisphere, extending partly onto the lateral aspect into the dorsal parietal lobe. The encroachment onto the lateral aspect is not obvious in the inflated brain image, which slightly distorts relative locations. A third cluster was located in the middle cingulate gyrus, in the medial frontal lobe. In the left hemisphere, one cluster was located in the intraparietal sulcus, mainly dorsal to but also partly overlapping the dorsal TPJ. A second cluster was mainly in the precuneus. A third cluster was in the middle cingulate. Although participants did not explicitly realize that one target location relative to the cue was predicted and another was nonpredicted, these six brain regions were significantly affected by the difference, responding with more activity in nonpredicted trials.

Results for Experiment 2

Posttest Questions. In experiment 2, in which subjects performed task 2, the cue stimulus was black instead of red. Because of that change, the black cue was backward masked by the array of black rings appearing immediately after it, rendering the cue perceptually invisible (8, 11). Many paradigms that seek to manipulate the awareness of a stimulus ask subjects whether they are aware of the stimulus on a trial-by-trial basis using interleaved trials. The downside of this type of measure is that it makes the stimulus task relevant, explicitly telling subjects that a stimulus may be present and that they should note its presence, altering their attention to it and potentially increasing the likelihood that subjects will become aware of it. The success of the present study depended on ensuring that the subjects were truly unaware of the cue in task 2. We chose, therefore, to test a separate set of subjects in task 2 and to leave the subjects uninformed about the cue stimulus until after the experiment so that they would be less likely to become aware of it. Therefore, in the instruction period, subjects were not told about the presence of the masked cue. They were given no explicit knowledge that it existed or that it predicted the location of the target. Only after completing the task, subjects were asked whether they had noticed the cue.

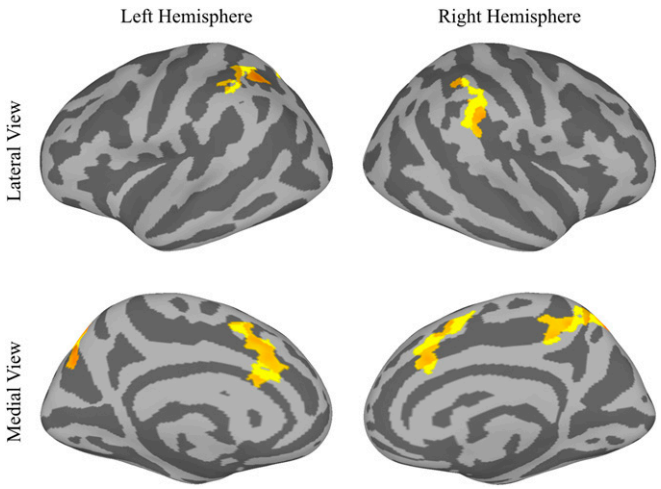

Fig. 3. MRI results in experiment 1 mapped onto the inflated, fsaverage6-standardized brain surface. Highlighted areas pass a $P < 0.05$ threshold, are cluster corrected for multiple comparisons, and show regions with significantly more activity in nonpredicted trials than in predicted trials.

Table 1. Brain activations in experiment 1

Anatomical region	*x*, *y*, *z* (RAI)	Δ β	Cluster size
Right supramarginal gyrus	82, −24, 18	0.070	403
Right precuneus	23, −67, 47	0.122	603
Right midanterior cingulate	10, 57, 24	0.053	447
Left intraparietal sulcus	−58, −48, 41	0.085	437
Left precuneus	−25, −68, 48	0.131	358
Left midanterior cingulate	−12, 38, 43	0.051	447

RAI coordinates are given relative to the fsaverage6 surface and indicate the location of the peak response (the node within the cluster with the largest contrast between nonpredicted and predicted trials). Δ β is the magnitude of the activation (β-value in percent signal change from baseline) for the nonpredicted trials minus the β-value for the predicted trials, for the voxel within a cluster that showed the greatest difference. Cluster size is given in square millimeters.

Many paradigms also test awareness of a stimulus using objective measures such as a forced-choice paradigm. We chose not to use this approach either, partly for the same reason—it would require telling subjects about the cue, increasing their likelihood of becoming aware of it. Moreover, objective measures of awareness do not necessarily address the question of subjective awareness (46). For these reasons, we relied on subjects' subjective reports after completing all trials.

After testing, no subjects reported having seen a black circle, appearing just before the mask, on any of the trials. After being shown a reduced-speed example of a trial, in which the cue was plainly visible, no subjects reported that they had seen anything that looked like the cue at any time during any of the trials. These results suggest that the mask and the procedure of using a separate set of subjects naïve to the cue successfully reduced, and probably eliminated, awareness of the cue. The findings are consistent with prior reports using the same masking technique to eliminate perceptual awareness of a stimulus (8, 11).

Task Performance. Fig. 4*A* shows the mean reaction times for nonpredicted and predicted trials (mean RT for predicted trials = 459.87 ms, SEM = 12.07; mean RT for nonpredicted trials = 460.74 ms, SEM = 12.55; mean accuracy for predicted trials = 97.87%, SEM = 0.39; and mean accuracy for nonpredicted trials = 97.84%, SEM = 0.38). Note that the reaction times are slightly shorter in task 2, without awareness of the cue, than in task 1, with awareness of the cue. The difference lies mainly in the nonpredicted trials, in which a longer reaction time in task 1 is expected. However, a longer latency in task 1 compared with task 2 also appears in the predicted trials, though this difference between tasks is small and not statistically significant (*t* test, $P > 0.05$). In our previous behavioral studies (8), when subjects were aware of an additional stimulus during the trials, even if they believed it to be task irrelevant, it appeared to add a processing cost, resulting in slightly longer reaction times. An overall processing cost of awareness, as subjects engage in an extra cognitive step, is not surprising. Any subtle baseline shift in reaction times between the tasks, however, does not affect the crucial analysis, which depends on a difference between conditions within each task. Whether a subject has long or short reaction times overall, if the subject is faster on predicted than on nonpredicted trials, then the subject has learned to predictively control attention. We computed the difference in reaction time between predicted and nonpredicted target locations. Fig. 4*B* shows that the ΔRT in task 2 was not significantly greater than 0. Thus, there was no evidence that attention was directed to the location

predicted by the cue in task 2 (mean ΔRT = 0.86 ms, SEM = 2.81, two-tailed t test, df = 16, t = 0.31, and $P = 0.762$).

We also directly compared the attention effect (ΔRT) between tasks 1 and 2 in a between-subjects analysis. The attention effect was significantly greater in task 1 than in task 2 (difference = 11.79 ms, SEM = 3.89, between-subjects, two-tailed t test, df = 33, t = 3.033, and $P = 0.0047$).

In an additional, alternative analysis, we analyzed reaction-time data from all individual trials from all subjects using a three-factor ANOVA, with subject as one factor, trial condition (nonpredicted or predicted) as a second within-subjects factor, and awareness status (task 1 versus task 2) as a third between-subjects factor. This analysis confirmed a significant interaction between trial condition and awareness status, reflecting that reaction times were significantly longer in nonpredicted trials than in predicted trials in the aware case, as compared with the unaware case (for the interaction between trial condition and awareness status, $df_1 = 1$, $df_2 = 20209$, F = 10.3227, and $P = 0.0013$).

These findings are consistent with our previous experiments, in which removing visual awareness of the cue significantly impaired the ability of the attention controller to shift or redirect attention with respect to the cue (8).

MRI Results. In experiment 2, in which subjects performed task 2, even though subjects were not aware that the cue predicted the target and were not even aware that any cue had been presented at all, the brain nonetheless processed the cue and the predictive relationship between the cue and the target. Fig. 5 shows the cortical areas where activity was significantly greater during nonpredicted trials than during predicted trials. Two significant clusters were found in the left hemisphere (see Table 2 for coordinates). One cluster was located in the left angular gyrus; the second cluster was located in the left medial prefrontal cortex. No significant activity, however, was found in the right hemisphere. Of particular relevance to our hypotheses, the right TPJ showed no significant activity. When subjects were not aware of the cue, the right TPJ was not significantly active in the contrast between nonpredicted and predicted trials. This result confirms our hypothesis.

It is always possible that some activity was present in the right TPJ, just beneath statistical threshold. To examine that possibility, we defined a region of interest (ROI) around the right TPJ (~2,000 mm^2, based on the definition of the TPJ used in ref. 47). Testing within an ROI reduces the multiple-comparison correction and therefore can be more sensitive to weaker activations

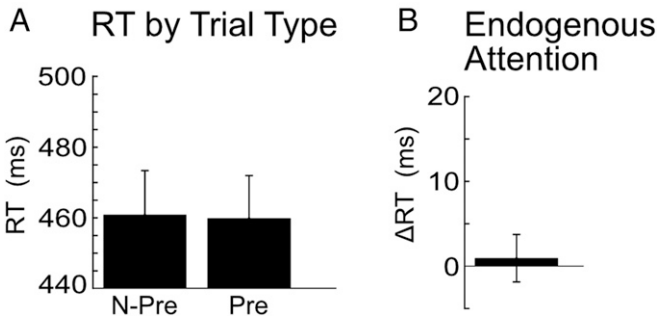

Fig. 4. Behavioral results from experiment 2, in which subjects were unaware of the cue. Data from 17 subjects. Error bars show SE among subjects. (*A*) The *y*-axis shows average reaction time for each target location relative to the cue. Targets were 85% likely to appear to the predicted side of the cue (Pre) and 15% likely to appear to the nonpredicted side of the cue (N-Pre). (*B*) The *y*-axis shows endogenous attention effect: ΔRT = (mean reaction time in nonpredicted trials) − (mean reaction time in predicted trials).

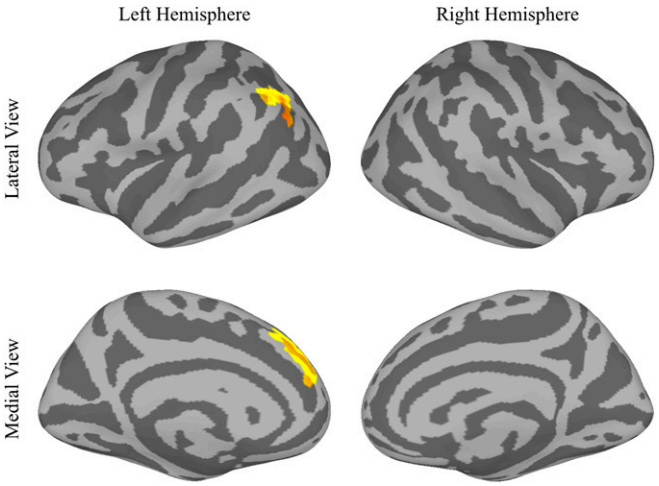

Fig. 5. MRI results in experiment 2 mapped onto the inflated, fsaverage6-standardized brain surface. Highlighted areas pass a $P < 0.05$ threshold cluster corrected for multiple comparisons and show regions with significantly more activity in nonpredicted trials than in predicted trials.

(for small volume corrections, see ref. 48). Yet even within this ROI, no clusters of activity related to the nonpredicted versus predicted contrast were obtained that passed a $P < 0.05$ significance. We did not even obtain any clusters in the ROI that passed a lax, $P < 0.10$ threshold. We saw no evidence of any right TPJ activity associated with the nonpredicted versus predicted contrast, not even weak or subthreshold activity.

Finally, we used a between-subjects analysis to compare the right TPJ activity (in the nonpredicted versus predicted contrast) between tasks 1 and 2. In the ROI around the right TPJ, we found a region in the supramarginal gyrus in which activity was significantly greater in task 1 than in task 2 ([nonpredicted − predicted for experiment 1] − [nonpredicted − predicted for experiment 2], significant cluster of 211 mm^2 that passed a cluster-corrected, $P < 0.05$ level of significance). Thus, in task 1, in which awareness of the cue allowed for the endogenous control of attention, the right TPJ showed significant activity in association with the use of that cue to control attention; in task 2, in which a lack of awareness of the cue led to a lack of endogenous attention control, the activity in the right TPJ was reduced and no longer detected, and the signal in the right TPJ was significantly greater in task 1 than in task 2.

Discussion

Previous behavioral experiments (8) provided two related findings that served as the basis for the present study. First, when people were aware of a visual cue, they implicitly used the cue to predictively guide their attention to a target. Second, in

Table 2. Brain activations in experiment 2

Anatomical region	*x*, *y*, *z* (RAI)	Δ β	Cluster size
Left angular gyrus	−41, −68, 44	0.060	583
Left superior frontal gyrus	−7, 37, 44	0.112	449

RAI coordinates are given relative to the fsaverage6 surface and indicate the location of the peak response (the node within the cluster with the largest contrast between nonpredicted and predicted trials). Δ β is the magnitude of the activation (β-value in percent signal change from baseline) for the nonpredicted trials minus the β-value for the predicted trials, for the voxel within a cluster that showed the greatest difference. Cluster size is given in square millimeters.

an otherwise similar paradigm, when people were not aware of the cue, they no longer engaged in the same implicit process of predictively guiding attention. In the present study, subjects performed the two tasks while brain activity was measured in an MRI scanner. We argued that any brain area that showed activity in the first task and not the second might participate in constructing that awareness-dependent, predictive model and using it to endogenously guide attention.

We did not look for a greater overall activity in task 1 than in task 2, which would have represented a nonspecific result subject to changes in baseline. Instead, we examined a contrast between two specific conditions. In predicted trials, the cue predicted a target location, and the target then appeared at that location. In nonpredicted trials, the cue predicted a target location, and the target then appeared elsewhere. Nothing distinguished the two trial types other than the predictive relationship between cue and target. Any brain area that responds differently to these two conditions must necessarily be sensitive to that predictive relationship between the cue and the target. Such a brain area reflects the brain's learning of the predictive information.

We hypothesized that the right TPJ, specifically, would show this pattern of response in task 1 and not in task 2. The hypothesis was confirmed. The right TPJ was active in task 1, in association with the cue-target relationship, and not measurably so in task 2. A direct comparison between the two tasks within the right TPJ showed a significant difference.

In addition to the activity found in the right TPJ, we also obtained a similar pattern of activity (present in task 1 and absent in task 2) in other cortical areas, including the precuneus and the medial frontal cortex. These areas—TPJ, precuneus, and medial frontal cortex—have been described before as part of a social cognitive, theory-of-mind network (13, 15–17). We previously suggested that the theory-of-mind network, in which the right TPJ is a central node, may contribute to constructing models of attention, both the attention of other people and one's own attention (1). Although in the present study we focused our predictions on the right TPJ, arguing that it should be a hotspot involved in building predictive models for the control of attention, we recognize that the effects we see here are reflected in a network of related areas, not one area.

In the following, we consider five possible alternative interpretations of the findings and then explain our own interpretation.

One possible explanation for the null result in the TPJ, in task 2, is that, because subjects were unaware of the cue, the cue was so reduced in stimulus efficacy that it affected no area of the brain and thus also did not affect the right TPJ. After all, if subjects are not aware of something, how can it impact the brain? The cue might as well not be there. However, this explanation is ruled out by the data. First, we know from prior data (8) that the black cue, though outside awareness, affects subjects by drawing exogenous attention to itself. In that prior study, we compared reaction times, when the target appeared at the same location as the cue versus when the target was displaced from the cue, and found a robust, repeatable difference. There is, therefore, no doubt that the black cue affects the subjects. Second, in the present data during task 2, the cue measurably affected the brain. We found two areas, the left angular gyrus and the left medial prefrontal cortex, that showed significantly more activity in the nonpredicted trials than in the predicted trials. Therefore, even though the subjects were not aware of the cue and did not use the cue to guide endogenous attention, these two brain areas still processed the predictive information that the cue provided about the target. At the same time, we found no evidence that the right TPJ was involved in that processing in task 2. We do not know why, without awareness, these two left hemisphere areas might have begun to process the relationship between the cue and target. Perhaps, when one mechanism shuts down—a mechanism associated with awareness of the cue and activity in the TPJ—a different mechanism begins to take over the processing of the cue's environmental contingencies. Hopefully, future experiments can explore this unexpected, nonconscious processing. In any case, the cue in task 2 did measurably affect subjects; it just did not affect endogenous attention control or the TPJ.

A second alternative explanation of the results is that eye movements were systematically different between nonpredicted and predicted trials, leading to a false signal. Although we did not measure eye movement in the present study, we believe this explanation is unlikely. In our prior behavioral study on these tasks (8), we did measure eye position and found no systematic differences between conditions or tasks, as subjects tended to fixate on the center of the display during each trial. Eye movement and eye position could not explain the effect on attention.

A third alternative explanation is that the right TPJ responds to surprise. In task 1, perhaps the TPJ became more active to nonpredicted trials than to predicted trials because subjects were surprised when the target appeared at a nonpredicted location. In this interpretation, in task 2, without awareness of the cue, subjects experienced no surprise on nonpredicted trials, and thus, the activity pattern disappeared. The right TPJ is indeed known to become active in association with unexpected, surprising, and statistically uncommon events (20–28). However, this explanation is ruled out in the present study. Even when subjects were aware of the cue (in task 1), they did not explicitly notice the trial statistics and therefore never realized that anything surprising or unexpected might have happened on nonpredicted trials. Thus, the present findings cannot be explained as a response to explicit surprise. Could the TPJ respond to implicitly processed, rare events, even when the subjects do not explicitly notice or experience surprise? If that were the case, then we would have seen elevated activity in task 2 during nonpredicted trials. Nonpredicted trials were just as rare in task 2 as in task 1. Moreover, in task 2, at least some brain areas showed evidence of processing the statistical relationship between the cue and target. Thus, at some implicit level, subjects knew about the difference between the more frequent predicted trials and the less frequent nonpredicted trials. Yet in task 2, the TPJ showed no evidence of responding to the less frequent trial type. For these reasons, neither explicit surprise nor the implicit coding of rarity can explain the right TPJ pattern of results.

A fourth possibility is that the TPJ generally monitors predictive events in the environment. In that interpretation, unrelated to attentional control, surprise, or awareness, the TPJ encoded the predictive relationship between cue and target. This possibility is ruled out for the same reason as in explanation three. The predictive relationship between the cue and the target was present in task 2, and at least some areas of the brain were sensitive to those predictive statistics, showing that at some level the subjects processed them. Yet the right TPJ did not respond to those predictive statistics.

A fifth alternate explanation is that the TPJ activity is related to exogenous shifts of attention. One of the most common and influential suggestions about the right TPJ is that it is involved in directing exogenous attention, as part of the ventral attention network (20–27). Of the explanations thus far, this possible explanation comes closest to accounting for the present pattern of results. We argue, however, that it is still not adequate. In task 1, on predicted trials, the likely sequence of events is as follows: Attention is exogenously drawn to the cue; attention is then endogenously shifted to the predicted location; the target appears at that location; there is probably a further rise in attention at target onset; and finally, the subject responds to the target. On nonpredicted trials, the likely sequence is the following: Attention is exogenously drawn to the cue; attention is endogenously shifted to the predicted location; the target appears at a different location; attention is exogenously drawn to the new location; and then, the subject responds to the target. Thus, the greater activity in the TPJ

during nonpredicted trials, in task 1, could be caused by the additional, final spatial shift of exogenous attention to acquire the target. Findings like these, over many experiments in varied forms (20–27), have led to the hypothesis that the right TPJ helps control exogenous attention.

We suggest, however, that this explanation, that the TPJ becomes active in association with exogenous attention or that it is part of a controller for exogenous attention, does not explain the results. First, on theoretical grounds, we suggest that exogenous attention does not require a dedicated controller. Exogenous attention is a bottom-up, externally imposed shift of attention. Only endogenous attention requires a dedicated control system. Second, the right TPJ is not active in association with exogenous shifts of attention per se. If it were, it would respond to any stimulus that appears, given the inevitable exogenous attention drawn to the sharp onset of a visual stimulus. The TPJ would be as responsive to the onset of visual stimuli as the primary visual cortex is responsive to visual stimuli. Such visual responsiveness is not reported in the literature. For example, in the present study, we know that the initial onset of the cue robustly draws exogenous attention (see ref. 8 for the experimental conditions that established that shift in exogenous attention). Whether the cue is red and visible (task 1) or black and outside of awareness (task 2), it draws exogenous attention at onset. If the right TPJ responds to a shift in exogenous attention per se, then it should respond to the onset of the cue, averaged across all trial conditions. We performed that analyses on the present data, for each task, and found no significant right TPJ activity in response to the onset of the cue (no significant clusters within the right TPJ ROI at $P < 0.05$ threshold). We are aware of no evidence that the right TPJ responds consistently when attention is shifted exogenously. The reason why the exogenous attention hypothesis has become so commonly accepted for the right TPJ is that, in a range of experiments, including in the present task 1, the right TPJ becomes active during shifts of exogenous attention in the case that something unexpected, nonpredicted, or rare occurs (20–27). Not all of these experiments involve a spatial shift of attention. Some may involve other sudden, or unexpected, exogenously induced changes in attentional status. For example, in the oddball task, subjects view a series of centrally placed, identical stimuli, and then an oddball stimulus appears, breaking the pattern, presumably causing an increase in attention at the same location, and triggering right TPJ activity (26, 27).

We are therefore faced with an interpretational conundrum. The right TPJ activity found in the present study is not triggered by unexpected, surprising, or rare events by themselves. It is not triggered by learning predictive stimulus relationships by themselves. It is not triggered by exogenous shifts of attention by themselves. Some more complex combination of conditions is needed to explain the pattern of results. What makes the TPJ respond more on nonpredicted trials than on predicted trials but only in task 1 and not in task 2? We suggest that there remains one factor that varies across task conditions in a manner to adequately explain the pattern of results. In task 1, on nonpredicted trials, the attention control system makes a prediction (it shifts attention predictively), and the prediction turns out to be wrong, whereas on predicted trials, the attention control system makes the same prediction, and the prediction turns out to be right. In task 2, according to the behavioral data, the attention control system makes no such predictions. It does not make right or wrong predictions on either trial type. We suggest, therefore, that a relative rise in activity in the TPJ occurs specifically when a prediction related to attention turns out to be wrong. Why should a violated prediction lead to TPJ activity? When a prediction is wrong, we suggest that the error causes the underlying predictive model to be incrementally updated or at least reevaluated, and the TPJ becomes briefly more active during that process. In this suggestion, it is the error correction to the underlying predictive model of attention that drives TPJ activity. This explanation accounts for the pattern of data not only in the present study but also across previous studies that find activity in the right TPJ associated with changes in attention. In our proposed interpretation, the ventral attention system, including the right TPJ, is not a controller of exogenous attention. Instead, it is building or adjusting a model of attention. It is monitoring and making predictions about both exogenous and endogenous attention. Sometimes, an unexpected or rare event causes a shift of attention that does not fit the predictive model. Then, the right TPJ has a rise in activity. But an exogenous shift of attention, in the present interpretation, will not, by itself, drive TPJ activity, nor will an unexpected event that does not impact exogenous attention. The key event that causes TPJ activity is the momentary violation of a continuously computed, predictive model used for the control of attention.

In the present study, how are the results related to awareness? The key difference between task 1 and 2 is that subjects are aware of the cue in task 1 and not in task 2. The behavioral data show that awareness of the cue permits the predictive, endogenous control of attention; lack of awareness of the cue prevents the predictive, endogenous control of attention. Thus, the behavioral results, and the MRI results in the TPJ, are modulated by awareness of the cue. If AST is right, then that modulation has a simple explanation: Awareness is the model of attention. Awareness of the cue corresponds to a state in which the brain is modeling the attention that has been drawn to the cue.

We acknowledge that other interpretations may be possible. Moreover, it is unlikely that the right TPJ is limited to the functions suggested here, since it is a complex brain area with many subregions, within which a great range of other tasks have been found to evoke activity, including autobiographical memory tasks and social cognition tasks (16, 49–52). In addition, the processes described here presumably depend on a network of areas of which the TPJ is only one node. Nonetheless, these findings represent a step toward understanding how attention and awareness interrelate in the brain.

Data Availability. Brain Imaging Data Structure format fMRI data have been deposited in Princeton DataSpace (DOI: 10.34770/9425-b553) (53). All study data are included in the main text.

ACKNOWLEDGMENTS. This work was supported by the Princeton Neuroscience Institute Innovation Fund. A.G. was supported by the Wenner-Gren Foundation, the Swedish Society of Medicine, and the Foundation Blanceflor. We thank Argos Wilterson for his enthusiastic encouragement.

1. M. S. A. Graziano, S. Kastner, Human consciousness and its relationship to social neuroscience: A novel hypothesis. *Cogn. Neurosci.* **2**, 98–113 (2011).
2. M. S. A. Graziano, *Consciousness and the Social Brain* (Oxford University Press, Oxford, UK, 2013).
3. M. S. A. Graziano, Consciousness and the attention schema: Why it has to be right. *Cogn. Neuropsychol.* **37**, 224–233 (2020).
4. M. S. A. Graziano, T. W. Webb, The attention schema theory: A mechanistic account of subjective awareness. *Front. Psychol.* **6**, 500 (2015).
5. M. S. A. Graziano, M. M. Botvinick, "How the brain represents the body: Insights from neurophysiology and psychology" in *Common Mechanisms in Perception and Action: Attention and Performance XIX*, W. Prinz, B. Hommel, Eds. (Oxford University Press, Oxford, UK, 2002), pp. 136–157.
6. R. Shadmehr, F. A. Mussa-Ivaldi, Adaptive representation of dynamics during learning of a motor task. *J. Neurosci.* **14**, 3208–3224 (1994).
7. K. A. Thoroughman, J. A. Taylor, Rapid reshaping of human motor generalization. *J. Neurosci.* **25**, 8948–8953 (2005).
8. A. I. Wilterson *et al.*, Attention control and the attention schema theory of consciousness. *Prog. Neurobiol.* **195**, 101844 (2020).
9. T. W. Webb, H. H. Kean, M. S. A. Graziano, Effects of awareness on the control of attention. *J. Cogn. Neurosci.* **28**, 842–851 (2016).

10. Y. Tsushima, Y. Sasaki, T. Watanabe, Greater disruption due to failure of inhibitory control on an ambiguous distractor. *Science* **314**, 1786–1788 (2006).
11. Z. Lin, S. O. Murray, More power to the unconscious: Conscious, but not unconscious, exogenous attention requires location variation. *Psychol. Sci.* **26**, 221–230 (2015).
12. A. Guterstam, A. I. Wilterson, D. Wachtell, M. S. A. Graziano, Other people's gaze encoded as implied motion in the human brain. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 13162–13167 (2020).
13. Y. T. Kelly, T. W. Webb, J. D. Meier, M. J. Arcaro, M. S. A. Graziano, Attributing awareness to oneself and to others. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 5012–5017 (2014).
14. T. W. Webb, K. M. Igelström, A. Schurger, M. S. A. Graziano, Cortical networks involved in visual awareness independent of visual attention. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 13923–13928 (2016).
15. H. L. Gallagher *et al*., Reading the mind in cartoons and stories: An fMRI study of 'theory of mind' in verbal and nonverbal tasks. *Neuropsychologia* **38**, 11–21 (2000).
16. R. Saxe, N. Kanwisher, People thinking about thinking people. The role of the temporo-parietal junction in "theory of mind". *Neuroimage* **19**, 1835–1842 (2003).
17. S. J. van Veluw, S. A. Chance, Differentiating between self and others: An ALE meta-analysis of fMRI studies of self-recognition and theory of mind. *Brain Imaging Behav.* **8**, 24–38 (2014).
18. G. Vallar, D. Perani, The anatomy of unilateral neglect after right-hemisphere stroke lesions. A clinical/CT-scan correlation study in man. *Neuropsychologia* **24**, 609–622 (1986).
19. V. Verdon, S. Schwartz, K.-O. Lovblad, C.-A. Hauert, P. Vuilleumier, Neuroanatomy of hemispatial neglect and its functional components: A study using voxel-based lesion-symptom mapping. *Brain* **133**, 880–894 (2010).
20. M. Corbetta, J. M. Kincade, G. L. Shulman, Neural systems for visual orienting and their relationships to spatial working memory. *J. Cogn. Neurosci.* **14**, 508–523 (2002).
21. M. Corbetta, G. Patel, G. L. Shulman, The reorienting system of the human brain: From environment to theory of mind. *Neuron* **58**, 306–324 (2008).
22. J. P. Mitchell, Activity in right temporo-parietal junction is not selective for theory-of-mind. *Cereb. Cortex* **18**, 262–271 (2008).
23. G. L. Shulman *et al*., Interaction of stimulus-driven reorienting and expectation in ventral and dorsal frontoparietal and basal ganglia-cortical networks. *J. Neurosci.* **29**, 4392–4407 (2009).
24. G. L. Shulman *et al*., Right hemisphere dominance during spatial selective attention and target detection occurs outside the dorsal frontoparietal network. *J. Neurosci.* **30**, 3640–3651 (2010).
25. J. T. Serences *et al*., Coordination of voluntary and stimulus-driven attentional control in human cortex. *Psychol. Sci.* **16**, 114–122 (2005).
26. A. A. Stevens, P. Skudlarski, J. C. Gatenby, J. C. Gore, Event-related fMRI of auditory and visual oddball tasks. *Magn. Reson. Imaging* **18**, 495–502 (2000).
27. H. Kim, Involvement of the dorsal and ventral attention networks in oddball stimulus processing: A meta-analysis. *Hum. Brain Mapp.* **35**, 2265–2284 (2014).
28. J. J. Geng, S. Vossel, Re-evaluating the role of TPJ in attentional control: Contextual updating? *Neurosci. Biobehav. Rev.* **37**, 2608–2620 (2013).
29. D. H. Brainard, The psychophysics toolbox. *Spat. Vis.* **10**, 433–436 (1997).
30. M. Kleiner *et al*., What's new in Psychtoolbox-3? *Perception* **36**, 1–16 (2007).
31. N. K. Logothetis, J. Pauls, M. Augath, T. Trinath, A. Oeltermann, Neurophysiological investigation of the basis of the fMRI signal. *Nature* **412**, 150–157 (2001).
32. O. Esteban *et al*., fMRIPrep: A robust preprocessing pipeline for functional MRI. *Nat. Methods* **16**, 111–116 (2019).
33. K. Gorgolewski *et al*., Nipype: A flexible, lightweight and extensible neuroimaging data processing framework in Python. *Front. Neuroinform.* **5**, 13 (2011).
34. N. J. Tustison *et al*., N4ITK: Improved N3 bias correction. *IEEE Trans. Med. Imaging* **29**, 1310–1320 (2010).
35. B. B. Avants, C. L. Epstein, M. Grossman, J. C. Gee, Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain. *Med. Image Anal.* **12**, 26–41 (2008).
36. V. S. Fonov, A. C. Evans, R. C. McKinstry, C. R. Almli, D. L. Collins, Unbiased nonlinear average age-appropriate brain templates from birth to adulthood. *Neuroimage* **47**, S102 (2009).
37. Y. Zhang, M. Brady, S. Smith, Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE Trans. Med. Imaging* **20**, 45–57 (2001).
38. R. W. Cox, AFNI: Software for analysis and visualization of functional magnetic resonance neuroimages. *Comput. Biomed. Res.* **29**, 162–173 (1996).
39. M. Jenkinson, P. Bannister, M. Brady, S. Smith, Improved optimization for the robust and accurate linear registration and motion correction of brain images. *Neuroimage* **17**, 825–841 (2002).
40. D. N. Greve, B. Fischl, Accurate and robust brain image alignment using boundary-based registration. *Neuroimage* **48**, 63–72 (2009).
41. Y. Behzadi, K. Restom, J. Liau, T. T. Liu, A component based noise correction method (CompCor) for BOLD and perfusion based fMRI. *Neuroimage* **37**, 90–101 (2007).
42. J. D. Power *et al*., Methods to detect, characterize, and remove motion artifact in resting state fMRI. *Neuroimage* **84**, 320–341 (2014).
43. A. Abraham *et al*., Machine learning for neuroimaging with scikit-learn. *Front. Neuroinform.* **8**, 14 (2014).
44. B. Fischl, M. I. Sereno, R. B. Tootell, A. M. Dale, High-resolution intersubject averaging and a coordinate system for the cortical surface. *Hum. Brain Mapp.* **8**, 272–284 (1999).
45. M. K. Chung *et al*., Cortical thickness analysis in autism with heat kernel smoothing. *Neuroimage* **25**, 1256–1265 (2005).
46. P. M. Merikle, D. Smilek, J. D. Eastwood, Perception without awareness: Perspectives from cognitive psychology. *Cognition* **79**, 115–134 (2001).
47. K. M. Igelström, T. W. Webb, M. S. A. Graziano, Neural processes in the human temporoparietal cortex separated by localized independent component analysis. *J. Neurosci.* **35**, 9432–9445 (2015).
48. R. A. Poldrack, J. A. Mumford, T. E. Nichols, *Handbook of Functional MRI Data Analysis* (Cambridge University Press, 2011).
49. K. M. Igelström, T. W. Webb, Y. T. Kelly, M. S. A. Graziano, Topographical organization of attentional, social and memory processes in the human temporoparietal cortex. *eNeuro* **3**, ENEURO.0060-16.2016 (2016).
50. R. Cabeza, E. Ciaramelli, M. Moscovitch, Cognitive contributions of the ventral parietal cortex: An integrative theoretical account. *Trends Cogn. Sci.* **16**, 338–352 (2012).
51. R. M. Carter, S. A. Huettel, A nexus model of the temporal-parietal junction. *Trends Cogn. Sci.* **17**, 328–336 (2013).
52. S. Konishi, M. E. Wheeler, D. I. Donaldson, R. L. Buckner, Neural correlates of episodic retrieval success. *Neuroimage* **12**, 276–286 (2000).
53. A. I. Wilterson, S. A. Nastase, B. J. Bio, A. Guterstam, M. S. A. Graziano, Attention and awareness in the dorsal attention network. Princeton DataSpace. http://arks.princeton.edu/ark:/88435/dsp01xp68kk27p. Deposited 20 November 2020.