# Attention as Inference: Selection Is Probabilistic; Responses Are All-or-None Samples

Edward Vul, Deborah Hanus, and Nancy Kanwisher
Massachusetts Institute of Technology

Theories of probabilistic cognition postulate that internal representations are made up of multiple simultaneously held hypotheses, each with its own probability of being correct (henceforth, "probability distributions"). However, subjects make discrete responses and report the phenomenal contents of their mind to be all-or-none states rather than graded probabilities. How can these 2 positions be reconciled? Selective attention tasks, such as those used to study crowding, the attentional blink, rapid serial visual, and so forth, were recast as probabilistic inference problems and used to assess how graded, probabilistic representations may produce discrete subjective states. The authors asked subjects to make multiple guesses per trial and used 2nd-order statistics to show that (a) visual selective attention operates in a graded fashion in time and space, selecting multiple targets to varying degrees on any given trial; and (b) responses are generated by a process of sampling from the probabilistic states that result from graded selection. The authors concluded that although people represent probability distributions, their discrete responses and conscious states are products of a process that samples from these probabilistic representations.

*Keywords:* attention, cognition, probabilistic inference

Physical constraints prevent us from producing multiple different actions at once; action is necessarily all-or-none. No matter how unsure we are about whether to turn left or right, we can only move in one direction. And no matter how unsure we are of our beliefs, we can only vocalize a single utterance. This all-or-none constraint on human action is so obvious that researchers often build it into real-world decision procedures (e.g., voting) and design their experiments around it (e.g., N-alternative forced choice). It is not only our actions but also our conscious states that seem to be all-or-none: A Necker cube appears to be in one configuration or another, never in both simultaneously. Researchers have attempted to circumvent all-or-none reporting constraints by using Likert scales to tap into graded phenomenal experience. But even when people are asked to report graded degrees of awareness, they use the available scale in an all-or-none fashion, reporting that they are either aware or not aware, rarely "half-aware" (Sergent & Dehaene, 2004).

Such introspections have resulted in many "all-or-none" accounts of cognitive representation. We consider all-or-none representations to be those that consist entirely of Boolean-valued beliefs, that is, beliefs that are either true or false, but not in-

between. In choices from multiple discrete options, one or more options may be deemed true and the others false. An object either belongs to a category or it does not; a signal has either passed threshold, or it has not. In choices along continuously valued dimensions (e.g., brightness), all-or-none representations take the form of point estimates (e.g., 11.3 trolands). Although the content of a point estimate is continuous (11.3), its truth value is all-or-none (e.g., "It is true that the brightness of the signal was 11.3 trolands"). Such all-or-none accounts of mental representation have been postulated for signal detection (point estimates corrupted by noise; e.g., Green & Swets, 1966), memory (memory traces as point estimates; e.g., Kinchla & Smyzer, 1967), and for concepts and knowledge (as logical rules and Boolean-valued propositions; e.g., Bruner, Goodnow, & Austin, 1956).

However, other theoretical perspectives treat mental representations as probability distributions in which multiple alternative hypotheses are held simultaneously, each with a different graded truth probability. According to one recent framework for modeling cognition, mental tasks can be optimally solved by Bayesian inference (Chater & Oaksford, 2008; Chater, Tenenbaum, & Yuille, 2006). Indeed, a variety of experiments show that human behavior often reflects this optimality, which implies that people are doing something like Bayesian inference (Chater, Tenenbaum, & Yuille, 2006; Kersten & Yuille, 2003; Steyvers, Griffiths, & Dennis, 2006). Implicit in the claim that people perform Bayesian inference is the idea that human cognitive machinery operates over probability distributions that reflect the uncertainty of the world (Chater et al., 2006; Griffiths & Tenenbaum, 2006). Representations of probability distributions are not all-or-none Boolean values, but rather graded probabilities: Every possible decision (left or right), estimate (amount of light present), or state (Necker cube tilted up or down) is assigned a probability that may be any value

**Fn1** between 0 and 1.[1] Probabilistic accounts have been proposed for memory (Steyvers et al., 2006), signal detection (Whiteley & Sahani, 2008), categorization (Tenenbaum, 1999), and knowledge (Kemp, Bonawitz, Coley, & Tenenbaum, 2008; Vul & Pashler, 2008).

Although these probabilistic accounts have recently gained much favor in cognitive science for their mathematical elegance and predictive power (Chater & Oaksford, 2008), they conflict with the common intuition that conscious access is all-or-none. How can we have both probabilistic representations and seemingly all-or-none conscious experience?

We tackle the conflict between all-or-none subjective experience and probabilistic accounts of representations within the domain of visual selective attention. This domain is an ideal testing ground for several reasons. First, irrespective of debates about cognitive representation broadly construed, the representation underlying visual selective attention has been disputed, with some postulating all-or-none Boolean representations (Huang & Pashler, 2007; Huang, Treisman, & Pashler, 2007) and others suggesting graded representations (Reeves & Sperling, 1986; Shih & Sperling, 2002). Second, probing the fine line between conscious access and unconscious representation requires a domain that examines that interface: Although the link between conscious access and visual attention has long been discussed and debated (Baars, 1997; Koch & Tsuchiya, 2007; Lamme, 2003; Posner, 1994), the only clear consensus is that they are closely related. Finally, visual selective attention tasks are appealing because they afford precise manipulations and rigorous psychophysical measurements.

Thus, we used visual selective attention tasks here to study internal (short-term memory) representations and how subjects use them. First, we provide a theoretical framework, casting a large class of attentional selection tasks as problems of inference under uncertainty. We then describe experiments that test whether visual selective attention produces all-or-none representations, or graded representations, akin to the probability distributions implicated in Bayesian inference. Our evidence supports the latter view and suggests that conscious responses constitute all-or-none samples from these probability distributions.

## Visual Selective Attention

The term *visual attention* encompasses many disparate phenomena sharing the feature that people can selectively distribute resources among the elements of the visual world (e.g., memory: Chun & Potter, 1995; Vul, Nieuwenstein, & Kanwisher, 2008), perceptual fidelity (Carrasco, 2006; Posner, Snyder, & Davidson, 1980), feature integration (Treisman & Schmidt, 1982), and object formation (Kahneman, Treisman, & Gibbs, 1992). Here, we considered a class of tasks in which subjects are directed by a cue to select one or more elements for subsequent report (thus allotting memory capacity preferentially to some items over others). In a classic example of such a task, people are presented with a rapid serial visual (RSVP) stream of letters. One is cued (e.g., by virtue of being surrounded by an annulus), and the subject must select that letter, remember its identity, and report that letter identity later. Similarly, in spatial selective attention tasks, an array of letters may be presented in a ring around fixation, with one of them **F1** cued for subsequent report by a line (see Figure 1).
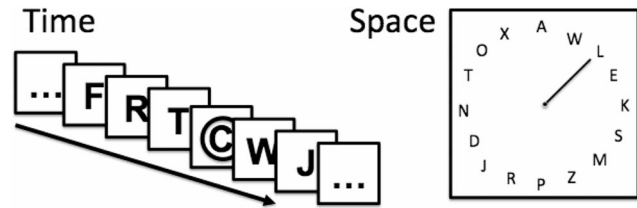


*Figure 1.* Prototypical experimental paradigms used in attentional selection tasks in time and space. One of many items is cued, and the subject must report that item.

Such tasks have been used to study the attentional blink (Chun, 1994, 1997; Chun & Potter, 1995; Raymond, Shapiro, & Arnell, 1992; Nieuwenstein & Potter, 2006; Vul, Hanus, & Kanwisher, 2008; Vul et al., 2008), crowding (He, Cavanagh, & Intriligator, 1996; Pelli, Palomares, & Majaj, 2004; Strasburger, 2005), illusory conjunctions (Prinzmetal, Henderson, & Ivry, 1995; Prinzmetal, Ivry, Beck, & Shimizu, 2002), change detection (Landman, Spekreijse, & Lamme, 2003), and short-term memory (e.g., partial report; Averbach & Coriell, 1961). In these experiments, researchers measure which items were reported and infer the properties of attentional selection (e.g., when it fails, and what its limits are). Rather than investigating the limits of attention in such tasks, here we are primarily concerned with the output of the selection process: the representation in short-term memory that attentional selection creates on any given trial of such an experiment.

Two main classes of theories address the issue of the representation that attention produces when selecting a particular object or region for storage in memory and subsequent report. According to one theory, items are selected through an *attentional gate* that defines a weighting function in space and time (Shih & Sperling, 2002). Therefore, on this account, the short-term memory representation resulting from selection is a weighted list, with items closer to the cue receiving a higher weight and those further from the cue receiving a lower weight. A contrasting recent theory postulates that items are selected by a *Boolean map* that defines some spatial regions as wholly selected and others as not selected, but does not include graded weights, or "half-selected" regions (Huang & Pashler, 2007). Therefore, on this account, the representation of possible items in short-term memory should be Boolean—an item will be either within the selected region, and remembered, or outside the selected region, and forgotten as a nontarget.

## Selective Attention as Inference Under Uncertainty

Theories of attention are usually cast at an algorithmic level, but it is also useful to consider Marr's (1982) computational theory

---

[1] There are, of course, other frameworks that have postulated graded representations. For instance, the level of activity in neural networks is graded. As we describe in the *Discussion* section, these other accounts are not mutually exclusive with representations of probability distributions, and may indeed be the algorithmic or neural implementations of the computational elements proposed by probabilistic models of cognition. Although throughout this article we motivate our experiments, and frame our results, at the computational Bayesian level, this need not preclude interpretations of the same results at the levels of algorithms or implementation.

level of explanation by asking what are the problems being solved in these tasks. Bayesian inference provides a useful framework, relating attentional selection to probabilistic cognition. Several groups have recently posed Bayesian accounts for mechanisms of attentional enhancement (Yu & Dayan, 2005), deployment of attentional enhancement or eye movements (Itti & Baldi, 2006; Najemnik & Geisler, 2005), or the integration of top-down influences with bottom-up saliency maps (Mozer, Shettel, & Vecera, 2005). Here, we apply the probabilistic approach to the "attentional selection" tasks we have discussed and cast these tasks in terms of inference under uncertainty.

What problem is being solved by visual selective attention in these tasks? Specifically, we want to know what the output of the attention mechanism ought to be given the nature of the problem. In a typical experiment on attentional selection, that problem entails reporting one feature or object (a "target"; e.g., the letter identity, *A*) that is distinguished from distractor items by some "cue" (e.g., an annulus)—a stimulus that identifies the spatial or temporal location of the target (see Figure 1). The spatiotemporal location is simply one or more dimensions along which different items are arrayed. Thus, attentional selection tasks amount to assessing which of the potential targets spatially or temporally co-occurs with the cue and then allocating short-term memory on the basis of the solution. To make the task challenging such that informative patterns of failure may be observed, the experimenter controls the discriminability of possible targets in time or space by taxing the system in different ways (e.g., close spatial or temporal packing of targets, brief display durations, etc). These conditions introduce spatial and temporal uncertainty about the locations of each possible target as well as the cue.

The subject's task, then, is to determine which target coincided with the cue, given some uncertainty about the spatiotemporal loca-tions of both the target and the cue. This task may therefore be considered inference under uncertainty, which is optimally solved by Bayesian inference. Given particular levels of uncertainty, the Bayesian solution to this problem entails, for each item and point in time, multiplying the probability that the letter occurred at that point in time by the probability that the cue occurred at that point in time and then integrating over time to obtain the probability that this letter coincided with the cue. The solution to this co-occurrence detection problem is a probability distribution over items, describing the likelihood that each item coincided with the cue. If this description is correct, and attentional selection is indeed solving the inference problem just described, then it should produce probability distributions over items likely to be the target (see Figure 2). We tested whether people **F2** represent such probability distributions over items in short-term memory.

## Within-Trial Gradation, Across-Trial Noise, and Representation

The typical experimental design in cognitive psychology precludes researchers from determining whether internal representations were all-or-none or graded on any one trial. The problem is caused by averaging across trials and subjects (e.g., Estes, 1956). Consider the task of reporting a cued letter from an RSVP sequence of letters. Subjects will not report the target correctly on all trials but will sometimes instead report the letter before or after the target, or occasionally another letter even farther away in the RSVP sequence (Botella, Garcia, & Barriopedro, 1992; Kikuchi, 1996). A histogram of such reports across trials will show a graded variation in the tendency to report items from each serial position (see Figure 3, bottom row) as expected, given the uncertainty **F3** inherent in the task.
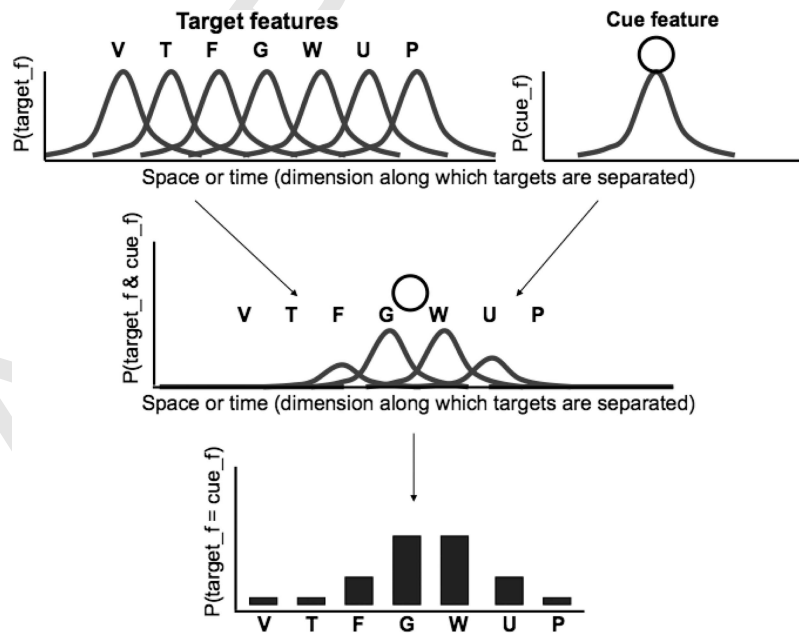


*Figure 2.* Selective attention as inference. Typical experimental paradigms make the task difficult by capitalizing on uncertainty in the spatial or temporal position of the targets and cues. The task thus amounts to inferring which target was likely to have co-occurred with the cue. A Bayesian solution to this task would result in a probability distribution over possible targets.
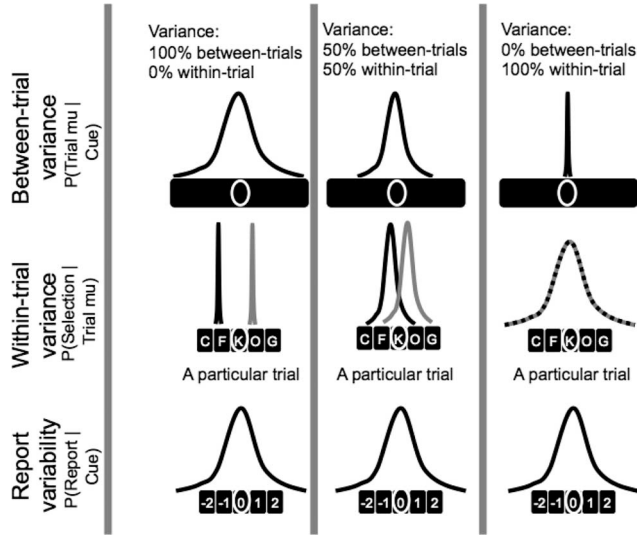
*Figure 3.* The final distribution of reports across trials is a combination of across-trial variation in where/when attention is deployed, and the within-trial gradation in the extent to which an item is selected on a given trial.

It is tempting to interpret this graded variation as indicating that selection itself is graded (Botella & Eriksen, 1992; Reeves & Sperling, 1986; Weichselgartner & Sperling, 1987). However, this conclusion does not follow, because variation in the items reported might reflect not gradations in the degree to which each item is selected on any given trial, but rather variation *across trials* in which items are selected. That is, the graded across-trial averages are consistent with the possibility that on each trial, subjects select items in an all-or-none-fashion, but which items are selected varies across trials due to variability, or noise, in the deployment of attention. This distinction is analogous to the classic dichotomy in signal detection theory: Is the variability in whether a stimulus is reported as visible due to *noise* that varies across trials (Green & Swets, 1966; Nieuwenstein, Chun, & van der Lubbe, 2005) or *uncertainty* that is represented on every trial (Vul & Pashler, 2008; Whiteley & Sahani, 2008)? Thus, across-trial histograms are not indicative of the properties of selection on any given trial.

Logically, the observed distribution of reports across trials is the combination of the across-trial variance and the within-trial gradation of selective attention. Figure 3 shows a few of these possibilities if the within-trial spread and across-trial gradation are both Gaussian. *Within-trial gradation* refers to the properties of selection on any one trial, that is, the representation in short-term memory resulting from selection. Across-trial variance, however, corresponds to the properties of this representation that change across trials. That is, given the within-trial distribution of selection on any given trial, how does it vary from one trial to the next (due to noise or to other factors)?

There are an infinite number of plausible combinations of within-trial gradation and across-trial variability in selection that could produce the same final pattern of results. The experiments presented in this article rule out many of these possibilities, but a few alternatives will remain. Before describing our experiments, it is worth laying out a few qualitatively different cases (see Figure 4).   **F4**
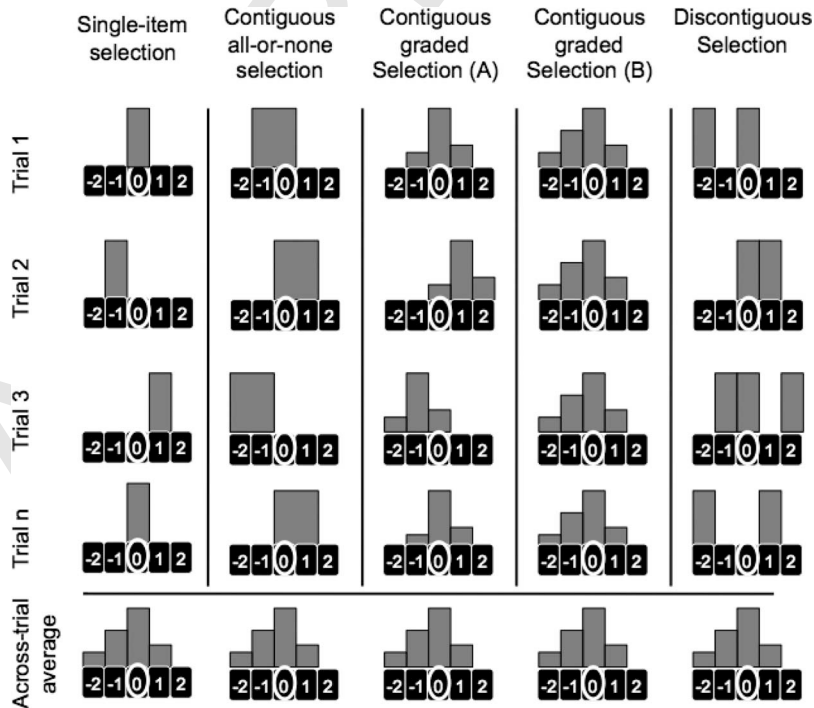


*Figure 4.* Several possible modes in which attentional selection may be operating to produce across-trial variation. Option 4 is consistent with subjects representing internal uncertainty as a probability distribution.

### Single-Item Selection

The simplest alternative is that subjects select and store in short-term memory only one letter per trial. For our purposes, this representation is essentially all-or-none, because only one letter is stored. On this account, the gradation in the frequency that a particular item is reported across all trials corresponds entirely to across-trial variability, which may be characterized as translation of the single-item selection window.

### Contiguous All-or-None Selection

Subjects may select multiple spatiotemporally contiguous items in an all-or-none fashion (an item is either selected, and stored in memory, or not). On this account, there must be across-trial variance to produce a *graded* distribution of final reports in the across-trial average. Furthermore, because the selected items are contiguous, this variability can only take the form of spatiotemporal translation of the selection window.

### Contiguous-Graded Selection

The alternative we advocate is that multiple contiguous items are selected, but the degree of "selection" varies across items on a given trial. This defines a weighting function over items, which may be described as a graded "attentional gate" (Shih & Sperling, 2002), or a probability distribution. On this account, there may also be a spatiotemporal translation from trial to trial (as shown in Figure 3), but that is not necessarily the case because the gradation of the selection window in this case may perfectly match the gradation seen in the final distribution of reports, as would be the case if there were no variability across trials.

### Complex Selection

One final possibility is that the items selected on a given trial need not be in a fixed spatiotemporal relationship (e.g., contiguous), as would be the case if several independent *single-item selection* episodes occurred on each trial. If this is the case, then across-trial variability is not constrained to be a mere translation of the attention window: It may take any form (e.g., on Trial A, Items –2, 0, and 1 are selected; on Trial B, Items –1 and 0 are selected). This account could describe any set of data because both (a) all other accounts are its subsets, and (b) it allows for as many degrees of freedom as there are data points. A similarly complicated account is that subjects may select a *variable number* of contiguous items; thus, the final response distribution will be a mixture of uniform distributions defined over different intervals of the presented items—this account also is sufficiently unconstrained that it could account for any pattern of data (by postulating specific proportions for different components of the mixture). As we describe later, our data suggest a more parsimonious account. For these reasons, we do not consider these alternatives again until the *Discussion* section.

### Within-Trial Representations, Attention, and Probability

The alternatives above propose different amounts of "within-trial gradation" and "across-trial variability" of representations. Within-trial gradation (rather than across-trial variability) of rep-

resentations has implications for selective attention, as well as probabilistic representation more broadly. First, within-trial gradation can determine whether selective attention operates in a graded or discrete fashion. Evidence for any amount of within-trial gradation of selection would conflict with recent theories of spatial selection that suggest that selection operates as a Boolean map, selecting regions of space in an all-or-none fashion (Huang & Pashler, 2007). However, evidence for any amount of across-trial variability in selection would call into question previous research using the distribution of reports across trials to infer the properties of selection on any one trial (Shih & Sperling, 2002).

Second, within-trial variability is also a measure of how people represent uncertainty on any given trial. A substantial amount of within-trial variability implies that subjects represent the uncertainty inherent in a particular task on every trial. This finding would suggest that internal representations may indeed be probability distributions. However, if we find only across-trial variability in reports, our results would suggest that many previous results showing that responses follow probability distributions appropriate to the inference in question may be an artifact of averaging across people or trials—the probability distributions exist across individuals or time, but not within one individual at a specific point in time (Mozer, Pashler, & Homaei, 2008).

Most important, if we find that attention operates in a graded fashion, then the results will have ramifications beyond the realm of visual selective attention to the nature of perceptual awareness. Introspection, as well as some data, suggests that awareness is discrete: We are either aware of something or we are not. Sergent and Dehaene (2004) tested this intuition by asking subjects to provide ratings of their degree of awareness of the target item in an "attentional blink" (Raymond et al., 1992) paradigm. Subjects reported bimodal degrees of visibility: Sometimes the target was rated as completely visible, sometimes completely invisible, but participants rarely provided intermediate ratings. These results suggest that conscious access may be a discrete phenomenon. A similar conclusion was reached from studies of the wagon wheel illusion under continuous light. In movies, a rotating wagon wheel can appear to move backwards due to aliasing arising from the discrete sampling frequency of the movie frames. Because the wagon-wheel illusion can be seen under continuous light, some have argued that perception is discrete: The wagon wheel moves backwards due to aliasing arising from discrete sampling of percepts from the environment (VanRullen & Koch, 2003). Given these findings, if the present studies find that selective attention is continuous, in that it produces graded representations, then this fact must be reconciled with the apparent all-or-none nature of conscious awareness.

In the experiments reported here, we measure the across-trial variance and within-trial spread of selection by asking subjects to make multiple responses on a given trial: Subjects first report their best estimate of the item that was cued and then make additional guesses about which item was cued. This method has been used previously in research on signal detection theory (Swets, Tanner, & Birdsall, 1961) and more recently to study representations of knowledge (Vul & Pashler, 2008). As in this previous literature, we consider the relationship between errors on the first response and on the second response. In our case, we consider the position of items reported in a selective attention task and evaluate whether two items reported on one trial are independent (as predicted if

they are samples from a probability distribution) or whether they share some variance (as predicted from across-trial variability). For example, if subjects incorrectly report an item appearing earlier in the RSVP list as the target, will a second guess from same trial likely be another item that appeared early in the list? If so, then there is some common error for the trial shared across guesses, indicating that there is some across-trial variability in which items are selected (thus, giving rise to a graded final distribution of reports). If the temporal positions of the intrusions reported in the two guesses are not correlated, then there is no common, shared error for a given trial, and the final distribution of reports is driven entirely by within-trial variability.

For *single-item selection*, we do not expect to find information in both guesses (even if the subject postpones reporting the selected item until the second guess, there will be no systematic relationship between the items reported on Guess 1 and Guess 2). For *contiguous all-or-none selection* to produce a graded final distribution of reports, variability must exist in the position of the selection window across trials. This translation would necessarily induce a correlation in the errors of two responses, and thus the contiguous all-or-none selection account mandates a correlation. Only the *contiguous-graded selection* account can produce a graded final distribution of reports without any across-trial variation (and thus correlation of errors).

Thus, we test for within- and across-trial variability of temporal selection in Experiment 1 and of spatial selection in Experiment 2. In both cases, we found that the there is no correlation in the temporal or spatial position of intrusions from multiple responses on one trial. This finding indicates that there is no across-trial variability, and therefore, the average distribution of final reports reflects the gradation of selection on any given trial. Thus, selection is continuous and graded, whereas responses act as samples from the graded representation. Our data indicate that attention selects a number of items to varying degrees on any given trial, creating a probability distribution over likely targets, and subjects make responses and subjective judgments by sampling items from the selected distribution while having no conscious access to the distribution itself.

## Experiment 1

First, we tested whether selective attention is graded: Are multiple items selected to varying degrees on a given trial, and does this within-trial spread of selection underlie the commonly observed final distribution of reports? Commonly adopted experiments with single-probe trials do not provide enough information to dissociate across-trial variance and within-trial gradation. To assess the spread of the items selected by attention on a given trial, we asked subjects to make four guesses about the identity of the cued target. By analyzing the distributions of subsequent guesses conditioned on the first guess, we can estimate the spread of selection within a given trial.

### Method

*Participants.*    Nine subjects from the Massachusetts Institute of Technology subject pool were recruited to participate. Subjects were between 18 and 35 years of age and were paid $10 for participation.

*Materials.*    On each trial, subjects saw an RSVP stream composed of one instance of each of the 26 English letters in a random order. Each letter was presented for 20 ms and was followed by a 47-ms blank (three and seven frames at a 150-Hz refresh rate, respectively), resulting in an RSVP rate of 15 items/s. Letters were white on a black background, capitalized, in size 48 Courier font. With our resolution (1024 × 768), monitor (Viewsonic G90f, ViewSonic, Walnut, CA), and viewing distance (roughly 50 cm), letters subtended roughly 2.5° visual angle.

On each trial, one cue appeared in the RSVP stream to indicate which of the letters was the designated target. The cue was a white annulus with an inner diameter of 2.8° and an outer diameter of 3.2°; thus, the cue appeared as a ring around the RSVP letter sequence. When a cue appeared, it was shown in the 47-ms blank interval between two letters (see Figure 5).    **F5**

Onset of the cue was randomly counterbalanced to appear either before the 6th, 8th, 10th, 12th, 14th, 16th, or 18th letter of the sequence. Subjects were asked to report whatever letter appeared immediately after, or at the same time as, the cue. The experiment was programmed in PsychToolbox (Brainard, 1997) on MATLAB 7 on a Windows XP computer.

*Procedure.*    Each subject began the experiment with two practice trials; the results of these trials were discarded. Following the practice trials, subjects completed three blocks of 70 trials each. Each block contained 10 instances of each of the seven possible cue onset positions, in a random order for each block.

At the end of each trial, subjects were asked to make four guesses about which letter they thought was cued by the annulus. Subjects reported the letters by pressing the corresponding keys on the keyboard. Duplicate letter reports were not accepted; thus, each guess was a unique letter.

Subjects were told that they would get 1 point if they reported the letter correctly on the first guess, 0.5 points on the second guess, 0.25 points on the third guess, and 0.125 points on the fourth guess. Feedback and scoring on each trial reflected this instruction. To motivate subjects to perform well on this task, in addition to the flat rate of $10 for participation, subjects were offered bonus cash awards for performance: $0.01 for each point scored (on average, subjects scored 160 points in a given session: $1.60 bonus).
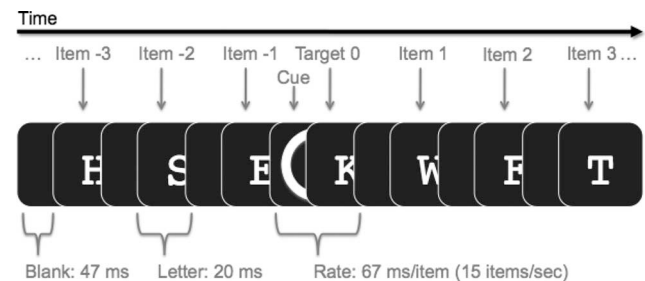


*Figure 5.*    Experiment 1 design: Subjects were asked to report one item cued in a rapid serial visual presentation, but were asked to make four guesses about the identity of that item. The decision to display the cue before, rather than during, the target was irrelevant to our main analysis of the relationship between multiple guesses on a single trial, and was done to match performance with some unrelated previous work (Vul, Hanus, & Kanwisher, 2008).

*Results*

Because there were no repeated letters on any trial, we could identify the exact serial position of the reported letters. From this information, we computed the distribution of guessed letters around the presented cue.

**F6**  Figure 6 (left panel) shows the empirical frequency with which a letter from each serial position was reported as a function of distance from the cue. That is, an *x* value of 0 corresponds to the cued letter (target); an *x* value of −1 is the letter that preceded the target; and an *x* value of 1 is the letter that followed the target. This is shown for each of the four guesses. The distribution of first guessed serial positions shows a precue intrusion pattern; that is, items preceding the cue (negative serial positions) are reported more often than items after the cued letter (positive serial positions). Effects such as this have been reported before under certain conditions (Botella et al., 1992; Kikuchi, 1996); presumably in our data, these effects are increased because the cue actually appears between the preceding distractor and the target.

Serial positions that are reported above chance may be identified in the right panel of Figure 6 as those points with log-likelihood ratios (log of the empirical frequency divided by chance frequency) above 0 (significance may be ascertained by the error bars, which correspond to 1 standard error of the mean, across subjects). These log-likelihood ratios for Guesses 2–4 suggest these guesses have roughly the same distribution of reports as the first guesses, given that the peak (Position 0, target) could not be guessed twice. However, this distribution also has an ever-increasing admixture of random chance reports. Because Guesses 3 and 4 are at, or close to, chance, all of our subsequent analyses look at just Guesses 1 and 2.

The fact that Guess 2 is above chance would seem to rule out the possibility of single-item selection because to have a reliable second guess, subjects must have selected more than one letter. This conclusion may also appear to follow from the observation that subjects produce a similar distribution on Guess 2 as Gess 1. However, these facts do not indicate how much of the variance seen in the distribution of reports on Guess 1 (what is normally measured in such tasks) is attributable to across-trial variance and within-trial gradation. This pattern of results may also arise if, on any given trial, subjects select one and only one letter, but on some trials, subjects pressed the wrong key on Guess 1 and responded with the actual selected letter on Guess 2 (or Guess 3, or Guess 4), thus raising performance on those subsequent guesses above chance. To determine whether this was the case for the second guess, we can look at the distribution of second-guess reports relative to the serial position of the first-guess report. If subjects only select one letter per trial, and either report it on the first or the second guess, then there should be no reliable relationship between the serial position of the first guess and the serial position of the second guess.

Figure 7 shows the frequency of Guess 2 reports as a function of  **F7** serial position distance from the letter reported on Guess 1. These data show that the second guess is likely to come from one of the four serial positions nearest to the first guess (these serial positions are reported above chance: all four *t*s(8) > 3.3, *p*s < .005). This indicates that subjects must be selecting at least two letters in proximal serial positions on any given trial. This pattern of results cannot arise from the single-item selection account, in which one and only one letter is selected on a single trial. Thus, we can say that multiple items are selected on each trial.

We must ascertain how much of the variance in reports arises from across-trial variability to assess whether the items selected on a given trial are selected in an all-or-none or in a graded fashion (the contiguous all-or-none and contiguous-graded selection accounts). A graded tendency to report particular serial positions across trials must arise from across-trial variability if selection takes the form of an all-or-none contiguous block on any given trial. However, if selection on a given trial may be graded, then there need not be across-trial variability to produce a graded
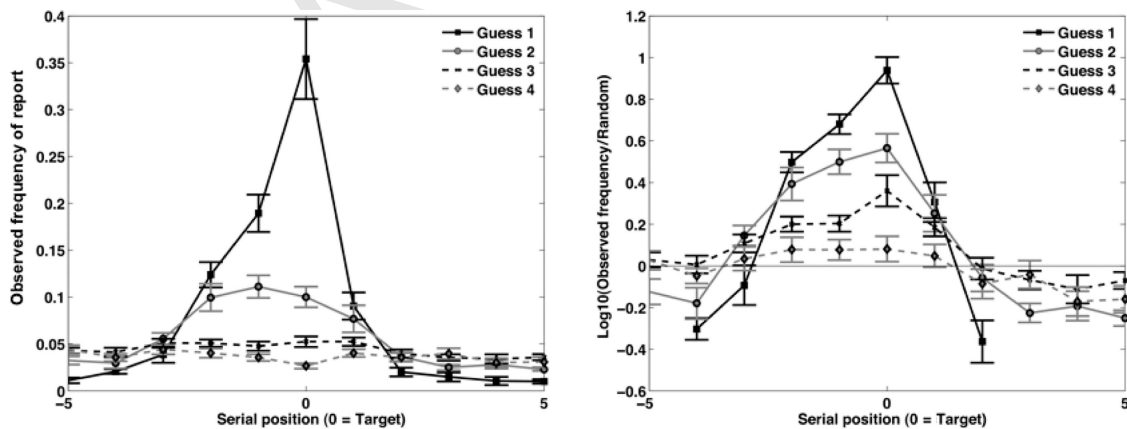


*Figure 6.* Experiment 1 results. The frequency that a given serial position is reported for each of the four guesses subjects make on each trial appears in the left panel. The log ratio of the empirical frequency of report, compared with the frequency of chance report appears in the right panel: This quantity effectively corrects for the decreasing chance rate of reporting particular items (if they are reported on prior guesses) and corresponds to how much above chance a particular serial position was reported on each guess. Error bars show 1 *SEM* across subjects.
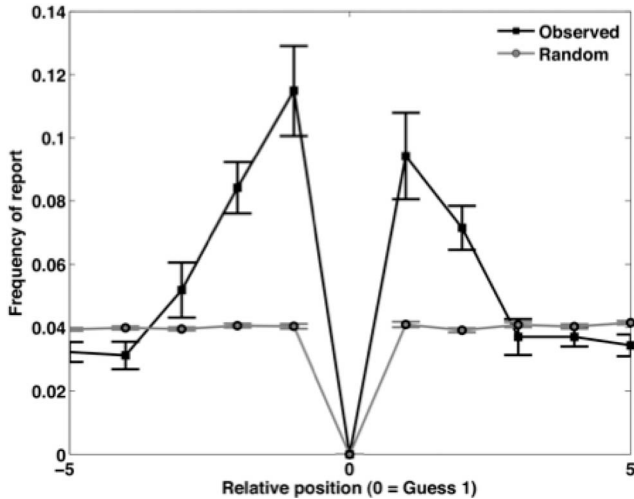
*Figure 7.* Experiment 1 results: The serial position of the item reported on Guess 2 relative to the serial position of the item reported on Guess 1. Subjects are likely to report two temporally adjacent letters on Guesses 1 and 2, indicating that multiple letters are selected on any given trial. Error bars show 1 *SEM* across subjects.

across-trial report frequency. Thus, the contiguous all-or-none account predicts a substantial amount of across-trial variability, as this is the only way that a graded distribution of errors may arise in the across-trial average.

To measure across-trial variability, we exploited the idea that across-trial variance in the form of temporal translation of the selected region should affect Guess 2 reports and Guess 1 reports similarly, such that Guess 2 reports should depend on the serial position of Guess 1 reports. If there is zero across-trial variance, then all guesses are sampled from the same distribution, which corresponds to the degree to which each letter is selected on every trial. Therefore, regardless of the absolute serial position of Guess 1, the distribution of absolute serial positions of Guess 2 should be unchanged. However, if there is substantial across-trial variance, then the guesses will be sampled from different distributions on different trials. Thus, on trials when Guess 1 was reported as, for example, the item two letters before the cue ($-2$), the distribution of reported Guess 2 serial positions should shift toward $-2$ (as it is sampled from the same uncentered distribution as Guess 1). Figure 8 provides an illustration of this *conditional-response dis-* **F8** *tribution* logic. Thus, we can estimate the across-trial variance by testing whether the distribution of Guess 2 reports is independent of Guess 1 reports.

Figure 9 displays this conditional-report distribution analysis: **F9** the distribution of Guess 2 reports conditioned on the serial position of Guess 1 reports. These conditional distributions are not substantially different from one another: They all appear to be sampled from the same distribution that we see in average Guess 1 reports. A crude way to assess whether Guesses 1 and 2 are dependent is to compare the average serial position reported for Guess 2 (within the range of $-1$ to 1) on trials where Guess 1 came from Serial Position $-2$ with trials where Guess 1 came from
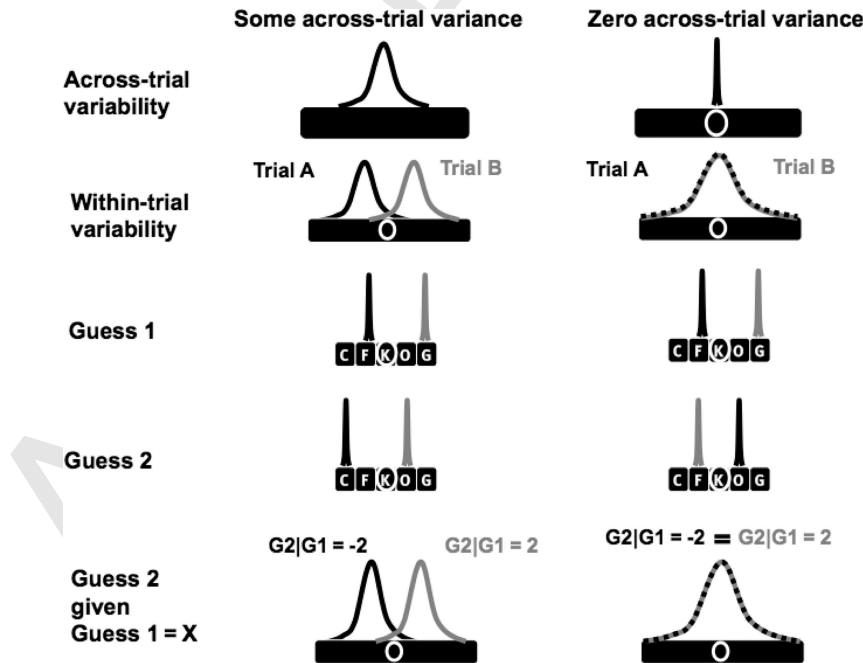


*Figure 8.* The logic behind the conditional analysis in Experiment 1. If there is a substantial amount of between-trial variance (Column 1), then, on some trials, earlier rapid serial visual positions will be selected (e.g., Trial A, black), and on other trials, later positions will be selected (e.g., Trial B, gray). Thus, on this account, Guesses 1 and 2 will be dependent, in that a guess from an early serial position of Guess 1 would predict more early stream reports for Guess 2. If there is zero across-trial variability (Column 2), then the selected distribution will be identical on every trial, and Guesses 1 and 2 will be independent. G2 = Guess 2; G1 = Guess 1.
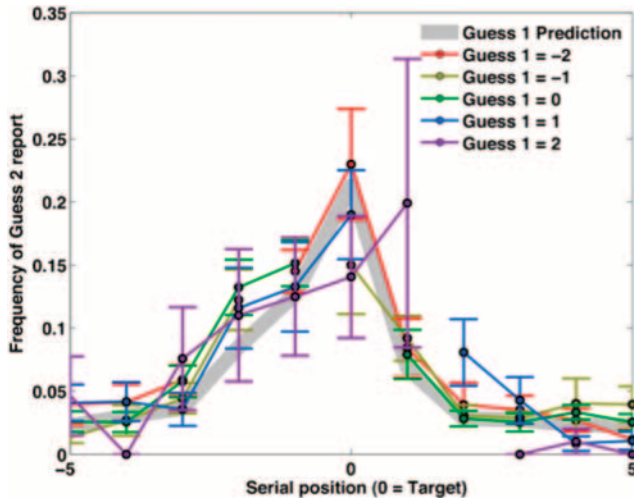
*Figure 9.* Experiment 1 results: Guess 2 reports conditioned on Guess 1. The distribution of Guess 2 reports does not change as a function of Guess 1 reports. This indicates that there is very little (if any) across-trial variance—Guess 1 and Guess 2 are independent. Moreover, the distribution of Guess 2 reports follows the distribution predicted by Guess 1 reports (thick gray line), indicating that Guess 1 and Guess 2 are identically distributed.

Serial Position 2. This comparison shows no significant difference, $t(8) < 1$, and the 95% confidence interval of the difference straddles 0 ($-0.74$ to 0.36).

Another test of independence is to evaluate the correlation between Guess 1 serial position and Guess 2 serial position. To make this test more conservative, we considered only trials on which Guess 1 and Guess 2 came from Serial Positions $-3$ through 3; thus, we discard most "noise" trials. Moreover, we discarded trials in which subjects reported the same absolute-value serial position for Guess 1 and Guess 2 (e.g., $-1$ and 1); thus, we removed the bias that would otherwise exist in this analysis because subjects cannot report the same letter twice. This left us with an average of 82 trials per subject. This analysis reveals no correlation between Response 1 and Response 2: an average correlation of $-.06$, with 95% confidence intervals across subjects between $-0.15$ and 0.02 (thus, if anything, there is a negative correlation). Thus, this analysis also shows that Guess 1 and Guess 2 are independent, with respect to their average serial position, as predicted if there were no (or very little) across-trial variability in the temporal position of the selection window.

Our claim that the conditional Guess 2 distributions are unchanged regardless of the serial position that Guess 1 came from can be more conservatively tested by asking whether the frequency of reports of any serial position differs between any of the five Guess 1 conditions. To test this, we computed 30 pairwise comparisons. For instance, one such comparison was the following: Probability of reporting Serial Position 2 on Guess 2 after Guess 1 was Serial Position 1, compared with the probability of reporting Serial Position 2 for Guess 2 when Guess 1 was Serial Position 0. We did such comparisons for every combination of the five Guess 1 report conditions, for Guess 2 reports in every serial position between $-2$ and 2 (where reports were above chance—note that this is more conservative than comparing all of the serial positions,

many of which are at chance for all conditions). Of those 30 comparisons, only 2 had $p < .05$, as would be expected by chance. Even if one adopts a lenient correction for multiple comparisons (Dunn-Sidak multiple comparison correction), none of the 30 comparisons are significant. Thus, we conclude that the distribution of letters reported in the second guess is independent of the serial position of the first guess. This would not be the case if there was any substantial across-trial variance resulting in different distributions from which reports are sampled trial to trial. Thus, we conclude that Guess 1 and Guess 2 errors are independent (Vul & Pashler, 2008).

Finally, we compared these conditional distributions of reports with the distribution we would expect if Guess 2 were another independent sample from the same distribution from which Guess 1 was drawn. We performed this simulation correcting for the increased rate of random guessing on Guess 2[2] as well as for the fact that the same letter could not be reported for Guesses 1 and 2. Along with the conditional distributions of report, Figure 9 also shows this Guess 1 model prediction (thick gray line). Deviations from the Guess 1 predictions are well within the errors of our measurement ($R^2 = 0.70$, $p < .00001$). This further bolsters our claim that all guesses are samples from the same underlying distribution that results from selection and that there is very little, if any, variability in selection across-trials.

We can also evaluate the extent to which Guess 1 and Guess 2 follow the same distribution by assessing quantile-quantile (QQ) plots of the observed conditional distribution and the distribution predicted by the model describing Guess 2 as another independent sample from the same distribution as Guess 1. If there is a shift in the distribution of Guess 2 reports toward the serial position of Guess 1 report, we should see an offset in the QQ plots around the target (0; shown in Figure 10, left panel). In contrast, the only deviations from a diagonal line we see occurs in the tails, where random uniform guessing causes nonsystematic deviations (see Figure 10, right panel).

These results further support our finding that Guesses 1 and 2 are independent and identically distributed, indicating that responses are samples from the same underlying representation.

## Discussion

To sum up these results, in Experiment 1 we found that Guess 1 and Guess 2 on a given trial tend to come from adjacent serial positions, indicating that selective attention in time selects multiple letters on a given trial (thus, ruling out the single-item selection hypothesis). Second, we found that Guess 1 and Guess 2 are independent, indicating that there is no shared across-trial variance between the two guesses, which rules out the contiguous all-or-none selection hypothesis and the contiguous-graded selection hypothesis with any substantial amount of across-trial variability. Finally, we also found that the conditional Guess 2 report distributions follow the predictions of a model of Guess 2 reports as another sample from the distribution of Guess 1 reports; thus, it

---

[2] We can adjust for random guessing by altering the proportions of a mixture model in which some proportion of guesses arises from a distribution as seen on Guess 1, and another proportion of guesses arises from a uniform distribution over all items.
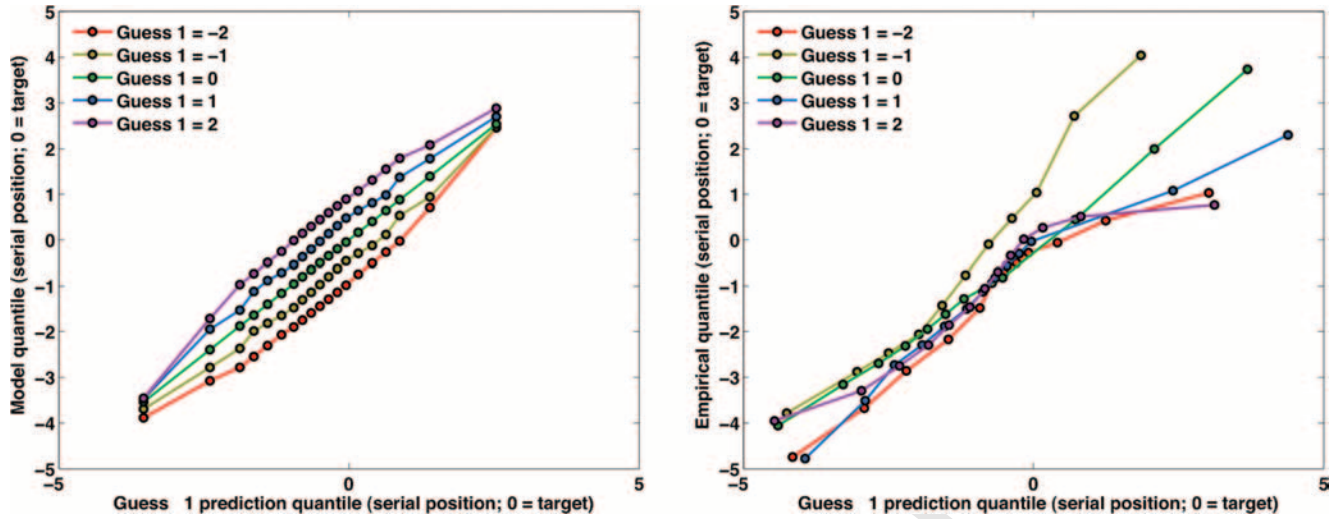
*Figure 10.* Quantile-quantile (QQ) plots that can be used to analyze the similarity between two probability distributions. The left panel represents what the QQ plots would look like if Guess 2 shifted in the direction of Guess 1 errors. The right panel represents the observed QQ plots; deviations from the diagonal are seen only in the extreme values.

seems that Guess 1 and Guess 2 are identically distributed. All together, these results support the hypothesis that on any given trial, attention "selects" a range of letters in a graded fashion, and the position of this selection window does not vary trial to trial. Responses have the statistical properties of independent and identically distributed samples from the graded selection distribution. A parsimonious account of these results describes selection as representing the uncertainty inherent in the inference about co-occurrence (the computational problem of the task) as a probability distribution over letters, from which responses are sampled.

## Experiment 2

We have shown that selection in time (temporal selection) can be best described as contiguous-graded selection with no detectable across-trial variability. In Experiment 2, we tested whether spatial selection also has the same properties. To do so, we used a paradigm that exchanges the roles of spatial and temporal dimensions of the RSVP experiment to create conditions that are comparable to RSVP, but in the spatial domain. Specifically, in RSVP, we display many letters in one location, separated in time: in Experiment 2, we displayed the same number of letters, at one point in time, separated in space. Thus, this design is similar to many historic iconic memory experiments (Averbach & Coriell, 1961).

### Method

*Participants.*   Eleven subjects from the Massachusetts Institute of Technology subject pool were recruited to participate. Subjects were between 18 and 35 years of age and were paid $10 for participation.

*Materials.*   On each trial, subjects saw the 26 English letters presented simultaneously in a circle in a random arrangement. Each letter subtended approximately 2° visual angle, and the circle

perimeter was at 6°. A line extending from fixation to the cued location served as the target cue. The cued location could be one of 13 points along the circle of letters (20°–353° degrees in the monitor plane, separated in steps of 27°). All display items were white on a black background, and letters were in capitalized Courier font (see Figure 11). **F11**

Each trial began with 1.5 s of fixation, then the cue was presented for 50 ms, followed by the letter array for 100 ms, followed again by the cue for 100 mc (see Figure 9).

The experiment was programmed in PsychToolbox (Brainard, 1997) on MATLAB 7 on a Windows XP computer.

*Procedure.*   Each subject began the experiment with two practice trials; the results of these trials were discarded. Following the practice trials, subjects completed five blocks of 78 trials each. Each block contained six instances of each of the 13 possible cue locations, in a random order for each block.

At the end of each trial, subjects were asked to make four guesses about which letter they thought was cued. Subjects reported the letters by pressing the corresponding keys on the keyboard. Just as in Experiment 1, duplicate letter reports were not accepted, and subjects were awarded 1, 0.5, 0.25, or 0.125 points if they guessed the cued letter correctly on Guesses 1–4, respectively. Again, as in Experiment 1, feedback and scoring reflected this instruction (in this experiment, the average bonus was $1.60).

### Results

Just as in Experiment 1, each letter appeared only in one (spatial) position on any given trial; thus, we could identify the exact location where any given reported letter appeared. We could then compute the empirical histogram of reports around the cue across trials for any given guess. Figure 12 (left panel) shows the **F12** empirical frequencies of reports for each guess, and Figure 12 (right panel) shows the logarithm of the ratio of observed: chance
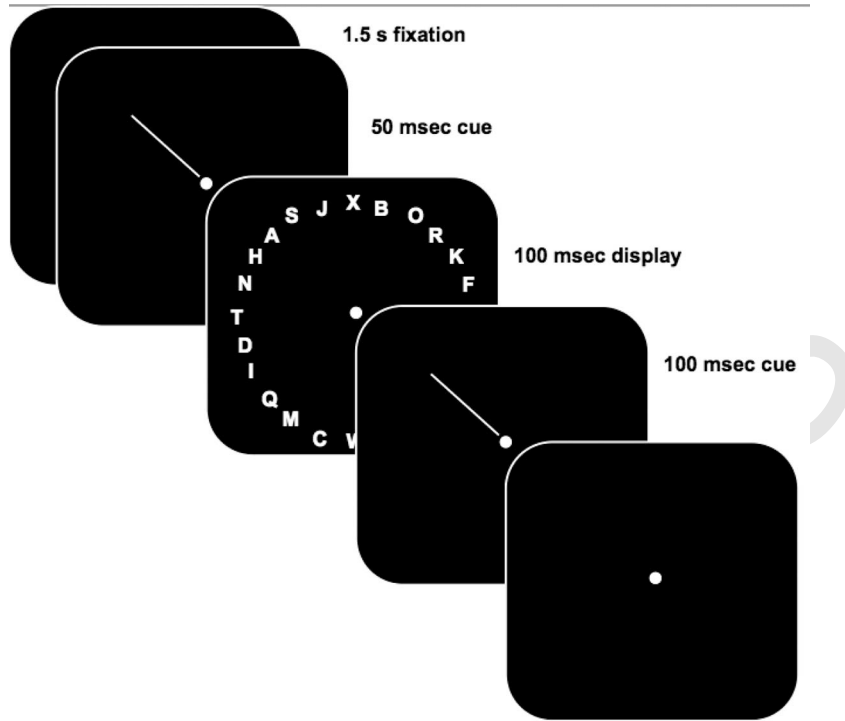
*Figure 11.* Experiment 2 design: A spatial version of rapid serial visual. All letters are presented at the same point in time, spread across space, and subjects must report the letter cued by the line and are given four guesses to do so.

frequencies. Just as in Experiment 1, the histogram of reports across trials shows substantial variability, and again, above chance reports on the second guess (above 0 log-observed:chance ratios; see Figure 12b).

To determine whether these results could arise from single-item selection or whether multiple letters were selected on a given trial, we again analyzed the distribution of Guess 2 reports around Guess 1. As can be seen in Figure 13, the letters reported for Guess **F13**
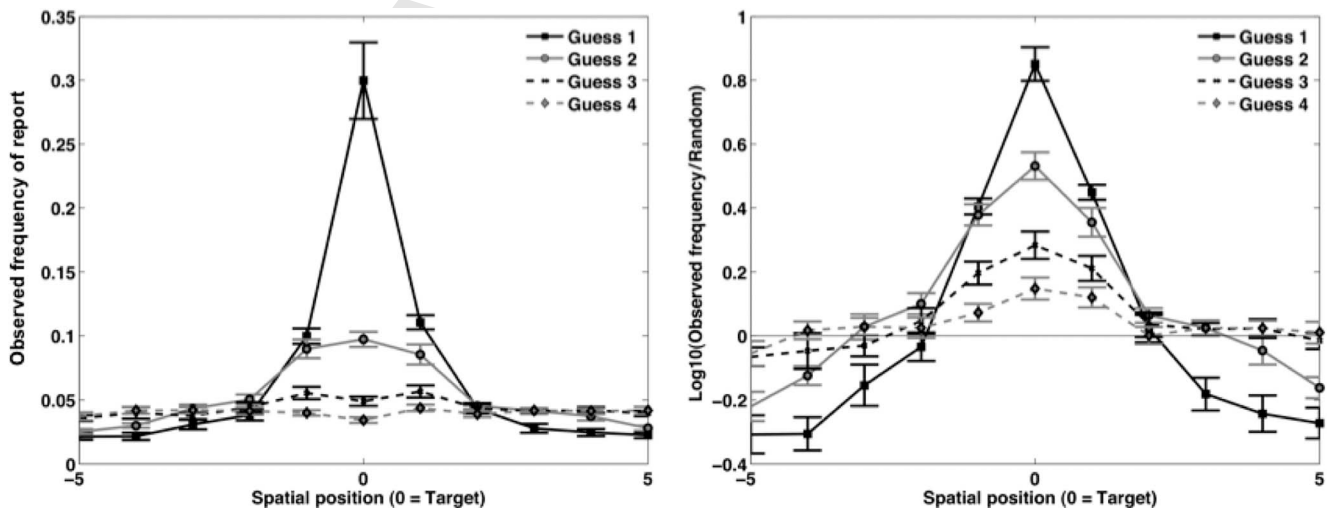


*Figure 12.* Experiment 2 results: The spatial distribution of reports for each of the four guesses subjects made on each trial. The *x*-axes correspond to spatial position, where 0 is the target, negative positions are counter-clockwise from the target, and positive positions are clockwise from the target. Panel 1 shows raw frequency data; Panel 2 shows the logarithm of the ratio of observed-to-chance report frequencies. Error bars are 1 *SEM*.
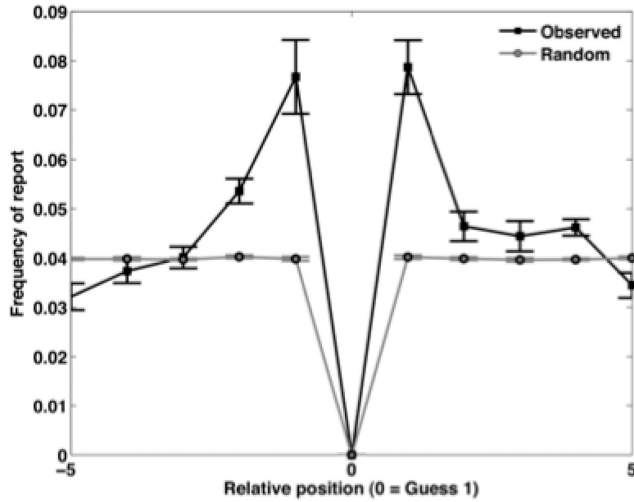
*Figure 13.* Experiment 2 results: Guess 2 reports around the position of the item guessed on Guess 1. Subjects are likely to report two spatially adjacent letters on Guesses 1 and 2, indicating that multiple letters are selected on any given trial.

2 tend to be adjacent to the letter reported on Guess 1 (for the four positions immediately adjacent to Guess 1, Guess 2 report frequency is above chance, all $ts(8) > 4$, $ps < .01$). This indicates that in space, just as in time, selective attention selects several letters on a given trial.

We used the same logic as in Figure 8 of Experiment 1 to test whether the selected letters are selected in an all-or-none or in a graded fashion. If they were selected in an all-or-none fashion, then across-trial variability (translation of the selection window) is required to produce the observed graded across-trial histograms. Thus, again, we looked at the distributions of Guess 2 reports conditioned on different Guess 1 reports. Figure 14 shows the results of this analysis. Just as in the temporal case, in the spatial case the distribution of Guess 2 reports does not depend on which
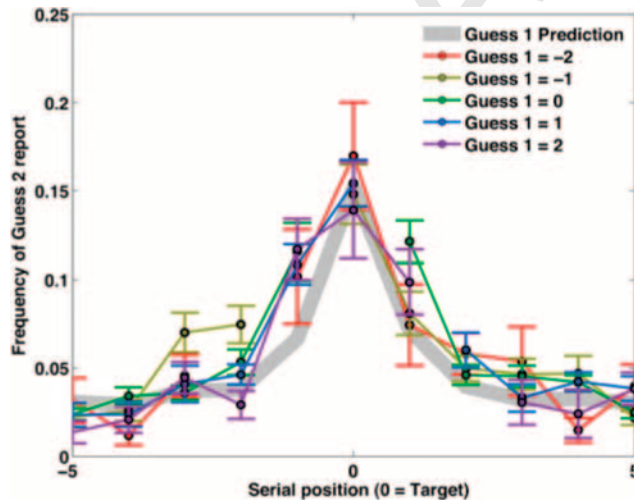
item was reported on Guess 1. We again compared the average reported position in the range of −1–1 when Guess 1 came from Serial Position −2 and when it came from Serial Position 2. We found no significant difference ($p = .86$, 95% confidence intervals [CIs] on the mean shift are −0.12–0.10). As in Experiment 1, we could also assess the independence of Guess 1 and Guess 2 by analyzing the correlation between Guess 1 and Guess 2 reports (using the same corrections as described in Experiment 1). Again, in the spatial selection case, just as in the temporal selection case, we found no significant correlation (95% CIs on the correlation coefficient are between −0.02 and 0.07, with an average of 102 trials included per subject). We can again assess whether the conditional distributions are identical by testing whether there are any significant differences in the frequency of any reported spatial positions within the range of −2 to 2 for each of the conditional distributions. To this end, we ran 30 pairwise comparisons, as in Experiment 1; although, 4 were significant, none survived a Dunn-Sidak multiple-comparisons correction. Just as in Experiment 1, the three analyses above indicate that Guess 1 and Guess 2 are independent, in that there is no evidence for any shared across-trial variance.

As in Experiment 1, we evaluated whether Guess 1 and Guess 2 are identically distributed by assessing whether conditional Guess 2 reports follow the same distribution as would be predicted by a model that describes Guess 2 as another sample from the Guess 1 distribution (modulo increased random guessing and the fact that the same letter cannot be reported twice). The correlation between the model prediction (shown in Figure 14) and the observed conditional report frequencies is very high ($r^2 = .88$, $p < .00001$). Finally, we can again assess the QQ plots for the predicted distribution and the observed distributions (for a prediction of what a nonindependent QQ plot would look like; see Figure 10a). Figure 15 shows these QQ plots: Again, the only observable deviation from a diagonal occurs in the noisy tails, but not around the target (the point with highest probability), indicating that the predicted and observed probabilities match very well.

Again, in the spatial case, just as in the temporal case, we see that selective attention selects a number of letters to varying



*Figure 14.* Experiment 2 results: Conditional Guess 2 reports as a function of spatial position.
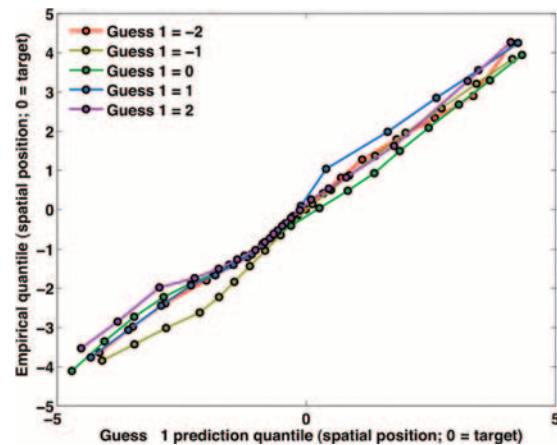


*Figure 15.* Experiment 2 results: Quantile-quantile plots for predicted conditional Guess 2 distributions and observed conditional Guess 2 distributions.

degrees on any given trial, reflecting the uncertainty inherent in the task. This conclusion in the spatial case is reminiscent of the crowding phenomenon (He et al., 1996; Pelli et al., 2004): People are worse at identifying a cued letter when other "crowding" letters are nearby. Our data show that in such circumstances, attention selects multiple adjacent letters, and the actual reported letter is a sample from this selected distribution. Our findings are consistent with accounts of crowding as a limit in the spatial precision of selective attention (He et al., 1996). However, for our purposes, in spatial selection, just as in temporal selection, multiple responses on a single trial have the statistical properties of independent, identically distributed samples from an internal probability distribution that reflects the uncertainty inherent in the task.

## Discussion

In two experiments, we tested the mechanisms of visual selective attention. Specifically, we asked whether multiple items are selected to different degrees on each trial, as predicted by Bayesian models of cognition in which mental representations consist of multiple simultaneous hypotheses, each with a different graded probability of being true. The fact that many studies have reported graded distributions of responses in the average over many trials does not answer this question because such distributions could arise either from selection of multiple items on each trial or from selection of a single discrete item on each trial, with some variability in the locus of selection across trials. To unconfound these two possibilities, subjects made multiple responses on each trial. In Experiment 1, we found that the temporal positions of intrusions from two guesses on one trial are uncorrelated. Because there was no correlation between errors on one trial, there is no shared spatial, or temporal, error between these two guesses. This observation means that there is no across-trial variance (or noise) in which items are selected, and therefore most of the variance seen in the final distribution of reports must occur within a given trial. Evidence of substantial within-trial variability indicates that subjects select a contiguous range of letters to varying degrees on every trial; thus, selective attention produces a representation equivalent to an internal probability distribution over likely targets (see Figure 3, right panel). In Experiment 2, we extended these results to the domain of spatial selection. There too, our data indicate that selection creates a graded probability distribution over a range of possible targets, and subjects make responses by sampling guesses from this distribution. Thus, it seems that errors in visual selective attention tasks arise due to a process of sampling responses from internal representations that reflect the uncertainty inherent in the task.

Our results connect to three other lines of research. First, Sergent and Dehaene (2004) assessed whether conscious access is discrete or continuous using an attentional blink (Raymond et al., 1992) paradigm: When two targets in an RSVP stream appear in close temporal proximity, the second target is often missed due to failures of attentional selection (Vul et al., 2008). Sergent and Dehaene (2004) asked subjects to report the visibility of the second target with a continuous scale and found that subjects used the scale in an all-or-none fashion. They reported either seeing or not seeing the target, without using any settings in between, suggesting that the target letter was not selected in a graded fashion. Our results suggest that subjects are not aware of the degree to which

a given item was selected (and thus, they cannot choose the most likely alternative), but instead they must sample alternatives for report. Thus, it appears that although selective attention operates continuously, we are only aware of discrete samples from our internal probabilistic states, indicating that conscious access is discrete, as Sergent and Dehaene claimed.

Second, the difference between continuous and graded selection and discrete conscious access bears on Boolean map theory (Huang & Pashler, 2007). Huang and Pashler described a series of elegant experiments that suggest that subjects can select regions of space only via Boolean maps—a region of space may be selected, or not selected, with no states in between. However, evidence for this claim comes from experiments that measure conscious "access" to the products of selection (e.g., mental rotation or transformation). There is no disagreement between our findings and those of Huang and Pashler: In our view, selection does not operate discretely, but rather continuously, selecting regions of space to varying degrees. However, access is discrete and reflects a sample from the selected distribution. Thus, continuous selection and discrete access are not in opposition if access is limited to a discrete sample from the selected distribution.

Our conclusions are also consistent with a third line of research: Shih and Sperling's (2002) proposed account of visual selective attention as a spatiotemporal "gate." This account can be seen as an algorithmic-level analysis in which we have offered an account at the level of computational theory (Marr, 1982). Our analysis of the computational problem entailed in selective attention tasks under uncertainty (detecting co-occurrences between cues and targets distributed over space or time) yields the same operations that Shih and Sperling's algorithmic level account proposed. The attentional gate proposed in their algorithm fulfills the computational role of uncertainty in the position of the cue. What Shih and Sperling referred to as spatiotemporal interference between items (interpretable as the persistence and point-spread functions of iconic memory) is computationally equivalent to what we refer to as the uncertainty about the spatiotemporal position of each letter. The process of multiplying the attentional gate function with the activation function of each letter and integrating over time is the same computation one would undertake to perform the appropriate inference about co-occurrence. In Shih and Sperling's algorithm, the result of this multiplication and integration produces activation strengths in short-term memory—these are computationally equivalent to a scaled probability distribution. Finally, the operation Shih and Sperling proposed of adding noise to this short-term memory strength and responding by taking the maximally activated letter may be equivalent to random sampling from a probability distribution (given certain conditions on the exact distribution of noise, e.g., variance scaling proportional to the activation strength; Ma, Beck, Latham, & Pouget, 2006). In short, the theoretical analysis of selective attention tasks that motivated our experiments is computationally isomorphic with the linear-systems account proposed by Shih and Sperling.

Several alternative accounts of our data cannot be ruled out by the present experiments. First, it could be the case that on each trial, multiple selection episodes are operating independently, each selecting one letter from a region around the target. We cannot rule out this account, as it could predict any pattern of data. However, on this account, an individual selection episode acts as the sampling process that we ascribe to postselective process of retrieval

from short-term memory; thus, instead of a probabilistic representation of the selected letters, as we advocate, this account must pose a probabilistic tendency to deploy selection episodes. Another alternative account is that there is complete certainty in the location of the cue, but there is substantial noise, or uncertainty, in the location of individual items, which are then coded with respect to their distance from the cue and reported accordingly. Both of these accounts are plausible alternatives that should be addressed in future work. Tentatively, we can say that other data from our lab (in which people are asked to report multiple features of one item) rules out the simplest version of this account (Rich & Vul, 2008). In general, completely ruling out "noise" in favor of "intrinsic uncertainty" as the source of variability in responses is impossible, as noise can be postulated to arise at any point in an arbitrarily complicated process model, thus making it consistent with just about any pattern of data. In our case, we think we have ruled out some intuitively simple accounts of noise in attentional selection, thus supporting the idea that in such tasks, intrinsic uncertainty coupled with a postselection sampling process are responsible for variability in subjects' responses.

There is an interesting tension in our data: We conclude that the gradation in the tendency to report a particular item reflects gradation in the degree to which each item is selected, rather than the average, across trials, of a set of all-or-none selection episodes of different items. We argue that this gradation reflects the result of uncertain inference about which item co-occurred with the cue. Thus, we postulate that the system represents uncertainty about where and when each item, and each cue, occurred. Usually, this uncertainty is purported to arise from noise that perturbs these measurements. However, we show no evidence of across-trial noise perturbing the spatiotemporal position of the cue (which would arise in translation of selection across trials). So why would there be uncertainty? This tension may be reconciled by supposing that the human visual system, through years of experience, has learned the amount of noise perturbing its measurements of the world, and the learned uncertainty persists in controlled laboratory settings when actual noise is eliminated through precise digital presentation. If so, then we predict that the uncertainty in selection (as measured by spatiotemporal variability of reported items) would decrease with sufficient training in laboratory or video game settings—this is a promising direction for future research.

In summary, our results provide evidence of a sampling process that connects graded internal probability distributions with discrete conscious access and responses. These results dovetail with findings from a very different domain: When subjects are asked to guess arbitrary facts about the world (e.g., "What proportion of the world's airports are in the US?"), multiple guesses from one subject contain independent error—thus, averaging two responses from one subject produces a more accurate guess than either response alone—a "crowd within" (Vul & Pashler, 2008). The "crowd within" results and the results in this article are both predicted by the idea that the mind operates via probabilistic inference (Chater et al., 2006) and solves complicated probabilistic inference problems by sampling (Vul, Goodman, Griffiths, & Tenenbaum, in press). Internal representations are graded probability distributions, yet responses about, and conscious access to, these representations is limited to discrete samples. Our mind appears to perform Bayesian inference without our knowing it.

## References

Averbach, E., & Coriell, A. (1961). Short term memory in vision. *Bell Systems Technical Journal, 40,* 309–328.

Baars, B. (1997). Some essential differences between consciousness and attention, perception, and working memory. *Consciousness and Cognition, 6*(2–3), 363–371.

Botella, J., & Eriksen, C. W. (1992). Filtering versus parallel processing in RSVP tasks. *Perception and Psychophysics, 51,* 334–343.

Botella, J., Garcia, M. L., & Barriopedro, M. (1992). Intrusion patterns in rapid serial visual presentation tasks with two response dimensions. *Perception & Psychophysics, 52,* 547–552.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision, 10,* 433–436.

Bruner, J., Goodnow, J., & Austin, G. (1956). *A study of thinking.* New York: Wiley.

Carrasco, M. (2006). Covert attention increases contrast sensitivity: Psychophysical, neurophysiological and neuroimaging studies. *Progress in Brain Research, 154,* 33–70.

Chater, N., & Oaksford, M. (2008). *The probabilistic mind: Prospects for Bayesian cognitive science.* New York: Oxford University Press.

Chater, N., Tenenbaum, J. B., & Yuille, A. (2006). Probabilistic models of cognition: Conceptual foundations. *Trends in Cognitive Sciences, 10,* 287–291.

Chun, M. M. (1994). Processing deficits in RSVP: The attentional blink and repetition blindness. Retrieved September, 4, 2009, from http://dspace.mit.edu/handle/1721.1/11651

Chun, M. M. (1997). Temporal binding errors are redistributed by the attentional blink. *Perception & Psychophysics, 59,* 1191–1199.

Chun, M. M., & Potter, M. C. (1995). A two-stage model for multiple target detection in rapid serial visual presentation. *Journal of Experimental Psychology, 21,* 109–127.

Estes, W. K. (1956). The problem of inference from curves based on group data. *Psychological Bulletin, 53,* 134–140.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics.* New York: Wiley.

Griffiths, T. L., & Tenenbaum, J. B. (2006). Optimal predictions in everyday cognition. *Psychological Science, 17,* 767–773.

He, S., Cavanagh, P., & Intriligator, J. (1996, September 26). Attentional resolution and the locus of visual awareness. *Nature, 383,* 334–337.

Huang, L., & Pashler, H. (2007). A Boolean map theory of visual attention. *Psychological Review, 114,* 599–631.

Huang, L., Treisman, A., & Pashler, H. (2007, August 10). Characterizing the limits of human visual awareness. *Science, 317,* 823–825.

Itti, L., & Baldi, P. (2009). Bayesian surprise attracts human attention. *Vision Research, 49,* 1295–1306.

Kahneman, D., Treisman, A., & Gibbs, B. J. (1992). The reviewing of object files: Object-specific integration of information. *Cognitive Psychology, 24,* 175–219.

Kemp, C., Bonawitz, E. B., Coley, J. D., & Tenenbaum, J. B. (2008). Inductive reasoning about causally transmitted properties. *Cognition, 109,* 175–192.

Kersten, D., & Yuille, A. (2003). Bayesian models of object perception. *Current Opinion in Neurobiology, 13,* 1–9.

Kikuchi, T. (1996). Detection of Kanji words in a rapid serial visual presentation task. *Journal of Experimental Psychology: Human Perception and Performance, 22,* 332–341.

Kinchla, R., & Smyzer, F. (1967). A diffusion model of perceptual memory. *Perception & Psychophysics, 2,* 219–229.

Koch, C., & Tsuchiya, N. (2007). Attention and consciousness: Two distinct brain processes. *Trends in Cognitive Sciences, 11,* 16–22.

Lamme, V. (2003). Why visual attention and awareness are different. *Trends in Cognitive Sciences, 7,* 12–18.

Landman, R., Spekreijse, H., & Lamme, V. A. (2003). Large capacity

storage of integrated objects before change blindness. *Vision Research, 43,* 149–164.

Ma, W. J., Beck, J. M., Latham, P. E., & Pouget, A. (2006). Bayesian inference with probabilistic population codes. *Nature in Neurosciences, 9,* 1432–1438.

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information.* New York: Henry Holt.

Mozer, M., Pashler, H., & Homaei, H. (2008). Optimal predictions in everyday cognition: The wisdom of individuals or crowds? *Cognitive Science, 32,* 1133–1147.

Mozer, M. C., Shettel, M., & Vecera, S. (2005). Top-down control of visual attention: A rational account. In Y. Weiss, B. Schoelkopf, & J. Platt (Eds.), *Advances in neural information processing systems* (pp. 923–930). Cambridge, MA: MIT Press.

Najemnik, J., & Geisler, W. S. (2005, March 17). Optimal eye movement strategies in visual search. *Nature, 434,* 387–391.

Nieuwenstein, M., Chun, M. M., & van der Lubbe, R. H. J. (2005). Delayed attentional engagement in the attentional blink. *Journal of Experimental Psychology: Human Perception and Performance, 31,* 1463–1475.

Nieuwenstein, M. R., & Potter, M. C. (2006). Temporal limits of selection and memory encoding: A comparison of whole versus partial report in rapid serial visual presentation. *Psychological Science, 17,* 471–475.

Pelli, D. G., Palomares, M., & Majaj, N. J. (2004). Crowding is unlike ordinary masking: Distinguishing feature integration from detection. *Journal of Vision, 4,* 1136–1169.

Posner, M. I. (1994). Attention: The mechanisms of consciousness. *Proceedings of the National Academy of Sciences, USA, 91,* 7398–7403.

Posner, M. I., Snyder, C. R., & Davidson, B. J. (1980). Attention and the detection of signals. *Journal of Experimental Psychology, 109,* 160–174.

Prinzmetal, W., Henderson, D., & Ivry, R. (1995). Loosening the constraints on illusory conjunctions: Assessing the roles of exposure duration and attention. *Journal of Experimental Psychology: Human Perception and Performance, 21,* 1362–1375.

Prinzmetal, W., Ivry, R. B., Beck, D., & Shimizu, N. (2002). A measurement theory of illusory conjunctions. *Journal of Experimental Psychology: Human Perception and Performance, 28,* 251–269.

Raymond, J. E., Shapiro, K. L., & Arnell, K. M. (1992). Temporary suppression of visual processing in an RSVP task: An attentional blink? *Journal of Experimental Psychology: Human Perception and Performance, 18,* 849–860.

Reeves, A., & Sperling, G. (1986). Attention gating in short-term visual memory. *Psychological Review, 93,* 180–206.

Rich, A., & Vul, E. (2008, August). *Can feature binding be explained by certainty in space and time?* Poster presented at the European Conference on Visual Perception, Utrecht, the Netherlands.

Sergent, C., & Dehaene, S. (2004). Is consciousness a gradual phenomenon? Evidence for an all-or-none bifurcation during the attentional blink. *Psychological Science, 15,* 720–728.

Shih, S., & Sperling, G. (2002). Measuring and modeling the trajectory of visual spatial attention. *Psychological Review, 109,* 260–305.

Steyvers, M., Griffiths, T., & Dennis, S. (2006). Probabilistic inference in human semantic memory. *Trends in Cognitive Sciences, 10,* 327–334.

Strasburger, H. (2005). Unfocused spatial attention underlies the crowding effect in indirect form vision. *Journal of Vision, 5,* 1024–1037.

Swets, J., Tanner, Jr., W. P., & Birdsall, T. G. (1961). Decision processes in perception. *Psychological Review, 68,* 301–340.

Tenenbaum, J. B. (1999). Bayesian modeling of human concept learning. In M. S. Kearns, S. A. Solla, & D. A. Cohn (Eds.), *Advances in neural information processing systems 11* (pp. 59–68). Cambridge, MA: MIT Press.

Treisman, A., & Schmidt, H. (1982). Illusory conjunctions in the perception of objects. *Cognitive Psychology, 14,* 107–141.

VanRullen, R., & Koch, C. (2003). Is perception discrete or continuous. *Trends in Cognitive Sciences, 7,* 207–213.

Vul, E., Goodman, N. D., Griffiths, T. L., & Tenenbaum, J. B. (in press). *One and done? Optimal decisions from very few samples.* Proceedings of the 31st Annual Meeting of the Cognitive Science Society, Amseterdam, the Netherlands.

Vul, E., Hanus, D., & Kanwisher, N. (2008). Delay of selective attention during the attentional blink. *Vision Research, 48,* 1902–1909.

Vul, E., Nieuwenstein, M., & Kanwisher, N. (2008). Temporal selection is suppressed, delayed, and diffused during the attentional blink. *Psychological Science, 19,* 55–61.

Vul, E., & Pashler, H. (2008). Measuring the crowd within: Probabilistic representations within individuals. *Psychological Science, 19,* 645–647.

Weichselgartner, E., & Sperling, G. (1987, November 6). Dynamics of automatic and controlled visual attention. *Science, 238,* 778–780.

Whiteley, L., & Sahani, M. (2008). Implicit knowledge of visual uncertainty guides decisions with asymmetric outcomes. *Journal of Vision, 8,* 1–15.

Yu, A. J., & Dayan, P. (2005). Inference, attention, and decision in a Bayesian neural architecture. *Advances in Neural Information Processing Systems, 17,* 1577–1584.