

Original Research Article

Attention control and the attention schema theory of consciousness[☆]

Andrew I. Wilterson^a, Casey M. Kemper^a, Noah Kim^a, Taylor W. Webb^b,
Alexandra M.W. Reblando^c, Michael S.A. Graziano^{a,*}

^a Department of Psychology, Princeton University, Princeton, NJ 08544, United States

^b Department of Psychology, University of California, Los Angeles, CA 90095, United States

^c Department of Psychology, Columbia University, New York, NY 10032, United States



ARTICLE INFO

Keywords:

Consciousness

Awareness

Exogenous attention

Endogenous attention

Internal model

ABSTRACT

In the attention schema theory (AST), the brain constructs a schematic, simplified model of attention. The model is associated with three cognitive processes: a model of one's own attention contributes to the endogenous control of attention, a model of the attention of others contributes to theory of mind, and the contents of these models leads to the common human claim that we contain a non-physical consciousness or awareness inside us. Because AST is a control-engineering style theory, it can make specific predictions in complex situations. Here, over six experiments, we examined interactions between attention and awareness to test predictions of AST. Participants performed a visual task in which a cue stimulus affected their attention, as measured by their reactions to a subsequent target stimulus. The task measured both exogenous attention drawn to the cue and endogenous attention directed to a target location predicted by the cue. When participants were not aware that the cue predicted the target, both exogenous and endogenous attention effects remained. In contrast, when participants were not visually aware of the cue itself, the exogenous attention effect remained and the endogenous effect was impaired. In an additional two experiments, when participants learned an implicit shift of attention, the learning generalized from trained spatial locations to adjacent, untrained locations. Each of these findings matched predictions of AST. The results support the interpretation that attention control relies partly on an internal model that is responsible for claims of awareness.

1. Introduction

The attention schema theory (AST) is a proposed explanation for how people claim to have a subjective consciousness (Graziano, 2013, 2019; Graziano and Kastner, 2011; Webb and Graziano, 2015). It proposes that three different phenomena – the control of attention, some aspects of social cognition, and the claim of subjective consciousness – are linked by one mechanism.

First, in the theory, attention – the manner in which processing resources are selectively focused – is controlled partly with the help of an internal model of attention (Webb and Graziano, 2015). That model is a representation, or a rich set of information, about attention itself. It includes information about the current state of attention and predictive information. Just as the brain constructs an arm schema to describe, predict, and thus help control the arm as it reaches toward objects, so the brain constructs an attention schema to describe, predict, and thus help control attention as it is directed toward items. In that analogy, endogenous or internally directed attention is like the motor system

controlling the arm – requiring an internal model of the arm – whereas exogenous or stimulus-driven attention is like an external object pushing the arm, a movement that does not require an internal control model. The model updates to keep track of ongoing changes, whether those changes are externally or internally induced; but the usefulness of the model is primarily for endogenous control.

Second, in the theory, an attention schema can also be constructed to model the attentional states of other people, thus helping to predict their behavior (Graziano, 2013, 2019; Guterstam et al., 2018; Kelly et al., 2014; Pesquita et al., 2016). Suppose you wish to predict John's behavior. The more John attends to item X, the more likely he is to react to X or remember X to guide future behavior. Thus, a model of attention – of what it is, what it does, what its typical dynamics are, and what specific state John's attention is in – would be useful for predicting his behavior. In that perspective, an attention schema is a component of theory of mind.

Third, in AST, the brain's model of attention supplies the information on the basis of which people claim to have consciousness (here also

[☆] This article is part of a Special Issue in honor of Prof. Charles G. Gross.

* Corresponding author.

E-mail address: graziano@princeton.edu (M.S.A. Graziano).

called awareness) (Graziano, 2013, 2019; Graziano and Kastner, 2011; Webb and Graziano, 2015). Attention and consciousness have suspiciously similar attributes. Attention is an emergent property in which the brain's processing resources "grasp" a specific set of items, enhancing the relevant signals and enabling the brain to make decisions and direct actions toward the attended items. Consciousness is typically seen as a subjective experience with the same properties noted above for attention. Consciousness "grasps" specific items, bringing them to some level of understanding and vividness, and empowering us to make decisions and direct actions toward those items. One major difference between attention and consciousness is that attention is a physical process of interacting neurons and electrochemical signals, whereas people report consciousness from the first-person perspective as an insubstantial, internal essence, a "feeling." In the proposal, when people report the presence of consciousness, the report is based on specific information in the brain, and the information derives from an imperfect model of attention. Consciousness of object X corresponds to the specific condition in which a person is processing object X, at least some of the person's attention is on X, and the attention schema depicts the attention that is on X. Consciousness, in this account, is literally what the brain makes of its own attention from the inside. Hence the similarity, but not-quite-perfect match, between attention and consciousness. The attention schema depicts the main dynamics of attention while leaving out any description of the underlying physical mechanism. On the basis of that incomplete and schematic information, people believe they have a physically insubstantial, private essence, a subjective experience, that takes possession of items. The theory addresses the problem of how an information-processing machine constructs a self-model that inaccurately describes an intangible conscious experience. On the basis of that self-model, the machine "thinks" and "claims" it has an intangible conscious experience. This approach to consciousness has been described in more detail in previous publications (Graziano, 2013, 2019; Graziano and Kastner, 2011; Webb and Graziano, 2015) and has similarities to other philosophical approaches to consciousness (e.g. Dennett, 1991; Frankish, 2016; Metzinger, 2009; for commonalities between AST and some other modern theories of consciousness, see Graziano et al., 2019).

A growing set of experiments suggests how AST may relate to functional networks in the brain. For example, specific cortical networks involved in theory of mind are active when people judge the awareness or attention of others (Kelly et al., 2014). Many of those same brain regions, especially in the temporoparietal junction, are also associated with self-reported awareness, suggesting that attributing consciousness to oneself and to others may depend on partially overlapping, theory-of-mind systems (Igelström et al., 2016; Kelly et al., 2014; Webb et al., 2016a,2016b). The temporoparietal junction, a hub of many overlapping networks, has become a topic of controversy because, on the one hand, it is associated with awareness, hemispatial neglect, and control of attention, but on the other hand it is associated with social cognition and theory of mind (Corbetta et al., 2000; Igelström and Graziano, 2017; Mitchell, 2008; Saxe and Kanwisher, 2003; Shulman et al., 2010; Vallar and Perani, 1986). AST makes sense of that partial overlap of properties, functionally linking one's own self-reported awareness to the control of attention and to social cognition.

Arguably the most direct way to test AST, however, is not by localizing where its components are represented in the brain, but instead by probing the specific, working relationship between awareness and attention, the heart of the theory. In AST, awareness and attention are not the same thing, but they are related in a specific, complex, and testable manner. Just as the arm schema is a detail-poor model of the arm that can sometimes make mistakes and become dissociated from the arm, so awareness is a detail-poor model of attention that can sometimes become dissociated from attention. When the arm schema is compromised – for example, if it becomes spatially misaligned from the arm, or if the arm schema disappears entirely in the case of a stroke – the control of the arm is impaired. When awareness becomes

misaligned from attention – for example, when you are attending to a stimulus but not aware of it – the control system should become impaired at regulating or adjusting the attention to that stimulus. It is now well established that attention to a stimulus can indeed occur without subjective awareness of the stimulus (e.g. Ansorge and Heumann, 2006; Hsieh et al., 2011; Jiang et al., 2006; Kentridge et al., 1999, 2004; Lambert et al., 1999; Lin and Murray, 2015; McCormick, 1997; Travers et al., 2018; Tsushima et al., 2006; Webb et al., 2016a,2016b). That condition of attention without awareness provides a basis for testing predictions of AST. Thus far, the evidence tends to support AST: without awareness of a stimulus, although attention to the stimulus may remain, the endogenous control of that attention is impaired. Here we briefly summarize three previous examples of this type of relationship that we view as supportive of AST, before describing the current set of experiments.

First, one of the most common and important control tasks that the attention system faces is to minimize attention to a distractor. Think of casting small amounts of attention to minor distractors around the room, such as an annoying mosquito, while focusing attention mainly on a conversation. One study suggests that when people are not aware of a distracting stimulus, *more* of their attention leaks toward that distractor and away from the primary target of the task, whereas when people are aware of the distractor, they are better able to minimize attention to it (Tsushima et al., 2006). This finding is inconsistent with some common assumptions about attention and awareness. For example, in one intuitive view, if enough attention is directed at a stimulus, then the stimulus is boosted or emphasized enough to enter awareness. Awareness, in that view, acts essentially as a heightened state of attention. But as intuitive as that view may be, it cannot explain the finding that awareness is sometimes associated with *less* rather than more attention. Instead, the finding is consistent with the control-theory interpretation of AST. In that interpretation, the brain builds a model of the attention that is focused on X, which we report as an awareness of X. A gap in the model – a failure to model the attention that is being siphoned away to a distractor – leads to two outcomes. First, we report being unaware of any distracting stimulus. Second, we show a poor ability to regulate and minimize the attention to the distractor. The controller cannot regulate the leak if it does not know about the leak.

A second study examined participants' ability to adapt the spatial distribution of attention to new environmental statistics. That ability was compromised in the absence of awareness of the relevant stimuli (Lin and Murray, 2015). Again, this finding is consistent with the control-theory interpretation. One of the main benefits of an internal model is that it can be efficiently adapted to new environmental statistics. When the internal model is compromised, adaptation is reduced.

In a third study, without awareness of a stimulus, attention drawn to the stimulus was not overall smaller or larger in magnitude, but showed greater fluctuations over time, possibly reflecting a reduction in endogenous control and stabilization of attention (Webb et al., 2016a,2016b).

Findings like the three noted above suggest that awareness and attention are not the same thing, but neither are they two entirely independent processes. Instead, awareness helps in the control and regulation of attention. Without awareness of a stimulus, attention to that stimulus is often still present, but is less well controlled in at least the three ways noted above. When it is useful to minimize the attention drawn to a stimulus, you are less able to do so; when it is useful to shift attention away from a stimulus to a new spatial location, you are less able to do so; and when it is useful to stabilize attention on a stimulus for an extended time interval, you are less able to do so. This loss of control in the absence of awareness supports AST, in which compromising awareness of a stimulus is the same as compromising an internal model of the attention that is focused on the stimulus, which in turn compromises the endogenous control over that focus of attention.

The purpose of the present set of six experiments was to

systematically test how awareness affects both exogenous and endogenous attention. We used a cued attention paradigm that allowed for measurements of both exogenous and endogenous aspects of visual-spatial attention in a single task. On each trial, a visual cue first appeared. Then, half a second later, a target stimulus appeared, to which the participants responded. The target could appear at the same location as the cue, or offset by 3.5° to the right or left of the cue. On most trials (70 %), the target was offset in one direction. Thus, the cue statistically predicted the location of the target. Participants showed two attentional effects. First, reaction times were shorter when the target appeared at the same location as the cue versus a location offset from the cue, indicating a simple exogenous effect in which the cue automatically drew attention to its own location relative to other locations. Second, reaction times were reduced when the target appeared at the predicted location offset in one direction from the cue, as compared to the opposite, non-predicted, offset location. This second effect suggested a type of endogenous attention control; an internal controller of attention was able to take advantage of the cue-target relationship to adjust or shift attention toward the predicted location. The paradigm therefore offered a tool for measuring an exogenous influence on attention and at least one form of endogenous control of attention in the same task. We tested how these two specific measures of attention were affected by awareness. First, we manipulated whether participants were aware of the cue-target contingencies. Second, we manipulated whether participants were visually aware of the cue.

1.1. Using AST to make predictions about attention and awareness

By considering how the present task interacts with the logic of AST, it is possible to construct a set of predictions. The predictions are complex and specific enough to put at least some parts of AST to a meaningful test, though of course they do not test all aspects of AST. At its heart, AST is a control engineering theory. To make predictions, we should pinpoint, in each circumstance, exactly what the endogenous controller of attention is supposed to be controlling. A controller needs a model of the thing it is controlling. If it is supposed to regulate, sustain, or shift the specific focus of attention that is currently on object X, then for best performance it should require a model of the attention on X, which implies that the person must be conscious of X. As noted above, in AST, consciousness of object X occurs when the attention schema depicts the attention that is focused on X. In contrast, if the endogenous controller of attention is not regulating or adjusting or shifting the attention that is on object X, then it should not require a model of the attention that is on X, and therefore should not require consciousness of X. With that principle in mind, the following predictions can be made.

1.1.1. Prediction 1: Awareness of the cue-target contingencies, and awareness of the cue itself, are not necessary for the exogenous attention effect

In the task, a visual cue appears and exogenously draws visual-spatial attention to itself. Normally, when exogenous attention is drawn to a stimulus, one becomes aware of the stimulus. In AST, this awareness occurs because the attention on the stimulus is incorporated into, or is depicted by, an attention schema. However, that internal model is not necessary for exogenous attention. While a control model is necessary for the endogenous control of attention, exogenous attention is stimulus driven, or externally guided. If the attention schema makes an error and fails to represent the exogenous attention drawn to the cue, then the endogenous control system, with a gap in its internal model, may be less able to adjust, maintain, or regulate that attention, failing to take it into account, as in the studies cited above (Lin and Murray, 2015; Tsushima et al., 2006; Webb et al., 2016a, 2016b). The initial exogenous component of the attention, however, should remain. From the control-theory point of view of AST, awareness of any part of the task, whether awareness of the cue-target contingencies or awareness of

the cue itself, should not be necessary for the exogenous attention effect.

1.1.2. Prediction 2: Visual awareness of the cue is necessary for the endogenous attention effect

The proposed attention schema is analogous to the body schema. In the case of the arm, there are two important components of the model. One is descriptive (what is the current state of the arm?), and the other is predictive (what should be done to move the arm as desired?). Both components are required to control the arm effectively. In the context of the present task, the attention schema needs a descriptive component, keeping track of the current state of attention, and a predictive component, encoding the cue-target contingencies that predict where attention should ideally be adjusted relative to its current location. The two components interrelate as follows: in each trial, spatial attention is exogenously pulled to the cue. The descriptive part of the attention schema can model that focus of attention on the cue. But the system must also learn to enhance attention at a slightly offset location relative to where it is being exogenously pulled, in anticipation of a target stimulus. We argue that the endogenous control system performs this task by keeping track of where spatial attention is being pulled exogenously, and adjusting according to the learned rule, extending the spatial focus of attention to one side. The control system, in that case, must contain descriptive information about the attention pulled to the cue, and must also contain predictive information about the correct adjustment to impose. In that argument, when participants are not aware of the cue, it is the first, descriptive part of the model, not the second, predictive part, that fails. Even if the system learns that attention should be adjusted in a particular way relative to where it is being exogenously drawn, it does not “know” that attention is being exogenously drawn somewhere, because the attention drawn to the cue has not been correctly modeled. Thus, without consciousness of the cue, the endogenous attention effect measured here should be impaired.

1.1.3. Prediction 3: Awareness of the cue-target contingencies is not necessary for the endogenous attention effect

All of the present predictions stem from a principle outlined above: according to AST, modifying, shifting, or adjusting the attention that is on object X in a controlled manner requires consciousness of X. A lack of consciousness of X implies that the attention schema has failed to model the attention on X, which implies that the control of that attention will be impaired. With that principle in mind, we can ask: if subjects are not conscious of the cue-target contingencies, what aspect of attention control will be impaired? Attention operates in many domains, not just in spatial and sensory domains but also in abstract cognitive domains (Chun et al., 2011). While performing the present task, participants can attend to the cue stimulus, to the target, to other items in the space around them, and also, in principle, they can attend to the abstract information about the cue-target relationship. We are all familiar with focusing attention on an abstract thought, an idea that is in our minds at the moment. If the task required participants to control or adjust that particular component of attention – perhaps to suppress it, enhance it, or shift it from that abstract concept to a different concept – then, according to AST, a model depicting that particular component of attention would be required, and thus consciousness of the cue-target contingencies would be required. However, the task does not involve any control or adjustment of that component of attention. In the present task, we measure how the endogenous controller adjusts visual-spatial attention. We do not measure how the endogenous controller adjusts the attention that might or might not be focused on the abstract idea of the cue-target relationship. Therefore, the controller does not need an internal model of attention on the cue-target relationship, for any aspect of the task measured here. Therefore, no aspect of the task measured here requires participants to be conscious of the cue-target relationship. Participants may learn the cue-target contingencies implicitly; participants may also use the cue-target contingencies

implicitly to guide visual-spatial attention; but according to AST, consciousness of those contingencies is not necessary for any aspect of the task.

One might ask: according to the theory, are people conscious of all the information contained in the attention schema? This common question stems from a mistaken understanding of the theory. The theory is not that information enters the attention schema and thus enters consciousness. The attention schema is not a magic receptacle that turns things conscious. It contains descriptive and predictive information about attention, none of which enters consciousness. Moreover, the information of which we *are* conscious is not within the attention schema. For example, when a person is conscious of visual stimulus X, the visual information about X is not a part of the attention schema. The attention schema adds extra information to the larger set of information relevant to object X, and on the basis of that added information, the person can say, “The visual stimulus comes with the property that I am conscious of it.” In analogy, a color-processing network adds extra information on the basis of which one can say, “Object X comes with the property of color.” The theory is essentially about the information required for people to lay claim to consciousness. It is based on the perspective that everything a person thinks, believes, and claims, derives ultimately from specific information in the brain, and therefore to understand consciousness – to understand the claim we make that we have a subjective feel inside us attached to the objects and thoughts we are processing – requires chasing down the source of the relevant information. In AST, a person can claim to be conscious of item X when the person is attending to X and the attention schema represents or depicts that attentional focus on X.

2. Introduction for experiment 1

The purpose of experiment 1 was to collect baseline data on the magnitude of the exogenous and endogenous attention effects when subjects were fully aware of all aspects of the task. Our goal was to be able to quickly and easily measure both exogenous and endogenous attention using the same probe task; therefore, we sought a cue-target interval at which both types of signals were present. Exogenous attention presumably peaks early whereas endogenous attention should remain over a longer interval. The exact timing of attention effects, including exogenous effects, inhibition of return, and endogenous effects, can vary considerably depending on the exact details of the paradigm (McCormick, 1997; Posner, 1980; Posner et al., 1985; Webb et al., 2016a, 2016b). In the present paradigm, as reported below, we found that with an interval of 500 ms, a positive exogenous effect was still present and a positive endogenous effect could also be measured.

3. Methods for experiment 1

3.1. Subjects

All subjects provided informed consent and all procedures were approved by the Princeton Institutional Review Board. In pilot experiments, we found that about 25 subjects were needed for statistical power for the attention effects. Expecting attrition and exclusions, in the present experiment we tested more than needed, hoping to arrive at 25–30 subjects. Twenty-nine subjects were tested (18–33 years old, 12 women, normal or corrected-to-normal vision). None were excluded for poor performance (all performed at > 80 % of trials correct). One was excluded due to a reaction time that was an outlier (using the Grubbs test for outliers, after confirming that the reaction time data was normally distributed). Thus, 28 were included in the final analysis.

3.2. Task

Visual stimuli were presented using Matlab and the Psychophysics Toolbox (Brainard, 1997; Kleiner et al., 2007) on an Acer Predator XB1

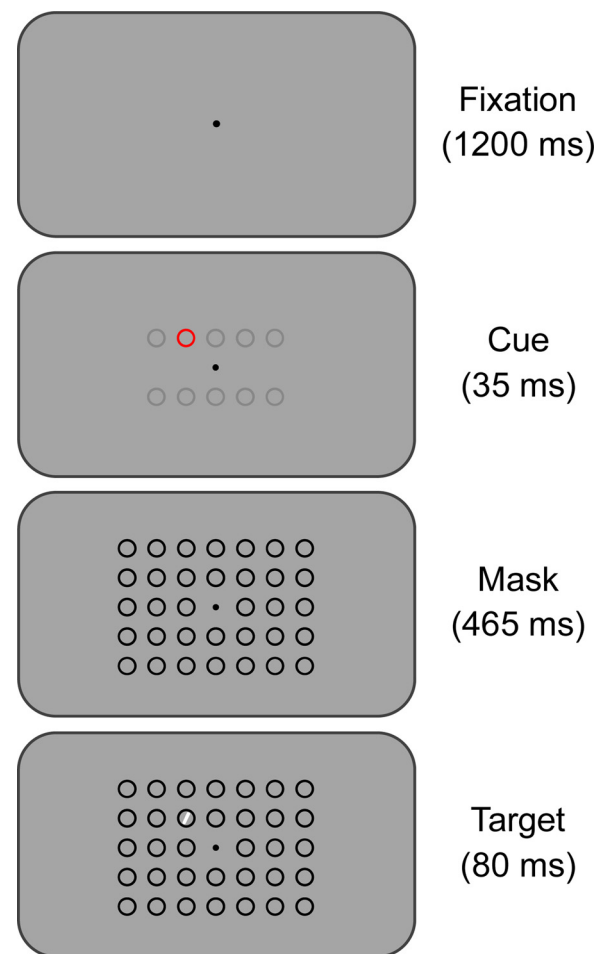


Fig. 1. Task paradigm for experiment 1. After the fixation point appeared, the red cue appeared in one of 10 possible locations (gray circles show possible locations for the cue and were not visible to the subject). Black distractor circles then appeared in a 5×7 grid. The target then appeared, either at the cue location or shifted one position to the right or left of the cue. Subjects discriminated the slant of the target in a reaction-time task.

monitor, with resolution of 2560 by 1440 pixels and a refresh rate of 144 Hz. Subjects sat stabilized by a chinrest 30 cm from the monitor and used key presses on a standard keyboard for behavioral responses.

Fig. 1 shows the task. 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 the point and to maintain fixation at that location throughout the trial. After 1200 ms, the cue stimulus appeared at a peripheral location. The cue was a red annulus (inner diameter 2.75° , outer diameter 3.0°). It 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, 7.0° vertically). The reason for the large number of possible cue locations was related to the subsequent experiments, in which multiple cue locations helped to obscure the cue-target contingencies. To ensure that all experiments were as comparable as possible, the same set of 10 possible cue locations was used in experiment 1.

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, arranged in a 7×7 grid at 3.5° spacing, excluding only the central position (Fig. 1, panel 3). The mask did not prevent participants from seeing the red cue, as indicated by the questionnaire presented to subjects after the experiment. The purpose of the mask was related to a manipulation used in experiment 3 and 4, to remove awareness of the cue. To ensure that all experiments were as comparable as possible, the

mask was used in all experiments including the present one.

After 465 ms (500 ms after the onset of the cue), the target stimulus was added to the display, centered in one of the black annuli, while the annuli remained on the screen. The target could appear in one of three places: either at the same position as the cue, one grid position to the left, or one grid position to the right. The target consisted of a thin white line visible against the neutral gray background. It was angled 10° from vertical, tilted either to the left or right. 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 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 the F key if the target was tilted to the left and the J key if it was tilted to the right. Subjects were allowed a maximum response window of 1000 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. If the subject responded, the response window was immediately terminated, and, after a 500 ms inter-trial-interval during which the screen was blank, the next trial began with the presentation of the fixation point. If no response occurred by the end of the response window, a “too slow” warning was presented on screen, a 1500 ms time-out period was imposed followed by a 500 ms inter-trial-interval, and then the next trial began. Non-responses were rare (< 1% of total trials). After a non-response trial, an additional trial of the same stimulus configuration was added to the randomized schedule of trials, such that the subjects always completed the requisite number of trials per condition. Every 40 trials, subjects were offered a short break.

The task included the following trial types. The cue could be located at any of 10 possible grid locations. The target could be at the same location as the cue, one grid location to the left of it, or one grid location to the right. The target could be tilted toward the left or right. This $10 \times 3 \times 2$ design resulted in 60 trial types. Although all trial types were presented, for purposes of analysis they were collapsed into three main conditions: the target could be presented either to the left, right, or at the same location as the cue. 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 70 % of trials (called “predicted” trials), to the left of the cue on 15 % of trials (called “non-predicted” trials), and at the same location as the cue on 15 % of trials (called “cue-location” trials). Trial types were otherwise counterbalanced and randomly interleaved. Whether the predicted direction was to the right or left was counterbalanced across subjects. Each subject performed 200 trials, taking about 10 min.

Before running any trials, subjects were instructed on the task and given 10 practice trials. During the instruction period, subjects were told explicitly that the target was more likely to appear to one, predicted side of the cue, and less likely to appear to the other side or at the location of the cue. Subjects were also told which specific side was the predicted one.

3.3. Analysis

The main analysis focused on reaction time data because of its greater sensitivity, consistent with previous experiments (McCormick, 1997; Posner, 1980; Posner et al., 1985; Webb et al., 2016a, 2016b). However, we also provide results for accuracy (% correct). The reaction time analyses presented here 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.

For each subject, we calculated two measures which we called the exogenous attention effect and the endogenous attention effect. To measure the exogenous attention effect, or how much attention was

drawn to the location of the cue, we computed $\Delta RT_X = [\text{sum of reaction times for all predicted and non-predicted trials}] / [\text{total number of predicted and non-predicted trials}] - [\text{sum of reaction times in cue-location trials}] / [\text{total number of cue-location trials}]$. A positive score indicates a faster average reaction time when the target was presented at the same location as the cue, as compared to the average reaction time when the target was presented to the two sides of the cue. (Note that because of the unequal numbers of trials, the mean RT for predicted trials and the mean RT for non-predicted trials cannot be combined with a simple average.) The corresponding accuracy result, $\Delta A_X = [\% \text{ correct across all cue-location trials}] - [\% \text{ correct across all predicted and non-predicted trials}]$, is also provided in the results section.

To measure the endogenous attention effect, or how much attention was biased toward the predicted location over the non-predicted location, we computed $\Delta RT_N = [\text{mean reaction time in non-predicted trials}] - [\text{mean reaction time in predicted trials}]$. A positive score on this measure indicates that the subjects utilized the uneven trial statistics and directed more attention to the predicted location. The corresponding accuracy result, $\Delta A_N = [\% \text{ correct in predicted trials}] - [\% \text{ correct in non-predicted trials}]$, is also provided in the results section.

The reason for using a 500 ms interval between cue onset and target onset is that in pilot studies we determined that in the present paradigm, at that interval, both the exogenous and endogenous attention effects could be obtained.

3.4. Posttest questions

After completing the attention trials, subjects were asked whether they had consistently seen the red circular cue that had been pointed out to them during the initial instructions. Subjects were also asked whether they had observed the statistical trend, which was explained to them during the instruction period, in which the target appeared more often to one side of the cue.

4. Results for experiment 1

4.1. Posttest questions

In the question period after testing, all subjects reported that they had no trouble seeing the briefly presented, red circular cue on each trial. All subjects also reported that they were aware of and had observed the statistical trend that had been pointed out to them in the initial instructions, in which the target appeared more often to one side of the cue.

4.2. Task performance

Fig. 2A shows the mean reaction times for the three main conditions. If the only attention effect was the presence of attention exogenously drawn to the cue, then we would expect the central, cued location to have the shortest reaction times, and the two side locations to have equal, longer reaction times. If the only attention effect was the presence of attention endogenously directed by the cue, then we would expect the predicted location (shown on the right) to have the shortest reaction times, and the other two locations to have equal, longer reaction times. However, a combination of the two patterns occurred, suggesting that both exogenous and endogenous attention effects may have been present.

The large error bars in Fig. 2A derive from between-subject variability. To assess the significance of the differences noted above, a within-subjects statistical comparison is needed. As described in the methods, two within-subjects difference scores were computed, ΔRT_X and ΔRT_N , to measure the exogenous and endogenous attention components.

As shown in Fig. 2B, the exogenous attention effect was statistically

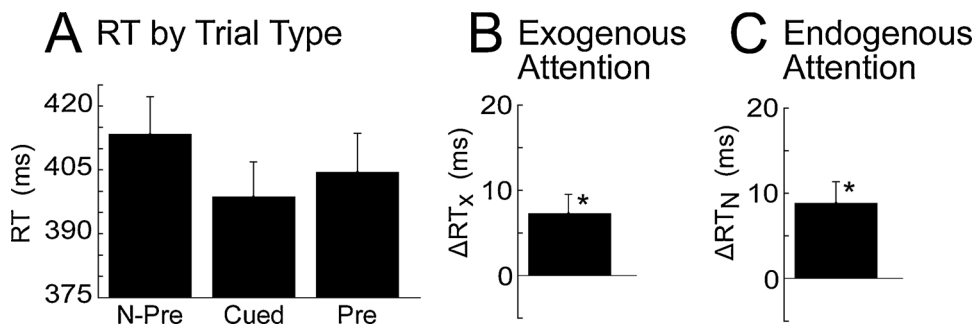


Fig. 2. Results for experiment 1, in which subjects were aware of all aspects of the task. Data from 28 subjects. Error bars show standard error among subjects. Star indicates a difference score significantly different from 0 (two-tailed *t*-test, $p < 0.05$). A. Y axis shows average reaction time for each target location relative to the cue. Targets were 70 % likely to appear to the predicted side of the cue (**Pre**), 15 % likely to appear to the non-predicted side of the cue (**N-Pre**), and 15 % likely to appear at the same location as the cue (**Cued**). B. Y axis shows exogenous attention effect: $\Delta RT_X = [\text{sum of reaction times across all predicted and non-predicted trials}] / [\text{number of predicted and non-predicted trials}] - [\text{sum of reaction times across all cue-location trials}] / [\text{number of cue-location trials}]$. C. Y axis shows endogenous attention effect: $\Delta RT_N = [\text{mean reaction time in non-predicted trials}] - [\text{mean reaction time in predicted trials}]$.

significant, since ΔRT_X was significantly greater than 0 ($\Delta RT_X = 7.29$ ms, SEM = 2.22, two-tailed *t*-test, $df = 27$, $t = 3.29$, $p = 0.003$; for accuracy data, $\Delta A_X = 2.61$ %, SEM = 0.88, two-tailed *t*-test, $df = 27$, $t = 2.95$, $p = 0.006$). In addition, as shown in Fig. 2C, the endogenous attention effect was statistically significant, since ΔRT_N was significantly greater than 0 ($\Delta RT_N = 8.91$ ms, SEM = 2.49, two-tailed *t*-test, $df = 27$, $t = 3.58$, $p = 0.001$; for accuracy data, $\Delta A_N = 2.17$ %, SEM = 1.31, two-tailed *t*-test, $df = 27$, $t = 1.66$, $p = 0.109$).

5. Discussion for experiment 1

Experiment 1 showed that the task used here was successfully able to measure two aspects of visual-spatial attention: first, an exogenous effect in which the onset of a cue stimulus automatically drew attention to itself; and second, an endogenous effect in which the control system was able to use information about the cue-target relationship in order to shift spatial attention toward the target location predicted by the cue. The absolute magnitude of the attention effects was notably small: less than 10 ms for both exogenous and endogenous effects. In many cued attention tasks, the reaction time differences can be much greater, ranging up to 50 ms (e.g. Webb et al., 2016a,2016b). However, the small magnitude of the effect here was expected for the following reason. Most cued attention tasks measure the reaction time differences between relatively distant targets placed to the left or right side of the participant's midline. They measure how spatial attention differs between two spatially distant locations. Here we measured reaction time differences between targets that were spaced 3.5° apart, only minimally deviated from the location of the visual cue. Thus the experiment measured how spatial attention varied across extremely small shifts of location. Any reaction time differences are expected to be small in absolute magnitude. Yet it is important not to confuse a numerically small effect with a statistically small effect. Both the exogenous and endogenous effects were statistically robust, with medium to large effect sizes (Cohen's *D* for the exogenous effect = 0.621 and for the endogenous effect = 0.676, both considered to be medium-large effects).

The results of experiment 1 provide a baseline in which subjects were aware of all aspects of the task. As described next, experiments 2 through 4 tested the consequences of removing awareness of two aspects of the task.

6. Introduction for experiment 2

The purpose of experiment 2 was to remove subjects' awareness of the cue-target contingencies and measure the effect on attention. Subjects were still 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, instead of the two cue locations used in more typical attention paradigms, is that in pilot experiments, we

found that with many cue locations, subjects were unable to notice the statistical relationship between cue and target on their own. The visual impression was of a complicated, flashing, 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 mask them. When questioned after testing, no subjects correctly identified the cue-target relationship, and when told the correct one, all subjects confirmed that they had been unaware of it. While most subjects realized that the target tended to appear near the cue, none realized that the target was more likely to appear to one side of the cue as opposed to the other side. Thus, while being visually aware of the relevant stimuli, subjects could not form an explicit cognitive strategy to direct attention preferentially to one side of the cue. For the cue to have an endogenous effect on attention, biasing it toward the predicted location, the subjects would need to learn the cue-target contingencies implicitly. At least some previous studies found that attention can be influenced by stimulus relationships that subjects learn implicitly and fail to notice explicitly (e.g. Howard et al., 2008; Lambert and Sumich, 1996). As described in the **Introduction**, on the basis of AST we hypothesized that awareness of the cue-target contingencies would not be necessary for the exogenous attention effect (prediction 1) or the endogenous attention effect (prediction 3).

7. Methods for experiment 2

Methods for experiment 2 were the same as for experiment 1 in all respects except in the following ways. The predictive relationship between the cue and the target was not told to the subjects prior to testing. The purpose of the cue was not explained. As far as the subjects knew, the cue was task-irrelevant. After completing the attention trials, subjects were given a verbal questionnaire. They were asked whether they had seen the red circular stimulus during the trials, and whether they had noticed any pattern or relationship between it and the target stimulus.

As in experiment 1, we aimed for 25–30 subjects and tested more in anticipation of possible exclusions. Thirty-four subjects, not tested in experiment 1, were tested in experiment 2 (18–27 years old, 21 women, normal or corrected-to-normal vision). Two were excluded due to poor performance (< 80 % of trials correct). Seven were excluded because they reported having trouble seeing the cue consistently. It is possible that, when subjects were no longer told that the cue was behaviorally relevant, they tended to ignore it, and thus found it more difficult to notice consistently. Twenty-five subjects were included in the final analysis.

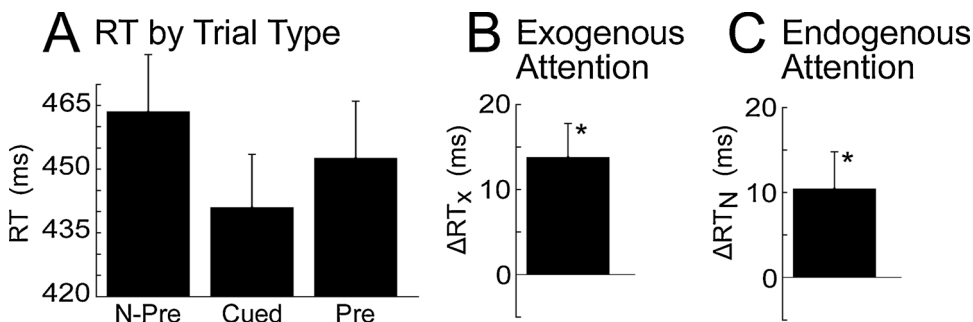


Fig. 3. Results for experiment 2, in which subjects were unaware of the cue-target contingencies. Data from 25 subjects. Error bars show standard error among subjects. Star indicates a difference score significantly different from 0 (two-tailed *t*-test, $p < 0.05$). A. Mean reaction time for targets on the non-predicted side of the cue (N-Pre), at the location of the cue (Cued), and on the predicted side of the cue (Pre). B. Exogenous attention effect. C. Endogenous attention effect.

8. Results for experiment 2

8.1. Posttest questions

Of the 25 subjects included in the analysis, when asked after testing 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 said that they had noticed no other pattern, or they suggested patterns that were not in any way related to the actual pattern (for example, guessing that the target sometimes appeared above or below the cue). Therefore, any preferential directing of attention to the predicted location was likely to be the result of an implicit process and not an explicit strategy.

8.2. Task performance

Fig. 3A shows the mean reaction times for the three main conditions: non-predicted trials, cue-location trials, and predicted trials. Fig. 3B shows that, once again, attention was exogenously drawn to the cue, since ΔRT_x was significantly greater than 0 ($\Delta RT_x = 13.83$ ms, SEM = 3.97, two-tailed *t*-test, $df = 24$, $t = 3.49$, $p = 0.002$; for accuracy data, $\Delta A_x = 0.54$ % correct, SEM = 0.79, two-tailed *t*-test, $df = 24$, $t = 0.69$, $p = 0.498$). Subjects also showed an endogenous attention effect since, as shown in Fig. 3C, ΔRT_N was significantly greater than 0 ($\Delta RT_N = 10.43$ ms, SEM = 4.33, two-tailed *t*-test, $df = 24$, $t = 2.41$, $p = 0.024$; for accuracy data, $\Delta A_N = 0.78$ %, SEM = 1.00, two-tailed *t*-test, $df = 24$, $t = 0.78$, $p = 0.442$). Thus, even though subjects reported no explicit knowledge of the relationship between cue and target, the control system for attention was able to learn the relationship and take advantage of it to help guide attention to the target. Both exogenous and endogenous attention survived the manipulation.

9. Discussion for experiment 2

In experiment 2, subjects were unaware of the cue-target contingencies. They did not know explicitly that the cue predicted which side the target would appear. Yet the subjects were still able to use the cue to guide attention to the predicted location. Evidently, the subjects did not need consciousness of the cue-target contingencies, and did not need to develop an explicit cognitive strategy, for the endogenous attention controller to benefit from the information conveyed by the cue. As predicted by AST, awareness of the cue-target contingencies was not necessary for either the exogenous or the endogenous attention effect.

It should be noted here that the term 'endogenous attention' may mean different things to different researchers. Some may consider it by definition an explicit, intentional, cognitive process of directing attention, in which case the effect obtained here does not qualify. We, however, are using the term 'endogenous' more broadly to refer to an internal control system that regulates attention, whether explicitly or implicitly. In the present paradigm, the attention control system can incorporate the cue-target contingencies and use them to direct the spatial locus of attention, and can do so without the subjects' explicit,

conscious knowledge.

10. Introduction for experiment 3

In experiment 3, just as in experiment 2, subjects were not aware of the cue-target contingencies. However, in addition, in experiment 3, the subjects' visual awareness of the cue was also removed. The method of masking involved a minimal change. Whereas in experiment 1 and 2, the cue was red and followed immediately by a black masking pattern, in experiment 3, the cue was black and followed by the same black masking pattern (for details, see **Methods for Experiment 3**). In these conditions, subjects did not even realize that a cue was presented. As described in the **Introduction**, based on AST, we predicted that without awareness of the cue, exogenous attention should still be drawn to the cue (prediction 1), but the endogenous control of attention with respect to the cue should be compromised (prediction 2). A similar result has been obtained in previous studies (e.g. McCormick, 1997; Travers et al., 2018).

11. Methods for experiment 3

The methods for experiment 3 were the same as for experiment 2 except in the following ways. The cue, which was red in experiment 2, was black in experiment 3. In the context of this paradigm, when a black cue is followed by a black mask, the cue becomes subjectively invisible (Lin and Murray, 2015). In the instruction period, subjects were not told about the presence of the cue. They were given no explicit knowledge that it existed or that it predicted the location of the target. All other aspects of the task remained the same. After completing the task, subjects were asked whether they had noticed the black cue appearing before the mask. They were then shown a reduced-speed example of a trial, in which the black cue was clearly visible, and asked if they had seen anything that looked like the cue during the trials.

Many paradigms test whether subjects are aware of a cue stimulus on a trial-by-trial basis. The difficulty with that type of measure is that it explicitly tells subjects a cue is present, drawing their attention to it, and increasing the likelihood that subjects will become aware of it. We chose, instead, not to inform subjects about the cue, so that the subjects would be less likely to become aware of it. Only after the full set of trials was complete were subjects asked about the possible presence of the cue. Many paradigms also test awareness 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, we do not believe objective measures of awareness necessarily address the question of subjective awareness (e.g. Merikle et al., 2001). For these reasons we relied on subjects' reports after completing all trials in the task.

As in experiment 1, we aimed for 25–30 subjects and tested more in anticipation of possible exclusions. Thirty-five subjects, not tested in experiments 1 or 2, were tested in experiment 3 (18–57 years old, 17 women). Two were excluded from analysis due to target discrimination accuracy below 80 %. Two were excluded because their vision could

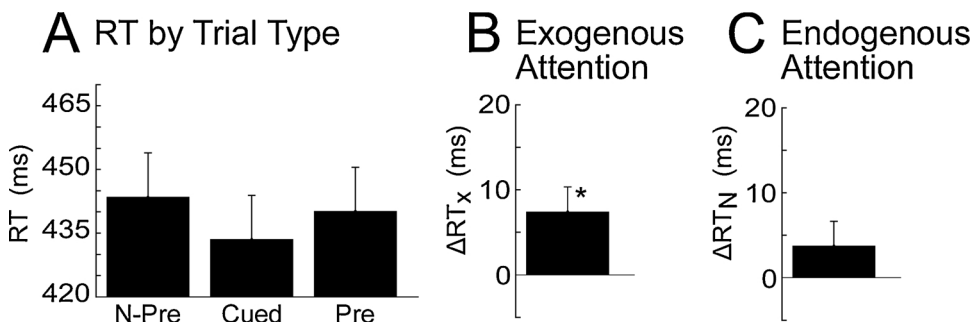


Fig. 4. Results for experiment 3, in which subjects were unaware of the cue-target contingencies and unaware of the cue. Data from 27 subjects. Error bars show standard error among subjects. Star indicates a difference score significantly different from 0 (two-tailed t -test, $p < 0.05$). A. Mean reaction time for targets on the non-predicted side of the cue (N-Pre), at the location of the cue (Cued), and on the predicted side of the cue (Pre). B. Exogenous attention effect. C. Endogenous attention effect.

not be corrected to normal. Four were excluded because, as noted below, they may have been partially visually aware of the cue. Thus 27 participants were included in the final analysis.

12. Results for experiment 3

12.1. Posttest questions

After testing, only one subject reported having seen the cue during at least some of the trials. After being shown a reduced-speed example of a trial, three additional subjects reported that they may have seen something that looked like the cue during the trials. However, they reported seeing it rarely (< 5 times throughout the task) and expressed surprise upon learning that it was present in every trial. These results suggest that the mask successfully reduced awareness of the cue and in most subjects eliminated it. We removed from analysis all four subject who had given any indication that they might have seen the cue.

12.2. Task performance

Fig. 4A shows the mean reaction times for the three main conditions: non-predicted trials, cue-location trials, and predicted trials. Fig. 4B shows that even though subjects were not aware of the cue, it still drew exogenous attention, since ΔRT_X was significantly greater than 0 ($\Delta RT_X = 7.43$ ms, SEM = 2.90, two-tailed t -test, $df = 26$, $t = 2.57$, $p = 0.016$; for accuracy data, $\Delta A_X = 0.82$ % correct, SEM = 0.76, two-tailed t -test, $df = 26$, $t = 1.08$, $p = 0.289$).

Subjects did not, however, show a statistically significant endogenous attention effect. As shown in Fig. 4C, ΔRT_N was not significantly different from 0 ($\Delta RT_N = 3.78$ ms, SEM = 2.88, two-tailed t -test, $df = 26$, $t = 1.31$, $p = 0.199$; for accuracy data, $\Delta A_N = 1.58$ %, SEM = 1.16, two-tailed t -test, $df = 26$, $t = 1.35$, $p = 0.186$). Thus, as predicted, in the absence of awareness of the cue, the exogenous effect survived and the endogenous attention effect was impaired.

13. Discussion for experiment 3

If the results of experiment 3 were to be taken in isolation, an easy interpretation would be available. Because participants were unaware that a cue was presented, they were therefore also unaware of the cue-target relationship. Therefore, they could not consciously direct their attention to the predicted location. This interpretation, as obvious as it seems, is clearly incorrect given the results of experiment 2. In experiment 2, subjects were not conscious of the cue-target relationship, and yet learned it anyway, shifting attention to the location predicted by the cue without knowing they were doing so. Consciousness of the cue-target relationship, and an explicit, conscious choice to move attention, are evidently not necessary for the endogenous attention effect. In experiment 3, therefore, the lack of an endogenous effect cannot be attributed to a lack of consciousness of the cue-target relationship. Something more complex must be going on. Somehow, visual awareness of the cue itself is necessary for an implicit, endogenous attention effect. As explained in the **Introduction** (Prediction 2), the control-

theory logic of AST predicts this specific outcome.

14. Introduction for experiment 4

Experiment 3 suggests that without visual awareness of the cue, the endogenous attention effect was compromised. However, it is possible that in experiment 3, we missed a small or subtle effect of endogenous attention due to a lack of statistical power, or that learning occurred in a slower manner requiring more trials to observe. To test thoroughly whether the endogenous attention effect can survive the loss of awareness of the cue, in experiment 4 we repeated the same paradigm as in experiment 3, but increased the number of subjects and tripled the number of trials from 200 to 600. We asked whether, by the final block of 100 trials, subjects showed any evidence of an endogenous attention effect, suggesting that the control system might still be able to learn the cue-target contingencies.

15. Methods for experiment 4

The methods were the same as for experiment 3 except in two ways. First, instead of 200 trials, subjects performed 600 trials. Second, the total number of subjects was increased to improve statistical reliability and therefore the sensitivity of the experiment to a possibly subtle learning effect. Forty-three subjects, not tested in experiments 1–3, were tested (18–22 years old, 29 women, normal or corrected-to-normal vision). Three were excluded from analysis due to poor performance. Four were excluded because they reported being aware of the cue on some trials. Thus 36 participants were included in the final analysis.

16. Results for experiment 4

16.1. Posttest questions

Since subjects were exposed to 600 trials over about 40 min, one concern was that with repeated exposure they may have begun to notice the cue stimulus by the end of the session. However, in the questions afterward, only one subject reported having seen the cue on a few trials. After being shown a reduced-speed example of a trial and asked whether during the task they had ever seen the cue apparent in that example, three additional subjects reported that they may have seen something that looked like the cue during some of the trials. These four subjects were removed from the analysis. All other subjects appeared to have remained unaware of the cue throughout the experiment.

16.2. Task performance

Fig. 5 shows the endogenous effect broken down into six blocks of 100 trials each. In the first five blocks, subjects showed no clear evidence of an endogenous effect, consistent with the result of experiment 3, in which no significant endogenous effect was found within 200 trials. In the final block, however, ΔRT_N was significantly greater than zero. To avoid multiple comparisons, we planned a statistical test of the

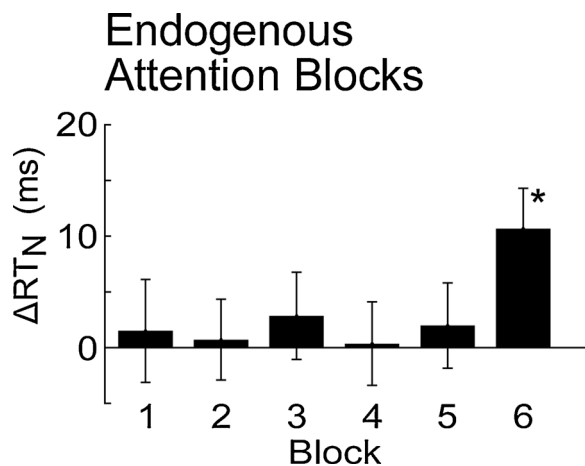


Fig. 5. Results for experiment 4. Subjects were unaware of both the cue-target contingencies and the cue, and were tested with 600 trials, three times more than in experiment 3. Data from 36 subjects. Error bars show standard error among subjects. Star indicates significant difference from 0, $p < 0.05$ (two-tailed t -test). The endogenous attention effect is shown broken into six, 100-trial blocks.

final block only, as the most direct test of the hypothesis that learning had occurred. In this final block, ΔRT_N was significantly greater than 0. (Planned comparison: $\Delta RT_N = 10.70$ ms, SEM = 3.62, two-tailed t -test, $df = 35$, $t = 2.92$, $p = 0.006$; for accuracy data, $\Delta A_N = 0.89$ %, SEM = 1.26, two-tailed t -test, $df = 29$, $t = 0.70$, $p = 0.490$.) As an alternative analysis, even when using a Bonferroni correction for 6 blocks (in which the calculated p value must be < 0.008 to pass a 0.05 alpha level), the final block still showed a statistically significant effect. The results suggest that with extended trials, the endogenous control system for attention may be able to eventually learn to use cue-target contingencies to direct attention to the target, even when subjects are unaware of the cue and unaware of the spatial rule that they are learning.

17. Discussion for experiment 4

On the basis of AST, we predicted that without awareness of the cue, the control system would be impaired – unable to effectively shift attention toward the location indicated by the cue. That prediction was confirmed in experiment 3. Experiment 4 shows that when testing more subjects and more trials, the endogenous attention effect might not be entirely eliminated. However, if it is present, it appears to be greatly delayed, by at least 500 trials. The main finding of experiments 3 and 4, therefore, is that the endogenous shifting of attention to the predicted location, which is robust and rapid in the presence of awareness of the cue, becomes impaired in the absence of awareness of the cue.

One possible interpretation is that two learning mechanisms are present. One mechanism, a much faster and more robust one, depends on awareness of the cue. A second learning mechanism, a slower and weaker one, may be able to proceed incrementally over many trials even without awareness of the cue. In that speculation, the essential difference between the two mechanisms is that one is model-based, providing the attention system with fast, flexible learning, whereas the other is model-free, providing a much slower, incremental learning (see Haith and Krakauer, 2013). Experiment 4, with its increased number of subjects and trials, may have begun to reveal that second, incremental learning mechanism. The possibility of two learning mechanisms is intriguing and worth further study. However, the present results merely hint at the possibility.

18. Introduction for experiment 5

The purpose of experiment 5 was to more thoroughly test the attention paradigm used in experiments 1–4, in order to prepare for a test of spatial generalization in experiment 6. Here we focused on the endogenous attention effect, and no longer measured the exogenous effect. We also added more potential cue and target locations, increasing the spatial coverage across the screen. Finally, by measuring eye position, we also examined whether the results could be explained by adaptation of overt attention (the motor adaptation of eye position or eye movement), or whether they were better explained by an adaptation of covert attention toward the predicted target location.

19. Methods for experiment 5

As in experiment 1, we aimed for 25–30 subjects and tested more in anticipation of possible exclusions. Thirty-seven participants, not tested in experiments 1–4, were tested here (18–38 years old, 28 women, normal or corrected-to-normal vision). Four were excluded from analysis due to failure of the eye-tracking equipment. Two were excluded because they reported being unable to see the visual cue clearly or consistently. One was excluded because of a mean reaction time that was an outlier (using the Grubbs test for outliers). Thus 30 participants were included in the final analysis. As in experiment 2, subjects were exposed to a red, clearly visible cue, but were not told about the cue-target contingencies.

Fig. 6 shows the paradigm. Unless otherwise specified, the trial events were the same as in experiment 2. The red annular cue could be in any of 22 possible locations around the screen. These locations formed a grid with 3.5° spacing between positions. The gray circles in Fig. 6 illustrate possible locations of the cue, but were not visible to participants. The grid locations eliminated the central position (so that the cue would not overlap the fixation point) and the positions to either side of center (so that the subsequent target would never overlap the fixation point). After the cue, a mask consisting of an array of black annuli was presented. The black annuli were arranged in a 7×7 grid at 3.5° spacing, excluding only the central position. The target stimulus was presented either one grid position to the left, or one grid position to the right, of the prior cue position.

The task included the following trial types. The cue could be located at any of 22 possible grid locations distributed around the screen. The target could be located one grid location to the left or the right of the cue. The target itself could be tilted toward the left or right. This $22 \times 2 \times 2$ design resulted in 88 trial types. Although all trial types were presented, for purposes of analysis they were collapsed into two conditions: the target could be presented either to the left or to the right of the cue. 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 “non-predicted” trials). Trial types were otherwise counterbalanced and randomly interleaved. Whether the predicted direction was to the right or left was counterbalanced across subjects. Each participant performed 384 trials.

Eye position was measured with an infrared eye tracker (SensoMotoric Instruments RED-500). Eye position was calibrated at the start of the experiment, and trials on which participants broke fixation were detected in subsequent analysis by a velocity cutoff ($50^\circ/\text{sec}$), which was sufficient to identify saccadic eye movement. Trials in which participants blinked also registered as a fast change in measured eye position and were identified by the same velocity cutoff. On average across participants, 13 % of trials were identified as saccade or blink trials. As described in the results, in some analyses, this 13 % of trials was removed.

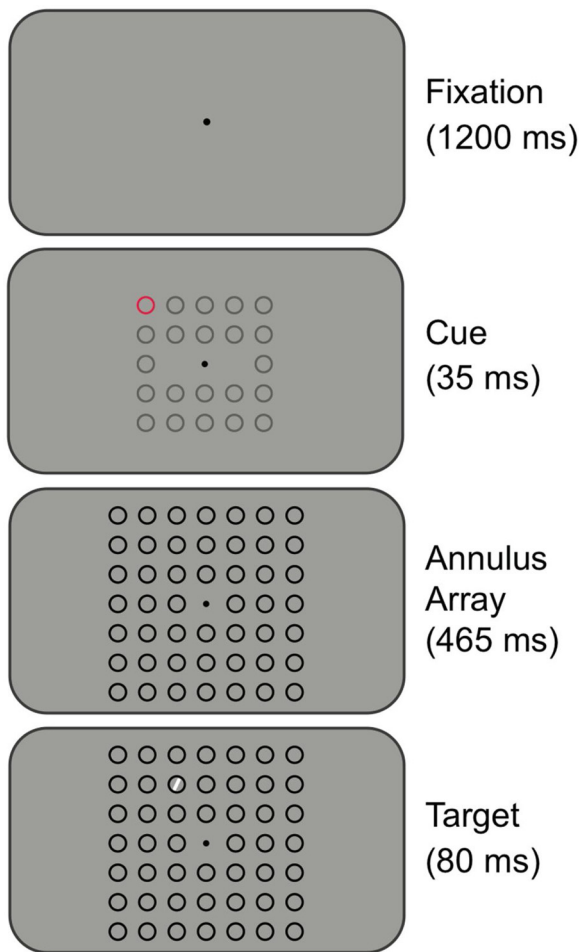


Fig. 6. Task paradigm for experiment 5. Similar to experiment 1 except more cue locations were used and the post-cue annulus array was larger. After the fixation point appeared, the red cue appeared in one of 22 possible locations (gray circles show possible locations for the cue and were not visible to the subject). Black distractor circles then appeared in a 7×7 grid. The target then appeared, shifted either one position to the right or left of the cue (85 % on predicted side, 15 % on non-predicted side). Subjects discriminated the slant of the target in a reaction-time task.

20. Results for experiment 5

20.1. Task performance

As shown in Fig. 7A, the mean ΔRT_N among the 30 participants was significantly greater than zero ($\Delta RT_N = 9.50$ ms, SEM = 2.57, two-tailed t -test, $df = 29$, $t = 3.70$, $p < 0.001$; for accuracy data, $\Delta A_N = 2.38$ %, SEM = 0.69, two-tailed t -test, $df = 29$, $t = 3.46$, $p = 0.002$). The positive result corroborates our previous finding. Even though subjects lacked explicit knowledge of the cue-target contingencies, the attention control system was able to learn those contingencies, such that attention was greater on the predicted side of the cue than on the non-predicted side. Unlike in experiments 1–4, in experiment 5, both the latency data (ΔRT_N) and the accuracy data (ΔA_N) showed a significant positive effect (see **Methods for Experiment 1** for the definition of these two metrics). The reason is probably that the cue was more predictive in experiment 5 than in the previous four experiments (85 % predictive, rather than 70 % predictive). Although the latency data is typically more sensitive to attentional differences, and the accuracy data relatively insensitive, in the present experiment the attention signal was apparently clear enough to be evident in both types of data.

20.2. Eye position and movement

It is possible that subjects broke fixation and made saccades toward the predicted location during the trial. It is also possible that subjects maintained a steady fixation, but learned to fixate on a position systematically offset from the central dot, in a process of motor adaptation. These motor effects might explain the pattern of results by bringing the better acuity of the fovea closer to the predicted target locations, thus resulting in shorter reaction times. To determine whether the adaptation involved overt attention or covert attention, we first used a velocity cutoff (50°/sec) to filter out trials in which any saccades or blinks occurred (a standard method in the oculomotor literature, e.g. Salvucci and Goldberg, 2000; Cooke and Graziano, 2003). For the remaining trials (87 % of trials on average across participants), no saccades or blinks occurred and fixation was maintained within less than 1 degree. Fig. 7B shows the results for the trials in which strict fixation was confirmed. Even in the confirmed fixation trials, the mean effect of attention was significantly greater than zero ($\Delta RT_N = 9.60$ ms, SEM = 2.58, two-tailed t -test, $df = 29$, $t = 3.72$, $p < 0.001$; for accuracy data, $\Delta A_N = 2.57$ %, SEM = 0.73, two-tailed t -test, $df = 29$, $t = 3.53$, $p = 0.001$). Thus eye movements during the trial did not account for the results.

We also tested whether a systematic bias in fixation location could account for the results. We tested whether participants who were trained on a rightward attention shift tended to fixate on a location displaced to the right of the center of the screen, compared to participants who were trained on a leftward attention shift. A bias in fixation location toward the entrained direction could potentially explain the difference in reaction times. We computed $\Delta X = [\text{mean } X \text{ location of eye position, measured during those trials in which no saccade occurred and fixation was confirmed, for all participants trained on a rightward attentional shift}] - [\text{the same for all participants trained on a leftward attentional shift}]$. For this analysis, we analyzed eye data within a time window from the onset of the cue to the participant's response, on each trial. ΔX was not significantly greater than zero, and was actually marginally below zero (mean $\Delta X = -0.06^\circ$; SEM = 0.27; two-tailed t -test, $df = 29$, $t = 0.22$, $p = 0.825$). We found no evidence that the training caused participants to fixate in a spatially biased manner.

20.3. Location relative to screen versus location relative to cue

Suppose a subject is trained on a rightward shift. The subject implicitly learns to pay more attention “to the right,” but to the right of what? One possibility is that the subject learns to attend more to locations that are on the right side of the screen, regardless of the location of the cue. Another possibility is that the subject learns a specific cue-target relationship; wherever the cue appears on the screen, the subject offsets attention to the right of the cue. It is also possible that both types of learning occurred.

We first examined whether subjects learned the cue-target relationship. Consider the seven columns of possible target positions (shown as black circles in the third panel of Fig. 6). Consider the central three columns. A target appearing at these locations could have been preceded by a cue to either side of it. On some trials, those target locations were to the predicted side of the cue, and on other trials the same target locations were to the non-predicted side of the cue. Were reaction times faster for predicted than for non-predicted trials? The usefulness of this analysis of the middle three columns is that target location is held constant while cue location is varied. (The analysis also excluded the middle, horizontal row. As shown in Fig. 6, in the middle three columns, no cues were presented in the middle row.) As shown in Fig. 7C, the results for this subset of the data confirm a significant effect of attention ($\Delta RT_N = 8.36$ ms, SEM = 3.44, two-tailed t -test, $df = 29$, $t = 2.43$, $p = 0.021$; for accuracy data, $\Delta A_N = 3.16$ %, SEM = 1.10, two-tailed t -test, $df = 29$, $t = 2.87$, $p = 0.008$). Attention was increased to the predicted side relative to the cue, even when target location on

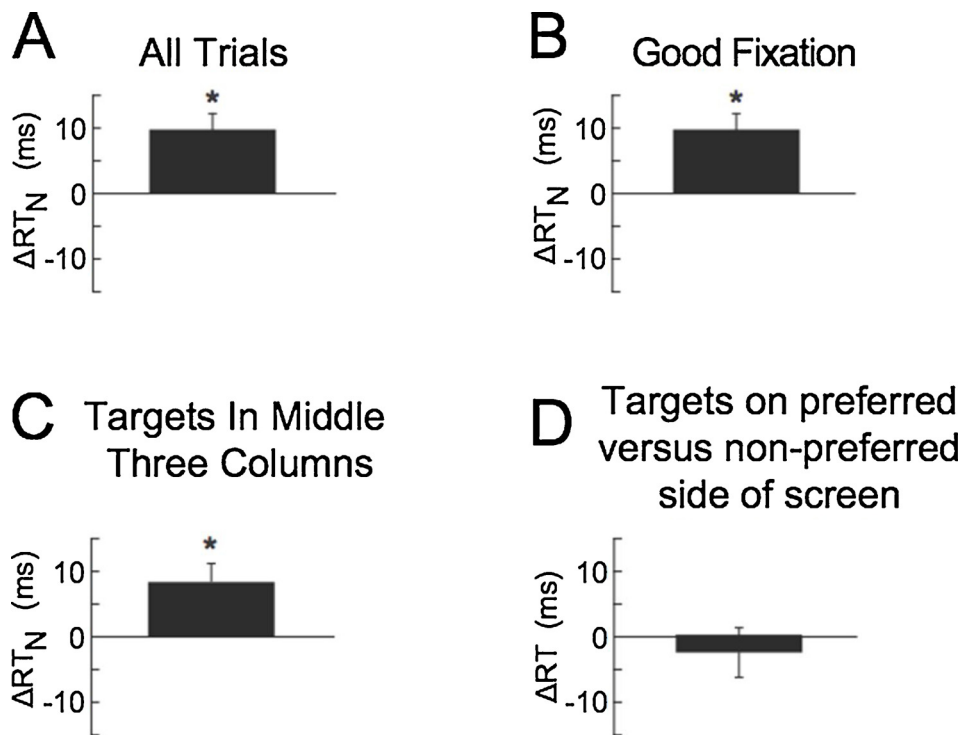


Fig. 7. Results for experiment 5. Data from 30 participants. A. Results for all trials. Y axis shows ΔRT_N = [mean reaction time for non-predicted trials] – [mean reaction time for predicted trials]. Error bar shows standard error among participants. Star indicates significant difference from 0, $p < 0.05$, in planned two-tailed t -test (see text for statistics details). B. Results for trials during which good fixation was confirmed. C. Results for trials during which the target appeared in the middle three columns, and good fixation was confirmed. D. Test of whether attention was stronger on one side of the screen. ΔRT = [reaction time for trials when the target appeared to the non-predicted side of the screen, regardless of cue location] – [reaction time for trials when the target appeared to the predicted side of the screen, regardless of cue location], again for trials when good fixation was confirmed.

the screen was held constant. Thus attention adapted to the cue-target spatial relationship.

We next examined whether participants also learned to increase attention overall to one side of the screen, which should manifest as shorter reaction times to targets on the more attended side. Consider again the seven columns of possible target positions. The end columns were eliminated from this analysis, because they do not provide a fair test. Targets in these positions could be preceded only by cues from one direction. The middle column was eliminated from the analysis because it is not revealing about the two sides of the screen. The analysis therefore focused on column 3 on the left side, and column 5 on the right side. (Again, the analysis also excluded the middle, horizontal row. As shown in Fig. 6, in the middle three columns, no cues were presented in the middle row.) We computed a difference score for each subject, ΔRT = [mean reaction time when the target appeared on the non-predicted side of the screen, e.g. left side of screen for participants trained on a rightward shift] – [mean reaction time when the target appeared on the predicted side of the screen, e.g. right side of screen for participants trained on a rightward shift]. If subjects learned to pay more attention to the predicted side of the screen, the difference scores should be significantly greater than zero. As shown in Fig. 7D, the mean difference score was not significantly different from zero (ΔRT = -2.56 ms, SEM = 3.33, two-tailed t -test, $df = 29$, $t = -0.771$, $p = 0.447$; for accuracy data, ΔA = -2.93 %, SEM = 2.28, two-tailed t -test, $df = 29$, $t = -1.287$, $p = 0.208$).

21. Discussion for experiment 5

The results of experiment 5 confirm the main finding from experiment 2: the endogenous attention effect is robust even when subjects are not explicitly aware of the cue-target contingencies. The results also rule out several possible artifacts. Subjects did not adapt overt attention, such as saccadic eye movements or fixation location. Instead, covert attention was adapted. Participants also did not learn a simple spatial distribution of attention, such as attending more to one side of the screen. Instead, the attention control system adapted to a specific relationship between the cue and the target, even though subjects were not explicitly aware of that relationship.

22. Introduction for experiment 6

In control theory, one of the most useful aspects of an internal model is that it can adapt to changing circumstances, thus giving the control system a layer of flexibility. For example, the internal model of the arm has been widely tested through adaptation paradigms (Gandolfo et al., 1996; Shadmehr and Moussavi, 2000; Shadmehr and Mussa-Ivaldi, 1994; Thoroughman and Shadmehr, 2000). Over repeated trials, people rapidly adapt to a force field applied to the arm or to a spatial shift in the visual location of the target. That adaptation is thought to occur within the brain's predictive model of the arm. One concern in these adaptation experiments is that participants might learn to adjust a few specific directions of reach without learning anything about intermediate directions. If adaptation is entirely local to the specific, trained directions and cannot generalize to intermediate locations, then the evidence does not point to a general rule learned by an internal model. By implication, the adaptation must be occurring within limited, low-level processes. The proposal of an adaptable internal model of the arm was not widely accepted until the results were shown to generalize to intermediate reach directions (Imamizu et al., 1995; Shadmehr and Moussavi, 2000; Thoroughman and Taylor, 2005). This generalization is limited. For example, training on one side of the reaching workspace will not necessarily generalize to the opposite side of the workspace. The internal model is apparently not a simple, one-size-fits-all model. However, the finding that some degree of spatial generalization can occur was interpreted as strong evidence that a deeper model of the arm can be adapted.

We asked the same generalization question in our attention paradigm. In experiment 6, participants were first trained on 16 cue locations scattered across the screen, selected randomly among the 22 possible locations. For each cue location, the target was more likely to appear to one side of the cue (the predicted side, 85 % of trials) and less likely to appear to the other side of the cue (the non-predicted side, 15 % of trials). One possibility is that the attention system underwent 16 different, spatially local adaptations. Such a piecemeal adaptation might be possible if what is adapted lies at the visual processing end, in a retinotopic map. Alternatively, a more general rule might have been learned at a deeper level, in which case the learning should be apparent

when tested at the 6 previously untrained cue locations. The purpose of Experiment 6 was to test whether the attentional learning was entirely local, specific to the 16 trained locations, or whether it showed generalization to the 6 nearby, intermediate locations.

23. Methods for experiment 6

The methods for experiment 6 were the same as for experiment 5 except in the following ways.

Expecting any possible generalization effect to be subtle – more subtle than the primary attention effect measured in the previous experiments – we increased the number of subjects from the previous target sample size of 25. Forty-six participants were tested (18–51 years old, 32 women, normal or corrected-to-normal vision). One participant was excluded from analysis due to poor performance on the task (less than 80 % of trials correct). One participant was excluded because he reported being unable to see the cue consistently. Thus, 44 participants were included in the final analysis.

23.1. Training phase

Sixteen of the 22 possible cue locations were used. These 16 trained locations were different for each participant and were selected randomly. The cue was equally likely to appear in any of the 16 possible positions. The target was displayed either one grid unit to the right or to the left of the cue. Trial types were proportioned such that the cue statistically predicted one target direction 85 % of the time. Whether the predicted direction was to the right or left of the cue was counterbalanced across participants. Participants performed 384 training trials.

23.2. Generalization phase

Following training, participants performed 308 trials in the generalization phase. Participants were not told that any aspect of the task had changed. Cues were presented at all 22 locations, 16 previously trained locations (accounting for 60 % of the trials) and 6 previously untrained locations (accounting for 40 % of the trials). When the cue appeared in a previously untrained location, the target was equally likely to appear in a position one grid unit to the right or left of the cued location. Thus these 6 cue locations never participated in any training in which the cue-target relationship was spatially biased. On trials in which the cue appeared in the 16 previously trained locations, the same spatially biased proportions of target placement used in the training phase were presented.

24. Results for experiment 6

24.1. Training phase

The mean ΔRT_N among the 44 participants, during the training phase, was significantly greater than zero ($\Delta RT_N = 10.47$ ms, SEM = 2.00, two-tailed t -test, $df = 43$, $t = 5.25$, $p < 0.001$; for accuracy data, $\Delta A_N = 1.64$ %, SEM = 0.68, two-tailed t -test, $df = 43$, $t = 2.40$, $p = 0.021$). Reaction times were faster when the target was to the predicted side of the cue, indicating that attention was spatially shifted to that side of the cue. The results replicate the previous experiments and show that the phenomenon is robust. Fig. 8A shows the mean ΔRT separated into four blocks for an easier comparison to the results in the generalization phase.

24.2. Generalization phase

When the control of the arm is studied in an adaptation paradigm, the extent of learning can be measured on a trial-by-trial basis (e.g. Shadmehr and Mussa-Ivaldi, 1994). In a very few trials, sometimes as

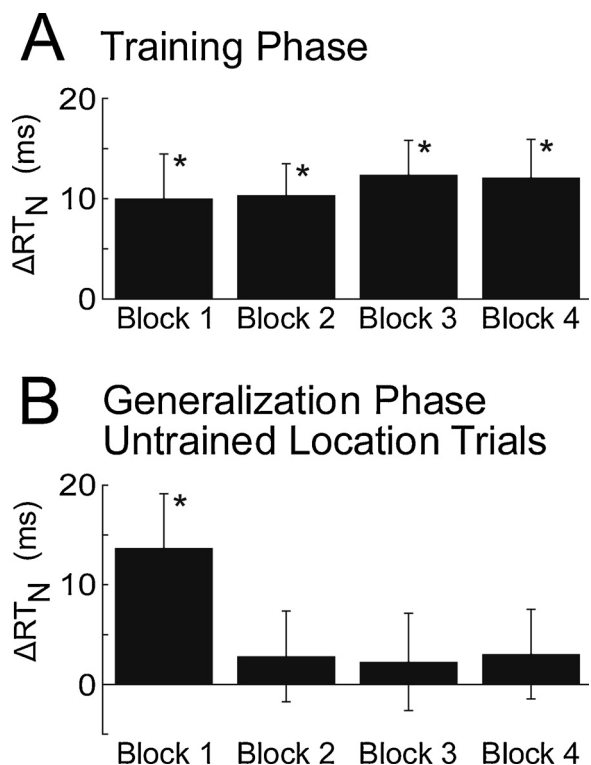


Fig. 8. Results from experiment 6. Data from 44 participants. A. Training phase, in which 16 cue locations were trained. Y axis shows the ΔRT_N = [mean reaction time for non-predicted trials] – [mean reaction time for predicted trials]. Error bars show standard error among participants. Stars indicate significant difference from 0, $p < 0.05$, in planned two-tailed t -test (see text for statistics details). Data divided into four blocks of 96 trials each. B. Post-training, generalization phase, showing data for the six cue locations that were previously untrained. Data divided into 4 blocks of 77 trials each.

few as five or ten, an adaptation curve is measured and the full extent of learning is complete. In a test of an after-effect, when the forcefield or spatial shift is removed, the prior learning can be observed only within the first few trials and then quickly disappears during the repetition of unbiased control trials. In the present experiment in the attention domain, however, trial-by-trial measurement of learning is not possible. Instead, to obtain a statistically reliable measure of the attention effect, a block of about 50–100 trials is necessary, over which learning may saturate. Judging from the results in Fig. 8A, the entire learning process occurs within the first block of trials and the learning curve cannot be pragmatically obtained. Moreover, here we trained subjects in one phase of trials and then tested the extent of learning in a second phase of trials. That second phase, extending over 308 trials, is not just a measure of how much learning occurred in the previous phase, but is also likely to contribute its own training effect, adding complexity to the pattern of results. The results must therefore be interpreted with these cautions and complexities in mind.

Fig. 8B shows the results in the generalization phase, for the 6 locations that were not included in the initial training phase. To avoid entraining an attention effect at these locations, cues at these locations did not predict whether the target would appear to the left or the right of the cue. Thus, these cue locations were never subjected to any spatially biased training. The results are divided into four blocks of 77 trials each, to show the changes in performance over time. Once again, it is important to understand that because each block includes many trials, learning is likely to have occurred within the block. Thus, in the first block of the generalization phase, we may see evidence of an attention effect lingering from the training phase. But during that first block, it is likely that exposure to the new statistics, in which the cue

does not predict the target, would undo the previous learning and remove the attention effect. Because of these considerations, we predicted that if generalization occurred, it would be observed as an attention effect present in the first block and would fade in subsequent blocks. The prediction that generalization occurred, therefore, rested on one crucial test: was a significant attention effect present in the first block for the previously untrained cue locations?

Fig. 8B shows that, as predicted, a significant adaptation effect was obtained in the generalization phase, in the first block of trials, but not in subsequent blocks (two-tailed *t*-test, block 1, $\Delta RT = 13.70$ ms, $SEM = 5.45$, $df = 43$, $t = 2.49$, $p = 0.017$; block 2, $\Delta RT = 2.81$ ms, $SEM = 4.55$, $df = 43$, $t = 0.61$, $p = 0.545$; block 3, $\Delta RT = 2.25$ ms, $SEM = 4.88$, $df = 43$, $t = 0.46$, $p = 0.650$; block 4, $\Delta RT = 3.04$ ms, $SEM = 4.51$, $df = 43$, $t = 0.67$, $p = 0.508$; for accuracy data, two-tailed *t*-test, block 1, $\Delta A = 2.46$ % ms, $SEM = 1.23$, $df = 43$, $t = 1.98$, $p = 0.054$; block 2, $\Delta A = 1.06$ %, $SEM = 1.14$, $df = 43$, $t = 0.91$, $p = 0.362$; block 3, $\Delta A = 1.37$ %, $SEM = 1.31$, $df = 43$, $t = 1.03$, $p = 0.309$; block 4, $\Delta A = 1.71$ %, $SEM = 1.12$, $df = 43$, $t = 1.51$, $p = 0.137$). Thus the adaptation from the previous training phase was transferred to cue locations that had never been trained. That adaptation effect was, as expected, transient. It disappeared after the first block of trials, indicating that the participants adapted to the spatially unbiased statistics of the new cue locations after exposure to many trials.

25. General discussion

This series of six experiments examined whether some aspects of attention and awareness adhered to the predictions of AST. While the results support a complex and specific set of predictions, they do not mean that the theory is confirmed. The same results could in principle be explained in alternative ways, and the theory is much more extensive than the predicted properties examined here. These results add to a growing literature that suggests, in our interpretation, that AST remains plausible. Attention and awareness are indeed separable, yet they do indeed interact in a complex and specific manner. Awareness seems to enhance or permit the internal control of attention, as an internal model of attention is predicted to do. When new spatial dynamics of attention are learned, that learning generalizes across space to adjacent locations, at least to some degree, again as expected if attention depends on an adaptable internal model.

In the first four experiments, several predictions of AST were tested. Exogenous attention attracted to a cue required neither awareness of the cue-target contingencies, nor awareness of the cue. Endogenous attention, directed on the basis of that cue, did not require awareness of the cue-target contingencies, but almost entirely disappeared without awareness of the cue. Experiment 4 provided evidence that there may still be an extremely slow, minimal learning that occurs without awareness of the cue, allowing attention to be redirected to the location predicted by the cue. However, the main finding on the endogenous attention effect is that, while it was robust when people were aware of the cue, it almost entirely disappeared when people were not aware of the cue, a finding consistent with previous reports (McCormick, 1997; Travers et al., 2018).

In the final two experiments, attention was adapted over a broad range of locations across the screen. That adaptation generalized and affected performance at adjacent, untrained locations. This type of spatial generalization is reminiscent of the spatial generalization seen in other adaptation paradigms, such as in the control of the arm, indicative of a deeper, adaptable, internal model. Again, the result matched a specific prediction based on AST.

25.1. Control theory and an internal model of attention

Covert attention seems like an intangible entity compared to a physical object such as an arm or an eyeball. When attention shifts,

nothing physical is moving. Moreover, unlike a body part which is subject to physical constraints, visual attention can move in complex ways, spreading, changing shape, or adjusting intensity. It can move through dimensions not tested here. Spatial visual attention is only one limited window on the issue. Feature attention, attention in other sensory domains, and even attention to cognitive events, are possible. Yet despite the added complexity of attention – perhaps even because of it – the principles of control engineering may be just as useful for controlling this intangible, amorphous thing as for controlling a body part. We argue that without an attention schema – without model-based information about the most general, surface properties of attention, its dynamics as it transitions between states, its consequences, and a constantly updating representation of its current state – the endogenous control of attention would be difficult, if not impossible.

In many ways, the present results resemble classic examples of spatial adaptation in the motor domain. For example, the vestibulo-ocular reflex uses sensory input (vestibular signals that indicate head rotation) to guide a controlled output (counter-rotation of the eyes to keep the visual world stabilized). The spatial relationship between input and optimal output can be shifted with prism glasses, and the system will adapt to that offset such that the same vestibular input will evoke a spatially adjusted eye-movement output (Robinson, 1976). Reaching can be adapted to prism displacement in a similar manner (Harris, 1965; Martin et al., 1996). Normally, the visual target triggers a spatially accurate reach. The prism introduces a spatial offset, and the system learns a new transfer function between input and output. The general principle that reaching involves an internal model, and that the internal model can be adapted through training in the presence of a visual displacement or in the presence of a force field on the arm, has been intensively studied (Cunningham, 1989; Gandolfo et al., 1996; Krakauer, 2009; Mazzoni and Krakauer, 2006; Pine et al., 1996; Shadmehr and Moussavi, 2000; Shadmehr and Mussa-Ivaldi, 1994; Taylor et al., 2014; Thoroughman and Shadmehr, 2000; Thoroughman and Taylor, 2005). The spatial adaptation of saccadic eye movement has also been studied in similar ways (Doré-Mazars and Collins, 2005; Frens and Van Opstal, 1994).

In classical adaptation paradigms, awareness of the relevant stimuli is not necessary. For example, subjects adapt to a gentle forcefield applied to the arm, even though they are unaware of the forcefield. In the present experiments, awareness played a specific role: awareness of the cue was necessary for effective adaptation to the spatial shift. We suggest that awareness plays a specific role in adaptation in the attention domain: the reason is that a lack of awareness of a stimulus is tantamount to a lack of an internal model of the attention focused on that stimulus, which affects the control of attention in the same way that compromising the internal model of the arm would affect the control of the arm.

25.2. Three theories of awareness

In this final section, we briefly consider three different theories of awareness. Each theory is summarized and its ability to explain the attention-awareness relationship is assessed.

One prominent theory of awareness is the global workspace theory (GW) (Baars, 1988; Dehaene, 2014; Dehaene and Changeux, 2011). In GW, information is boosted and stabilized by exogenous or endogenous attention mechanisms until it reaches the global workspace, where it becomes available to many systems including speech, decision-making, movement control, and memory. In this perspective, information that has entered the global workspace has entered subjective awareness. Awareness corresponds to the highest level of attention in the brain, in which information has been stabilized or boosted to a threshold level, sometimes called ignition, that makes it available to many systems around the brain.

At least in its simplest form, GW predicts a relationship between attention and awareness that does not accommodate the data. If the

theory is correct, then awareness should be equivalent to the upper end of the attention range. It should never be possible to have two comparable situations, in one of which someone is aware of stimulus X and paying little attention to it, and in the other of which the person is unaware of X and paying more attention to it. If the amount of attention in the first situation is sufficient to produce awareness of X, then the larger amount of attention in the second situation should be too. But that relationship is not consistent with the data. In some tasks, awareness of a stimulus is associated with *less* attention drawn to it, and a lack of awareness of the stimulus is associated with *more* attention drawn to the stimulus (Tsushima et al., 2006). In other tasks, a lack of awareness of a stimulus is associated with a change in the time course of attention rather than an overall drop in the magnitude of attention (Webb et al., 2016a,2016b). The present findings also show that without awareness, exogenous attention is not necessarily of lower magnitude. Findings like these show that awareness is not simply the highest end of the range of attentional enhancement. Attention without awareness is not simply attention that is too weak to boost the stimulus into the global workspace. Something more complex relates attention to awareness, in a manner not fully captured by GW at least in its simplest form. We are not arguing that GW is ruled out by the data on attention and awareness. Rather, to be consistent with the data, GW would need some elaboration or added mechanism. Below we will suggest a possible addition that brings GW in line with the AST.

A second prominent theory of awareness is the higher-order thought theory (HOT) (Gennaro, 2012; Lau and Rosenthal, 2011; Rosenthal, 2006). HOT depends on the insight that the claim, “I am aware of the stimulus,” contains more information than the claim, “There is a stimulus.” In the theory, the claim of awareness requires higher-order information about one’s own internal processes in addition to lower-order information about the stimulus. Awareness derives from that higher-order representation. At least in its simplest form, HOT says nothing specific about how attention relates to awareness. Again, the theory would need extra elaboration to accommodate the findings on attention and awareness.

AST was specifically constructed to address the relationship between attention and awareness (Graziano, 2013, 2019; Graziano and Kastner, 2011; Webb and Graziano, 2015). In AST, when a person reports being conscious of X, it is because at least three conditions have been met. First, the person is focusing some attention on X. As a result, X can be processed in greater depth and can influence widespread systems around the brain such as decision-making, memory, and action output systems. This condition resembles the “ignition” condition in GW. Second, an attention schema depicts that state of attention on X. That depiction is efficient – detail-poor – and presents a picture of a non-physical, mental essence that is mentally grasping or experiencing. That meta representation, or representation of the process of attention, fits the same general category as the higher-order representations in HOT. Third, higher cognition has access to the larger set of information representing X and representing the state of attention, and based on that larger picture can report, “I have a conscious experience of X.” Though not originally described in these terms, AST could be viewed as a fusion of HOT and GW. It is arguably the simplest possible way to unify HOT and GW, in that it posits a higher-order representation of the global workspace. In AST, the brain attributes a subjective awareness to itself because that construct serves as a useful, if simplified, model of attention, especially the highest level of attention that most impacts behavior, the global workspace.

AST does a better job than GW or HOT at accommodating the data on awareness and attention. It specifically addresses why attention becomes less well regulated, rather than weaker, when awareness is absent. For example, why would the lack of awareness of a stimulus sometimes be associated with an *increase* in attention to it (Tsushima et al., 2006)? Neither HOT nor GW, by itself, gives a specific explanation. However, AST accounts for it. It is a straightforward case of a model being used for control, and the absence of the model leading to

poor control. In the task, the stimulus in question is a distractor, and to perform the task optimally, attention to it should be minimized. But that ability to regulate and minimize attention is compromised when subjects are unaware of the distractor. In another study, without awareness of a cue stimulus, attention to the cue was not overall smaller or larger in magnitude, but showed greater fluctuations over time, possibly reflecting a reduction in control (Webb et al., 2016a,2016b). In the present study, exogenous or stimulus-driven attention survived the lack of awareness of the cue, but the endogenous control of attention – shifting attention relative to where it was drawn to the cue – was drastically impaired.

The results across many experiments are therefore converging on a general pattern: without awareness, attention is still possible, and it is not necessarily overall reduced in magnitude, but it is significantly less well controlled. The magnitude of attention is less consistent, the controller is less able to suppress attention to distractors, and the controller’s ability to use contingencies to efficiently move attention is compromised. That pattern is broadly consistent with three central components of AST: first, the common human claim that we have an awareness inside of us derives from a specific information set constructed in the brain (or else we would not be able to make the claim); second, that information set serves as a detail-poor, but useful model of attention; and third, the model is used to enhance the control of attention. As a result, when awareness is compromised, the control of attention is compromised. The idea that attention might benefit from a control model was also supported by a recent computational study of an artificial attention system, in which the addition of a control model – an attention schema – enhanced the stability and efficiency of the system (van den Boogaard et al., 2017).

We are not arguing that the data rule out HOT or GW. Instead, we note that AST is a useful way to incorporate ideas from HOT, GW, and AST into a single framework, and the unified framework is able to accommodate the growing pattern of data on attention and awareness.

Acknowledgments

This work was supported by the Princeton Neuroscience Institute Innovation Fund. We thank Jordan Taylor for advice about testing spatial generalization, and Argos Wilterson for his enthusiasm.

References

- Ansorge, U., Heumann, M., 2006. Shifts of visuospatial attention to invisible (metacontrast-masked) singletons: clues from reaction times and event-related potentials. *Adv. Cognit. Psychol.* 2, 61–76.
- Baars, B.J., 1988. *A Cognitive Theory of Consciousness*. Cambridge University Press, New York.
- Brainard, D.H., 1997. The psychophysics toolbox. *Spat. Vis.* 10, 433–436.
- Chun, M.M., Golomb, J.D., Turk-Browne, N.B., 2011. A taxonomy of external and internal attention. *Annu. Rev. Psychol.* 62, 73–101.
- Cooke, D.F., Graziano, M.S.A., 2003. Defensive movements evoked by air puff in monkeys. *J. Neurophysiol.* 90, 3317–3329.
- Corbetta, M., Kincade, J.M., Ollinger, J.M., McAvoy, M.P., Shulman, G.L., 2000. Voluntary orienting is dissociated from target detection in human posterior parietal cortex. *Nat. Neurosci.* 3, 292–297.
- Cunningham, H.A., 1989. Aiming error under transformed spatial mappings suggests a structure for visual-motor maps. *J. Exp. Psychol. Hum. Percept. Perform.* 15, 493–506.
- Dehaene, S., 2014. *Consciousness and the Brain*. Viking Press, New York.
- Dehaene, S., Changeux, J.-P., 2011. Experimental and theoretical approaches to conscious processing. *Neuron* 70, 200–227.
- Dennett, D.C., 1991. *Consciousness Explained*. Little-Brown., New York.
- Doré-Mazars, K., Collins, T., 2005. Saccadic adaptation shifts the pre-saccadic attention focus. *Exp. Brain Res.* 162, 537–542.
- Frankish, K., 2016. Illusionism as a theory of consciousness. *J. Conscious. Stud.* 23, 1–39.
- Frens, M.A., van Opstal, A.J., 1994. Transfer of short-term adaptation in human saccadic eye movements. *Exp. Brain Res.* 100, 293–306.
- Gandolfo, F., Mussa-Ivaldi, F.A., Bizzi, E., 1996. Motor learning by field approximation. *Proc. Nat. Acad. Sci. U.S.A.* 93, 3843–3846.
- Gennaro, R., 2012. *The Consciousness Paradox: Consciousness, Concepts, and Higher-Order Thoughts*. MIT Press, Cambridge, MA.
- Graziano, M.S.A., 2013. *Consciousness and the Social Brain*. Oxford University Press, New York.

- Graziano, M.S.A., 2019. Attributing awareness to others: the attention schema theory and its relationship to behavioral prediction. *J. Conscious. Stud.* 26, 17–37.
- Graziano, M.S.A., Kastner, S., 2011. Human consciousness and its relationship to social neuroscience: a novel hypothesis. *Cogn. Neurosci.* 2, 98–113.
- Graziano, M.S.A., Guterstam, A., Bio, B.J., Wilterson, A.I., 2019. Toward a standard model of consciousness: reconciling the attention schema, global workspace, higher-order thought, and illusionist theories. *Cognit. Neuropsychol.* <https://doi.org/10.1080/02643294.2019.1670630>. published online.
- Guterstam, A., Kean, H.H., Webb, T.W., Kean, F.S., Graziano, M.S.A., 2018. An implicit model of other people's visual attention as an invisible, force-carrying beam projecting from the eyes. *Proc. Nat. Acad. Sci. U.S.A.* 116, 328–333.
- Haith, A.M., Krakauer, J.W., 2013. Model-based and model-free mechanisms of human motor learning. In: In: Richardson, M., Riley, M., Shockley, K. (Eds.), *Progress in Motor Control: Neural Computational and Dynamic Approaches*, vol. 782. Springer, New York, pp. 1–21.
- Harris, C.S., 1965. Perceptual adaptation to inverted, reversed, and displaced vision. *Psychol. Rev.* 72, 419–444.
- Howard, J.H., Howard, D.V., Dennis, N.A., Kelly, A.J., 2008. Implicit learning of predictive relationships in three-element visual sequences by young and old adults. *J. Exp. Psychol. Learn. Mem. Cogn.* 34, 1139–1157.
- Hsieh, P., Colas, J.T., Kanwisher, N., 2011. Unconscious pop-out: attentional capture by unseen feature singletons only when top-down attention is available. *Psychol. Sci.* 22, 1220–1226.
- Igelström, K., Graziano, M.S.A., 2017. The inferior parietal lobule and temporoparietal junction: a network perspective. *Neuropsychologia* 105, 70–83.
- Igelström, K., Webb, T.W., Kelly, Y.T., Graziano, M.S.A., 2016. Topographical organization of attentional, social and memory processes in the human temporoparietal cortex. *eNeuro* 3 ENEURO.0060-16.2016.
- Imamizu, H., Uno, Y., Kawato, M., 1995. Internal representations of the motor apparatus: implications from generalization in visuomotor learning. *J. Exp. Psychol. Hum. Percept. Perform.* 21, 1174–1198.
- Jiang, Y., Costello, P., Fang, F., Huang, M., He, S., 2006. A Gender- and Sexual Orientation-Dependent Spatial Attentional Effect of Invisible Images. *Proc. Nat. Acad. Sci. U.S.A.* 103, 17048–17052.
- Kelly, Y.T., Webb, T.W., Meier, J.D., Arcaro, M.J., Graziano, M.S.A., 2014. Attributing awareness to oneself and to others. *Proc. Nat. Acad. Sci. U.S.A.* 111, 5012–5017.
- Kentridge, R.W., Heywood, C.A., Weiskrantz, L., 1999. Attention without awareness in blindsight. *Proc. Biol. Sci.* 266, 1805–1811.
- Kentridge, R.W., Heywood, C.A., Weiskrantz, L., 2004. Spatial attention speeds discrimination without awareness in blindsight. *Neuropsychologia* 42, 831–835.
- Kleiner, M., Brainard, D., Pelli, D., 2007. What's new in Psychtoolbox-3? *Perception* 36, 1–16.
- Krakauer, J.W., 2009. Motor learning and consolidation: the case of visuo-motor rotation. *Adv. Exp. Med. Biol.* 629, 405–421.
- Lambert, A.J., Sumich, A.L., 1996. Spatial orienting controlled without awareness: a semantically based implicit learning effect. *Q. J. Exp. Psychol. A* 49, 490–518.
- Lambert, A., Naikar, N., McLachlan, K., Aitken, V., 1999. A new component of visual orienting: implicit effects of peripheral information and subthreshold cues on covert attention. *J. Exp. Psychol. Hum. Percept. Perform.* 25, 321–340.
- Lau, H., Rosenthal, D., 2011. Empirical support for higher-order theories of consciousness. *Trends Cogn. Sci. (Regul. Ed.)* 15, 365–373.
- Lin, Z., Murray, S.O., 2015. More power to the unconscious: conscious, but not unconscious, exogenous attention requires location variation. *Psychol. Sci.* 26, 221–230.
- Martin, T.A., Keating, J.G., Goodkin, H.P., Bastian, A.J., Thach, W.T., 1996. Throwing while looking through prisms. II. Specificity and storage of multiple gaze-throw calibrations. *Brain* 119, 1199–1211.
- Mazzoni, P., Krakauer, J.W., 2006. An implicit plan overrides an explicit strategy during visuomotor adaptation. *J. Neurosci.* 26, 3642–3645.
- McCormick, P.A., 1997. Orienting attention without awareness. *J. Exp. Psychol. Hum. Percept. Perform.* 23, 168–180.
- Merikle, P.M., Smilek, D., Eastwood, J.D., 2001. Perception without awareness: perspectives from cognitive psychology. *Cognition* 79, 115–134.
- Metzinger, T., 2009. *The Ego Tunnel: The Science of the Mind and the Myth of the Self*. Basic Books, New York.
- Mitchell, L.P., 2008. Activity in the right temporo-parietal junction is not selective for theory-of-mind. *Cereb. Cortex* 18, 262–271.
- Pesquita, A., Chapman, C.S., Enns, J.T., 2016. Humans are sensitive to attention control when predicting others' actions. *Proc. Nat. Acad. Sci. U.S.A.* 113, 8669–8674.
- Pine, Z.M., Krakauer, J.W., Gordon, J., Ghez, C., 1996. Learning of scaling factors and reference axes for reaching movements. *Neuroreport* 7, 2357–2361.
- Posner, M.I., 1980. Orienting of attention. *Q. J. Exp. Psychol.* 32, 3–25.
- Posner, M.I., Rafal, R.D., Choate, L.S., Vaughan, J., 1985. Inhibition of return: neural basis and function. *Cogn. Neuropsychol.* 2, 211–228.
- Robinson, D.A., 1976. Adaptive gain control of the vestibulo-ocular reflex by the cerebellum. *J. Neurophysiol.* 39, 954–969.
- Rosenthal, D., 2006. *Consciousness and Mind*. Oxford University Press, Oxford, UK.
- Salvucci, D.D., Goldberg, J.H., 2000. Identifying fixation and saccades in eye-tracking protocols. *Proceedings of the 2000 Symposium on Eye-Tracking Research and Applications* 71–78.
- Saxe, R., Kanwisher, N., 2003. People thinking about thinking people: fMRI investigations of theory of mind. *Neuroimage* 19, 1835–1842.
- Shadmehr, R., Moussavi, Z.M., 2000. Spatial generalization from learning dynamics of reaching movements. *J. Neurosci.* 20, 7807–7815.
- Shadmehr, R., Mussa-Ivaldi, F.A., 1994. Adaptive representation of dynamics during learning of a motor task. *J. Neurosci.* 14, 3208–3224.
- Shulman, G.L., Pope, D.L., Astafiev, S.V., McAvoy, M.P., Snyder, A.Z., Corbetta, M., 2010. Right hemisphere dominance during spatial selective attention and target detection occurs outside the dorsal frontoparietal network. *J. Neurosci.* 30, 3640–3651.
- Taylor, J.A., Krakauer, J.W., Ivry, R.B., 2014. Explicit and implicit contributions to learning in a sensorimotor adaptation task. *J. Neurosci.* 34, 3023–3032.
- Thoroughman, K.A., Shadmehr, R., 2000. Learning of action through adaptive combination of motor primitives. *Nature* 407, 742–747.
- Thoroughman, K.A., Taylor, J.A., 2005. Rapid reshaping of human motor generalization. *J. Neurosci.* 25, 8948–8953.
- Travers, E., Frith, C.D., Shea, N., 2018. Learning rapidly about the relevance of visual cues requires conscious awareness. *Q. J. Exp. Psychol.* 71, 1698–1713.
- Tsushima, Y., Sasaki, Y., Watanabe, T., 2006. Greater disruption due to failure of inhibitory control on an ambiguous distractor. *Science* 314, 1786–1788.
- Vallar, G., Perani, D., 1986. The anatomy of unilateral neglect after right-hemisphere stroke lesions. A clinical/CT-scan correlation study in man. *Neuropsychologia* 24, 609–622.
- van den Boogaard, E., Treur, J., Turpin, M., 2017. A neurologically inspired neural network model for graziano's attention schema theory for consciousness. *International Work Conference on the Interplay Between Natural and Artificial Computation: Natural and Artificial Computation for Biomedicine and Neuroscience*, Part 1 10–21.
- Webb, T.W., Graziano, M.S.A., 2015. The attention schema theory: a mechanistic account of subjective awareness. *Front. Psychol.* 6 article 500.
- Webb, T.W., Kean, H.H., Graziano, M.S.A., 2016a. Effects of awareness on the control of attention. *J. Cogn. Neurosci.* 28, 842–851.
- Webb, T.W., Igelström, K., Schurger, A., Graziano, M.S.A., 2016b. Cortical networks involved in visual awareness independently of visual attention. *Proc. Nat. Acad. Sci. U.S.A.* 113, 13923–13928.