PREDICTABLE EFFECTS OF VISUAL SALIENCE IN EXPERIMENTAL DECISIONS AND GAMES*

XIAOMIN LI AND COLIN F. CAMERER

Bottom-up stimulus-driven visual salience is largely automatic, effortless, and independent of a person's "top-down" perceptual goals; it depends only on features of a visual stimulus. Algorithms have been carefully trained to predict stimulus-driven salience values for each pixel in any image. The economic question we address is whether these salience values help explain economic decisions. Our first experimental analysis shows that when people pick between sets of fruits that have artificially induced value, predicted salience (which is uncorrelated with value by design) leads to mistakes. Our second analysis uses evidence from games in which choices are locations in images. When players are trying to cooperatively match locations, predicted salience is highly correlated with the success of matching ($r =$ .57). In competitive hider-seeker location games, players choose salient locations more often than predicted by the unique Nash equilibrium. This tendency creates a disequilibrium "seeker's advantage" (seekers win more often than predicted in equilibrium). The result can be explained by level-k models in which predicted stimulus-driven salience influences level-0 choices and thereby influences overall perceptions, beliefs, and choices of higher-level players. The third analysis shows that there is an effect of visual salience in matrix games, but it is small and statistically weak. Applications to behavioral IO, price and tax salience, nudges and design, and visually influenced beliefs are suggested. *JEL Codes:* D91, C91, C72.

* Support was provided by the Behavioral and Neuroeconomics Discovery Fund (PI Camerer), Tianqiao and Chrissy Chen Center for Social and Decision Neuroscience, Alfred P. Sloan Foundation (G201811259), and NIMH Conte Center P50MH094258. Thanks to audiences at the Caltech Graduate Proseminar, Sloan/NOMIS Conference on Decision and Cognition, IAREP/SABE (Middlesex), Columbia University, Peking University, UC-Berkeley Behavioral Economics, Virtual Process Tracing Conference, Bocconi IGIER, and the Pitt Behavioral and Experimental Economics seminar. We especially thank Elke Weber, Vince Crawford, Jacqueline Gottlieb, Richard Thaler (for the Schelling list idea), Adam Sanjurjo (for "The Pearl" tip), the editor and several anonymous referees for helpful comments, Luca Polonio and Giorgio Coricelli for sharing the valuable gaze data in games, Anne Karing for a valuable image, Gidi Nave, Xintong Han, and Nina Solovnyeva (SURF 2020). Extra special thanks go to Eskil Forsell, Milica Mormann, and Alec Smith whose energetic prior research on this topic laid the foundation for this article, although their own work was never published.

The Quarterly Journal of Economics (2022), 1849–1900. https://doi.org/10.1093/qje/qjac025.
Advance Access publication on June 11, 2022.

I. Introduction

Features of a stimulus that grab attention are called "salient." Of the different types of externally triggered sensory salience, visual salience is the best understood and is clearly important given the amount of information that people process through the visual system. This investigation is about whether one type of visual salience can be predicted and can help explain choices in experimental economic decisions and games.

Many economists have studied attention and salience recently, as part of growth in the foundations of behavioral economics. Notable contributions include salience theory (Bordalo, Gennaioli, and Shleifer 2012b, 2013a, 2013b), a related model of focusing (Kőszegi and Szeidl 2013), and theories of rational (Sims 2003, 2006; Caplin and Dean 2015; Caplin, Dean, and Leahy 2019; Caplin et al. 2020; Kőszegi and Matějka 2020; Mackowiak et al. 2020) and dynamic inattention (Schwartzstein 2014; Gagnon-Bartsch, Rabin, and Schwartzstein 2018). The SAM algorithm salience is different from these economic models in content and purpose. We defer the comparison of those models to Section VII.

To begin with, there is an important distinction between "bottom-up" and "top-down" salience (e.g., Baluch and Itti 2011; Chun, Golomb, and Turk-Browne 2011).[1]

Bottom-up salience is what the human visual system notices most quickly and automatically. Bottom-up salience is also called "stimulus-driven"—the term we use from now on—because it depends only on the properties of a stimulus. Stimulus-driven

1. There is an ongoing debate in attention science about how sharp the bottom-up versus top-down distinction is. Awh, Belopolsky, and Theeuwes (2012) give the example of the history of selective attention to a feature, which seems to influence future attention. That influence is not purely stimulus-driven (because it depends on previous attentive behavior, not just the stimulus itself), nor is it accomplishing a goal. Another example is faces. Faces are considered to be bottom-up salient for humans, but they help achieve a variety of goals that are generally evolutionarily important (such as emotional communication, friend-foe detection, mate choice, and social learning). These goals might be even more important in a particular domain, like decoding facial emotion while watching a dramatic movie. So a person watching a movie sees faces that have both automatic bottom-up salience, and additional top-down salience to achieve the goal of understanding the movie. In general, the two processes together can be thought of as a family of filters that have been adaptively shaped by forces ranging almost continuously from evolutionarily conserved universal principles to others locally tuned by personal experience and valuation.

properties can be further divided into low- and high-level features (Judd et al. 2009). Low-level features are independent of object identity, meaning, and categorization; they include intensity, orientation, color, and motion. Higher-level features combine low-level features to identify and categorize objects and direct attention to objects that are familiar, semantically meaningful, and generally valued. Faces, people, and text are generally salient high-level features.

Many algorithms have been trained to predict stimulus-driven salience using large image sets and eye-tracking data from people who are "freely gazing" at the images for three to five seconds. These algorithms produce "salience maps" that closely match the actual gaze patterns.

In contrast to stimulus-driven attention, top-down attention is directed to achieve specific goals. We refer to top-down attention as "goal-directed" attention.[2] Goal-directed attention includes "extra-retinal[3] information such as intrinsic expectations, knowledge and goals" (Baluch and Itti 2011, 210).

To illustrate the distinction between stimulus-driven and goal-directed attention, consider the classic study by Yarbus (2013), done in 1967. He showed subjects a painting of people in a room. One group was told to freely gaze. Another group was told to "estimate the material circumstance" of the people in the painting. The third group was told to "estimate the ages of the people" in the painting. Eye-tracking showed that each of the three groups looked at somewhat different parts of the images.[4] Their gaze differences were due to differences in goal-directed attention. However, there was also a substantial overlap in measured attention. For example, people in both the free gaze and the "estimate the ages" goal conditions looked at faces in a similar way.

2. Stimulus-driven and goal-directed attention are also sometimes called "exogenous" and "endogenous" in attention psychology. Although we will not use this terminology, it is useful to emphasize the difference between stimulus-driven and rational (endogenous) attention models, discussed in Section VII.

3. "Extra-retinal" means that the information attended to because of goal-directed guidance is not input to the retina, but is instead represented in the visual cortex and other regions such as the superior colliculus (see Veale, Hafed, and Yoshida 2017); that information is in the proverbial "mind's eye" rather than coming from the retina.

4. Reversing the order of inference in Yarbus's early study, Haji-Abolhassani and Clark (2014) showed that perceptual goals could be inferred reliably from eye-gaze patterns.

This overlap indicates that the measured attention to faces was both stimulus-driven and goal-directed when the goal was age estimation.

The hypothesis tested in this article is whether stimulus-driven salience influences incentivized choices in three experiments involving decisions and strategic games. This type of salience lies outside of popular rational-inattention modeling, which is a specific mathematical derivation of optimal goal-directed attention (discussed further in Section VII). The results, therefore, provide evidence that goal-directed models (including rational inattention) are leaving out a type of attention—stimulus-driven salience—which is important behaviorally.

A preview of the first experiment illustrates the conflict between stimulus-driven salience and goal-directed perception. Subjects saw two sets of fruits, on the left and right halves of their computer screen. The two fruit sets were constructed to have different stimulus-driven salience, and different induced monetary value. The subjects' goal was to choose the set with the highest induced value, which requires goal-directed perception. Under time pressure, stimulus-driven salience sometimes shifted choices toward the high-salience options, even if those were low-value choices (see also Milosavljevic et al. 2012; Towal, Mormann, and Koch 2013).

The other two experiments test whether stimulus-driven salience influences strategic choices that are intended to accomplish goals of (i) either coordinated matching, or hiding and seeking, in location games and (ii) maximizing payoffs in normal-form matrix games. Predicted salience helps explain choices in the first set of location games and is weakly associated with low-level thinker choices (as classified by eye-tracking) in the second set of normal-form games.

The empirical analysis uses an algorithm called the salience attentive model (SAM).[5] SAM takes any 2-D color image as an input and predicts stimulus-driven attention—what most people will look at—in the first few seconds. The SAM algorithm is general, so it can be applied to any economic or social decisions influenced by images. Potential applications include advertisements; visual design features of "nudges"; televised political debates; e-commerce websites; virtual house tours; retail tags showing prices, promotions, or taxes; point-of-purchase displays;

5. SAM is the first of several acronyms we use repeatedly. They are summarized in Online Appendix Table J1.

social media; face-to-face interviewing; and graphical displays of information.

Here is the structure of the article. Section II presents the SAM algorithm. Section III describes the choice experiment pitting stimulus-driven salience against goal-directed attention. The results from location game experiments are described in Section IV and explained with cognitive hierarchy and level-k modeling in Section V. Section VI is about matrix games. Section VII describes several recent economic models of salience and attention and contrasts them with our approach. Section VIII concludes by speculating about other economic applications.

II. The Salience Attentive Model (SAM) Algorithm

Algorithms that take images as inputs, and output predictions about where people will look, have been an active area of research in visual neuroscience since the 1990s. A brief history will help clarify what the algorithms do (see Online Appendix B for more details).

The earliest algorithms included only low-level features (Itti, Koch, and Niebur 1998). Using these features as a starting point was motivated by decades of research on the cognitive neuroscience of perception, including animal and human neuroanatomy, and detailed understanding of functions and interaction of different parts of the human visual cortex.[6] We note these facts as an indication for readers of how much is known about basic aspects of the neural circuitry underlying attention and its connection to behavior, including the ability to causally change attention and subsequent behavior.

The early low-level algorithms were steadily improved by adding features that are higher level, and generally salient, such as faces (Cerf et al. 2007). In the hunt for better predictive ac-

6. Veale, Hafed, and Yoshida (2017) is an excellent review. An elegant recent example found that stimulus-driven salient features are associated with measured neural activity in a specific area of the visual cortex called V1 (Chen et al. 2016). V1 got that label because it is activated by retinal input earlier in time than other regions and detects only the simplest low-level features, such as orientation and direction. Krasovskaya and MacInnes (2019) review other examples of how well algorithmic salience is associated with measured neural activity in the visual cortex. Other studies show that microstimulating and lesioning specific regions of the brain (in nonhuman animals) can causally change goal-directed attention and behavior (Baluch and Itti 2011).

curacy, in 2014 state-of-the-art algorithms switched to a neural network structure in which there is less a priori specification of what salient features are (Vig, Dorr, and Cox 2014). These neural networks consist of multiple “layers” of connected discrete nodes. Each node in one layer receives weighted inputs from nodes in an earlier layer, and contributes weighted output as an input to nodes in a later layer. The initial input layer is based on a stimulus, and the final output layer encodes or “sees” an approximation of the stimulus. The network is “trained” by inputting stimuli—such as images—and propagating weighted inputs and outputs to eventually create a stimulus-specific output layer. That predicted output layer is then compared to the objective stimulus, and the connecting weights linking the different layer nodes are adjusted to improve accuracy. The SAM algorithm uses several modern variants of these methods to improve accuracy and training speed.[7] The network structure is usually “pretrained” using a borrowed “backbone” network that encodes low-level features. The network is then trained further to learn encoding of semantically meaningful objects that are commonly present in the image sets and are looked at by the training subjects (such as apples, prices, people, and text; see Cornia et al. 2018). The images in the SAM training sets were highly varied, and most subjects were students or others recruited at U.S. campuses (see Online Appendix Table B1 for details).

These algorithms have progressed quickly because researchers can try out new ideas on four popular open-access salience data sets (SALICON, MIT1003, MIT300, CAT2000). These are sets of images along with “ground truth” data on what people actually looked at in the first five seconds of free gaze,

7. In technical jargon, SAM is a convolutional neural network with a salience encoder using a long short-term memory structure. Convolution is a method that combines encoding at different spatial scales. Crudely speaking, if features are encoded at fine-grained spatial scales and also at supersets of those fine-grained scales, the object is big. The “long short-term memory” (LSTM) property is a kludge to retain memory so that back-propagation algorithms that adjust hidden-layer weights based on prediction errors do not overreact and create “vanishing gradients”—which are bad. SAM uses ResNet as its “backbone” (there is also a version with a VGG backbone). The backbone is the earliest part of the network (i.e., the layers closest to stimulus input, encoding low-level features). That part of the network typically has many layers and is therefore the most computationally demanding. It is used for low-level feature extraction from the input image. People nowadays mostly use established backbones such as ResNet or VGG, much like using a standard set of code then adding further code by hand.

recorded using eye-tracking and other high-quality methods for measuring visual attention.

SAM and similar algorithms are now highly accurate. The reported performance of SAM on the website MIT-salience is 0.88 using the AUC-Judd area-under-the-curve measure (Riche et al. 2013). An AUC of 0.50 is random and 1.0 is perfectly accurate. The SAM accuracy of 0.88 is a little better than earlier algorithms and approaches the accuracy of the best human-to-human benchmark, which is 0.92.[8]

Figure I shows an example image and its associated SAM saliency maps. The salience map assigns a salience value from zero to one to each pixel of the image. The salience map is typically shown as a "heatmap" in grayscale or in color, with warmer (redder) colors indicating higher salience.[9] We adopted the default parameters from the original approach and applied them to our image data set. There are no additional free parameters.[10]

To illustrate salience and strategic choice, Figure II, Panel A shows the map drawn by Schelling (1960) in a famous discussion of focality and "psychological prominence." The map shows small square houses, a pond in the lower left, two places marked x and y, and a river running horizontally through the lower third of the map. A bridge spans the river. Schelling wrote: "Two people parachute unexpectedly into the area shown, each with a map and knowing the other has one, but neither knowing where the other has dropped nor able to communicate directly. They must get together quickly to be rescued. Can they study their maps and 'coordinate' their behavior?" (1960, 56). Schelling said seven of the eight people (87.5%) who saw the map chose to rendezvous at the bridge.

8. The best human benchmark indicates how strongly two different sets of human fixation maps correlate for the same image. Each of the two sets contains many different individuals. Human-human accuracy is less than 1.0 because of idiosyncratic individual differences in their fixations, which make predictions from one group to another less than perfect (Judd, Durand, and Torralba 2012).

9. We use the standard color protocol "jet" in Matlab for all the heatmaps in this article.

10. Note that this CNN model, or any simpler variations of it, could be retrained on new data to understand different kinds of salience. Two studies have coded abstract features of strategies in two-person matrix games (e.g., minimax, equal payoffs, level-1) and fit machine learning models using those features to explain observed choices. Hartford, Wright, and Leyton-Brown (2016) is a neural network and Fudenberg and Liang (2019) is a random forest.

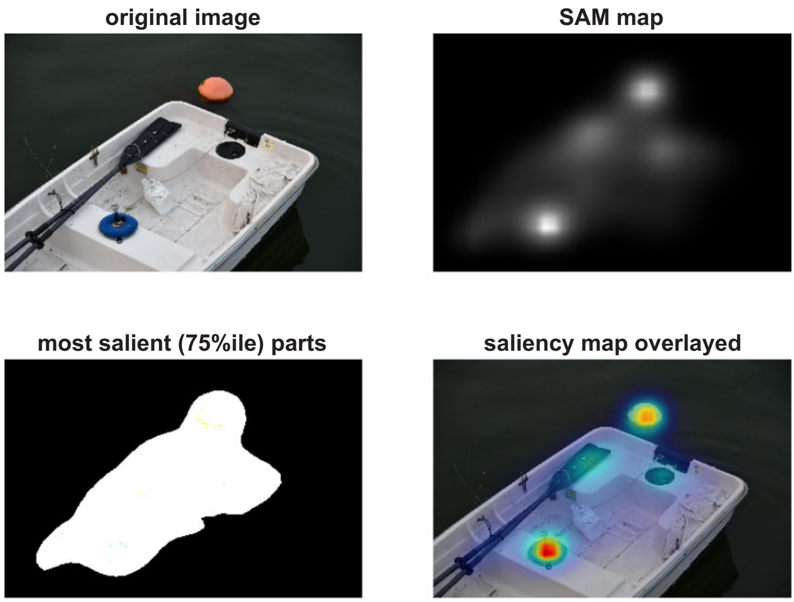

FIGURE I

A Salience Algorithm Example

Top left: An original image. Top right: The SAM saliency map, in which greater brightness indicates higher saliency. Bottom left: The area of the original image, which is 75% most salient. This area is generated from ranking all saliency values of each pixel. Bottom right: The original image with the saliency heatmap overlaid on it (warmer red colors indicate higher saliency). *Source:* Original photograph by Conor Wong Camerer (reproduced with permission).

In a larger incentivized experiment, $N = 61$ UCLA students earned \$1 if they matched. They chose the bridge 59% of the time (see Figure II, Panel B).[11] The SAM algorithm predicts that the bridge area, and the upper left road fork, are the most salient features (Figure II, Panel C).

Note that SAM does not predict that the "x spot" is salient, even though it was chosen by 25% of the subjects. For stimulus-driven algorithms, "x" is a special configuration of low-level features—two lines with diagonal orientation, that meet symmetrically in the middle. The x is also a high-level feature because it is a letter in many languages; that is, it is a recognized semantic object. However, the algorithm, as it was trained on other images,

11. These data were collected in conjunction with Milica Moormann and Alec Smith.

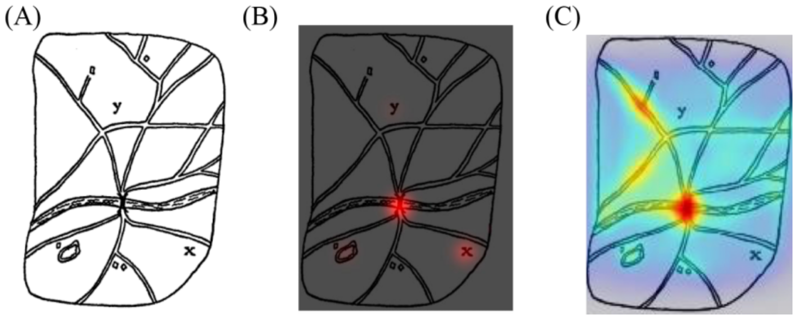

FIGURE II

Schelling's Map Revisited

Panel A: Original map. Panel B: Choice frequencies heatmap, where redness indicates choice frequency. Panel C: The SAM algorithm predicted salience heatmap. Panel A reproduced from Thomas C. Schelling, *The Strategy of Choice* (Cambridge, MA: Harvard University Press, 1960, 1980) Used by permission. All rights reserved.

was not originally capable of learning that "x" is familiarly known (to many UCLA subjects) to sometimes indicate locations of buried treasure on a map. So the x has minimal stimulus-driven salience, and SAM did not learn its goal-directed value for coordinating a meeting place on a map.

To distinguish the effects of purely visual salience and goal-directed attention further, we did an online experiment in which the 10 most prominent map locations were described in a verbal list. There was no accompanying visual map. Just as in the map experiment, matching the list choices of others gave a reward. The subjects' list choices were not the same as the map-based choices. The most popular choices were "x on the map" and "small house near the pond" (49% and 14%). Only 5% chose "bridge" (see Online Appendix F). This discrepancy shows that the popularity of the bridge choice depends on visual salience rather than its semantic content.[12]

II.A. *Explainable AI and the SAM Black Box*

Before proceeding, we note that the SAM algorithm is neither a model nor a mechanism, in the sense that economists typically use those terms. Neural network models (including SAM) are often

12. Rihn, Wei, and Khachatryan (2019) finds a related effect, that visual attention to a logo rather than text description of a type of plant changes valuation.

called “black boxes” because the basis of their predictions is in “hidden layers” that are difficult to interpret. One cannot readily do the comparative statics analysis that is useful in economics: for example, there is no simple mathematical way to easily compute how a change in an input image leads to a change in the outputted salience map.

However, an active area called “explainable AI” is concerned precisely with how to make opaque AI output more understandable (Hinton, Vinyals, and Dean 2015; Lipton 2018; Ras, van Gerven, and Haselager 2018; Arrieta et al. 2020; Belle and Papantonis 2020; Fan, Xiong, and Wang 2020).[13] Some progress has already been made in explainability for deep neural networks predicting visual salience. For example, He et al. (2019) used an image set in which a neural network predicts visual salience in a set of images. The categorical features in each image were also laboriously annotated by hand. That is, people looked at the images and coded the locations of vehicles, plants, animals, and so on. Then salience, as encoded at the middle-layer output of the neural network, was extracted (like examining a partially finished manufactured product). They found that the hand-coded categorized features were often correlated with the middle-layer salience predictions at these features’ locations. That correlation means that much of what the hidden middle layers were doing is learning the semantic categories of image features. In order of importance, 12 categories of features—a person’s head, “other”, an object, a person’s body part, etc.[14]—were most commonly encoded by the middle network layers.

The method just described is one way to measure the “feature relevance” of a predicted salience map. Feature relevance could be applied to all the images in our investigation as well, to improve explainability. For a set of maps like Schelling’s, each spatial location has one or more codeable features—the distance from the center, roads, forks in roads, ponds, rivers, houses, bridges, and so on (which were elements of the list version of the experiment). If these features and their locations are hand-coded, regressing

13. Igami (2020) explains the connection between some high-profile neural net training methods and structural estimation approaches invented in economics. This equivalence does not, however, guarantee the explainability of the content of the resulting neural networks.

14. The rest of the list is food, plant, symbol, vehicle, drink, animal head, and text.

the SAM salience values at each location against that location's features will measure how well the SAM salience values are approximated by a function of the coded features. A good fit means the black-box salience output is approximated by explainable features. The size and statistical strength of the regression coefficients indicate which features are most salient.

The Schelling map example sets up the empirical question in this article: How well does stimulus-driven salience—as predicted by SAM—predict actual choices in decisions and games? Does stimulus-driven salience get partially or entirely inhibited when there is also goal-directed attention?

We describe three experimental applications. They are:

i. Choices between visual images of two sets of fruits: The sets varied in induced values and in predicted salience. These data measure how often people picked lower-value sets because they were higher in stimulus-driven salience.
ii. Strategic choices of locations in visual images: In Schelling-style matching games, both players were rewarded if they matched by choosing the same location. In hider-seeker games, the hider wanted to mismatch and the seeker wanted to match. These data measure whether cognitive hierarchy or level-k structural models can fit data, and more ambitiously, make accurate cross-game predictions from the hider-seeker game to the matching game.
iii. Two-player 2×2 matrix games: These data measure whether stimulus-driven salience biases—which happen to predict looking at the top row and the left column in the matrices—can potentially explain strategy choices. This is a tough challenge for stimulus-driven theories because the experimental participants had a clear goal, to choose payoff-maximizing rows or columns. They may have ignored stimulus-driven salience entirely.

III. Decisions: Fruit Displays

III.A. *Study 1: Salience and Induced Value in Visual Fruit Displays*

The first experiment measured the empirical importance of visual salience in a simple setting that is lifelike. Subjects were shown two fruit sets presented on the left and right parts of an

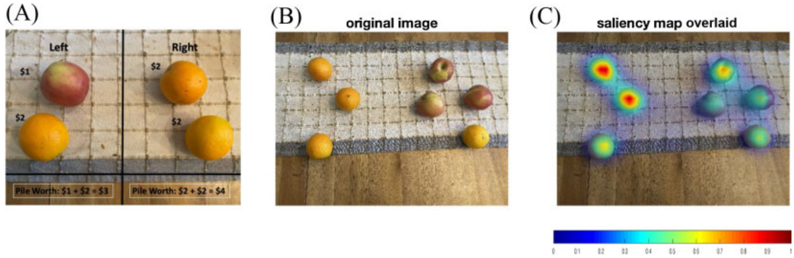

FIGURE III

Fruit Experiment Images

Panel A illustrates the rules of this task. Each fruit was worth a certain amount of dollars. The value of a set was the sum of all fruit values in that set. Panel B presents a sample image of an actual trial in this task, as subjects saw it (the dollar values were not shown). Panel C shows the SAM salience map for the sample image in Panel B. The left set was more salient than the right set in this example. All images used in this task had a salience distribution similar to this example, in that the salience peak is only distributed in one of the two sets. At the salience peak, the value of salience was 1 (the peak is located in the middle orange of the left set). In test images, the difference between the left and right salience peaks had an average difference of 0.23.

image, as shown in Figure III, Panel A. Each fruit type (e.g., apples or oranges) had a unique, predetermined induced monetary value (Smith 1976) that subjects learned before making choices. The induced values artificially created value, so there is an objectively best choice, and we can clearly judge if people are making mistakes.[15]

Ninety-seven participants did this study on Prolific (a European online data collection platform), following a preregistration process on the Open Science Foundation website (OSF).[16] All the participants were prescreened to have a prior approval rate of at least 70% based on their previous participation. Each subject was only allowed to participate in one experimental session (including pilot studies). Participation from mobile phones and tablets was not allowed to control for possible display effects.[17] There were five questions to check subjects' comprehension after the instruction session. We exclude individuals who failed more than one

15. We also hope that the induced monetary values swamped minor differences in intrinsic subjective value from personal or aesthetic preferences for fruits.

16. See https://osf.io/.

17. Even though computer screens also differ in size, phones and tablets have more variation in screen sizes.

question. See a full description of the experiment block design in Online Appendix G.G1 and Online Appendix Figure G1.

The total value of a fruit set is the simple sum of the values of all fruits in that set. The everyday analogue to this task is a retail vendor who is buying fruits at a wholesale market to resell and has in mind a retail price for each fruit. The retail price of the fruit induces value to the vendor. Subjects learned the induced values of different fruits before the main session of 20 choices.[18]

Although the vendor should optimally be computing resale value, the visually salient properties of fruit (such as color, intensity, and orientation), are hypothesized to influence stimulus-driven perception. The salience and value properties are independently controlled in the design.[19] In the choice sets, visual salience and fruit value were either positively or negatively correlated. The empirical question is whether subjects can ignore or inhibit visual salience, which is not generally correlated with induced value and could therefore lead to mistakes.

The main experiment included 20 images like those in Figure III. Choices were made with a 10-second time limit. Trials were balanced across induced values, numbers of fruits in the two sets, and whether the more salient set was on the left or right (see Online Appendix G.G2). Subjects earned money based on the induced value of the sets they chose in an incentive-compatible design (a 10% chance of earning the value of what they chose on one randomly selected trial).

The average difference between the most salient peaks in the two fruit sets was 0.23 on the 0–1 scale of salience. More ambitious designs could obviously covary the size of the salience difference and the size of value difference between the two sets. In half of the trials, SAM-salience and induced value are “congruent”—one set is higher in both salience and induced value. In the other half of the trials, they are “incongruent”—the high-salience set has a lower induced value or vice versa.

The dependent variable is 0-1 choice accuracy—did they choose the most highly valued set? With a 10-second time limit,

18. They experienced an untimed but incentivized session before the main session. More experimental details are in Online Appendix G.G1.

19. It is possible that stimulus-driven salience of fruits is correlated with their subjective value in the natural ecology—for example, brightness might be visually salient and also correlate with ripeness and fruit taste or nutrition. However, even if this is the case, by design stimulus-driven salience is uncorrelated with induced value, which is the only type of value a payoff-maximizing agent should attend to.

TABLE I
INFLUENCE OF SALIENCE-VALUE CONGRUENCY IN A SIMPLE CHOICE PROBLEM (FRUIT SETS)

	Dependent variable: Accuracy (0,1)				
	(1)	(2)	(3)	(4)	(5)
Congruency	0.83***	0.90***	0.89***	0.97***	1.26***
	(0.32)	(0.29)	(0.33)	(0.31)	(0.41)
abs(valueDiff)		0.80***		0.80***	0.77***
		(0.23)		(0.23)	(0.23)
Interaction:					− 0.55
Congruency*abs(valueDiff)					(0.63)
Constant	1.54***	0.78***	1.99***	1.25**	1.26**
	(0.10)	(0.17)	(0.61)	(0.57)	(0.57)
Covariates	No	No	Yes	Yes	Yes
Observations	1,382	1,382	1,307	1,307	1,307
Log likelihood	−644.7	−591.8	−607.5	−556.7	−556.3
Akaike inf. crit.	1,293	1,189	1,239	1,139	1,141

Notes. The congruency variable is the difference between the the maximum salience level of the more valuable set and the maximum salience level of the less valuable set (between 0 and 1). This variable will be positive if one option is both more salient and more valuable. abs(ValueDiff) is the absolute value of the induced value difference between left and right sets. Standard deviations are clustered on the per subject level. “Covariates” denotes whether the current model contains covariates: education, gender, income, and self-reported fruit preference (we ask them which fruit they prefer in everyday consumption: apples, oranges, or equal preference). The main effect estimates are not sensitive to these covariates, as is evident comparing specifications (1–2) to (3–4). * $p < .1$; ** $p < .05$; *** $p < .01$

choice accuracies were 85% and 79% in the congruent and incongruent conditions. This drop in accuracy, when salience conflicts with valuation, is highly significant (p-value = .002, two-sided t-test).

We test for the effect of salience, controlling for the value gain from choosing correctly, using a logistic regression of the form:

$$y_{ij} = \beta_1\left(S_j^H - S_j^L\right) + \beta_2 abs\left(V_j^L - V_j^R\right) + \beta_3\left(S_j^H - S_j^L\right)abs\left(V_j^L - V_j^R\right) + \beta_4 X_i + \epsilon_{ij}, \tag{1}$$

with robust standard errors clustered at the subject level. The variable y_{ij} is accuracy (a 0-1 dummy variable, for person i at image j); V_j^L and V_j^R are the monetary values of the left and right sets in image j, and $abs(V_j^L - V_j^R)$ is the absolute induced value difference (abs(valueDiff) in Table I). The congruency variable defined earlier is $S_j^H - S_j^L$, the difference in salience of the high- and low-valued sets. We are therefore regressing choice accuracy on congruency, absolute value difference, their interaction, and covariates.[20] The results are summarized in Table I. The induced value difference and congruency variables are both significantly associated with choice, with comparably large t-statistics (around 3–4).

There are two boundary conditions in which the effect of salience disappears. When the value difference is large the accuracy is 94% for both congruent and incongruent conditions (p = .91 for the test for a difference). When the value difference is small, the accuracy is lower and salience-value incongruence does have an effect (78% versus 69%, p = .01). (The Table I results pooled both types of images).

The second boundary condition is endogenous time allocation: When there is no time limit ($N = 22$),[21] participants in both conditions are near the ceiling of perfect accuracy (congruent 94% and incongruent 96%).

At this point, readers may be curious why subjects don't just ignore the stimulus-driven salience of the fruits. The reason is that in economics jargon, perceptions are not freely disposable. The visual perceptual system is highly evolved to distill a huge amount of visual input into a much smaller amount of useful information and not waste the small amount that seems useful. The fastest parts of that process occur implicitly (without conscious awareness) in less than a second. Inhibiting any rapid highly evolved implicit behavior is mentally difficult. One type of evidence about inhibition difficulty is that exogenous manipulation of attention—adding more "involuntary" attention to a choice object—increases later choice of that object (albeit by a small amount; see Shimojo et al. 2003; Armel, Beaumel, and Rangel 2008; Pachur et al. 2018; see Mormann and Russo 2021 for a contradictory view).[22]

A mechanistic explanation for why irrelevant salience affects choices comes from a popular class of psychological models for how attention and decision time influence choice. These

20. "Covariates" is yes when the current model contains covariates of education, gender, income, and self-reported fruit preference.

21. An additional group of subjects collected on Prolific did only the unlimited time experiment.

22. A related phenomenon is called the "mere exposure" effect in psychology. Mere exposure means that repeated presentation of one unfamiliar stimulus tends to slightly increase expressed likings for that stimulus, compared with similar stimuli with less exposure (see Zajonc 1968; Bornstein 1989 for meta-analytic review).

"accumulators" (or diffusion drift) models assume that over time perceptions and memory cumulate a running value of a latent numerical "evidence" variable (Ratcliff 1978; Ratcliff et al. 2016; Fudenberg, Strack, and Strzalecki 2018). A choice is made when the variable level crosses a mental threshold or barrier. In these models, if stimulus-driven initial perceptions enter the accumulator variable, there is no known mechanism that will fully erase their effect at low cost. If the time to decision can be endogenously chosen by the decision maker, then a very high threshold can be set which will dilute the early effect of stimulus-driven perceptions, but will not always fully inhibit that effect. (This is consistent with the absence of a salience effect in untimed trials.)

A different way to model why stimulus-driven perception influences choice comes from the signal-extraction model of Cunningham (2013). In that model, an "upstream" sensory system sends information to a more "informed" downstream system and the two kinds of information are integrated. However, the downstream system only has partial information about input to the sensory system. Intuitively, the brain may partially accumulate the stimulus-driven perception into a decision variable as if it might have come from value-driven attention.[23]

IV. Study 2: Matching and Hider-Seeker Location Games

This section reports new experimental data from location games. Schelling's map game is an example of a location game. In our general location games, two players saw a common visual image and simultaneously choose a location—a pixel. A circle was created around the pixels (with a radius of 108 pixels). The circle was about one-fifth of the screen width. The baseline circle size was chosen so that if players were choosing pixels randomly, they would match 7.1% of the time. (One experimental treatment below varied the circle size.)

In matching games, both players wanted to match by choosing locations that had overlapping circles. In hider-seeker games, seekers wanted to match and hiders wanted to mismatch. Interactions of the hider-seeker kind include predator-prey relations in nature. Human examples include choosing passwords to outwit hackers, other "coded" language and signals used in sports,

23. This kind of upstream-downstream integration is likely to be common in the brain, leading to illusions like the atmosphere illusion (people do not fully undo the effects of unusual foggy or clear days on distance perception).

gangs, and other rivalries to coordinate action with teammates and avoid detection by the other side. Industries such as fashion can have follower-leader dynamics (e.g., fashion leaders want to "hide" by choosing unique new designs, and outsiders want to "seek" by matching those designs which induce hider-seeker structure). Visual salience might conceivably play a role in some of these games.

The experiment had three blocks of games (see Online Appendix Figure H1): matching, the hider-seeker game in the role of seeker or hider, and the hider-seeker game in the opposite role of the one in the second block. The matching block always came first, followed by the hider and seeker blocks in a randomized order between subjects. During each block, there was a feedback sequence in which the choice the other player made was revealed to a player right after both choices, by showing the circle around the other player's pixel choice and the player's own circle. In a "no feedback" sequence, those results were not revealed.

The matching block had two sets of 20 images for each of the two feedback treatments (40 images in total). The hider-seeker game used a different set of 19 images for each of the two feedback treatments (38 images in total). For each image, subjects played once as a hider and once as a seeker. An additional short session of hider-seeker games followed in the last block (16 images) with a bonus payment 10 times higher than in the baseline, to test for effects of higher incentives.

There was unlimited time to read instructions but only six seconds to make a choice. Subjects got no payoff if they didn't respond before the known time limit (see the instructions in Online Appendix H). The results shown to subjects in the feedback condition were drawn from previous choices of actual subjects (using different previous subjects for each image).

One hundred fifty-one subjects participated, excluding a pilot data set for power analysis. Of these 151 subjects, 29 people (13 men, 16 women) participated in the lab, one at a time, in a small testing room where their eye movements were recorded. Fifteen of those subjects were from the Caltech community and 14 from the neighboring community (there were no differences in results between the two groups). The bonus payments were \$0.20, \$0.10, and \$0.40 in matching, hiding, and seeking games, respectively, for each "win" per trial (image). Participants were paid the cumulative monetary amount at the end of the experiment. In the lab experiments, all the visual images were displayed on a computer screen in 1920 × 1080 resolution. The other (N = 122) subjects participated online through Amazon Mechanical Turk (MTurk).[24] Images were randomly selected from a large image pool (273) with five categories (abstract art, city, faces, social scenes, and nature). The image set contained images with only one obvious salience center and more complex images that have multiple salience centers (Judd et al. 2009).

There were some behavioral differences between choices in the feedback and no-feedback conditions.[25] The largest effect is that the matching rate is higher with feedback than with no feedback (64% versus 35%). However, the seeker win rate in hider-seeker games is the same in both conditions (9%) and most other differences are not substantial. We therefore report only data from the feedback condition in this main text. The corresponding no-feedback results are in an Online Appendix D.

Figure IV shows examples of result screens that subjects saw during the experiment.

IV.A. Analysis and Results

Equilibrium game theory generates a statistical benchmark for what people might do.[26] In location games, strategies are pixels in x-y space (and their resulting circles).

For the matching coordination game, choices by the two players of any two pixels that create overlapping circles constitute a

24. Online experiments have the same instructions and block orders as the in-lab version, except that everything is shown in a web browser. This study was preregistered on the Open Science Framework (https://osf.io/yuqjg/) during data collection and before analysis. The sample size was predetermined before the data collection process, based on a pilot study (N = 29) carried out in March 2017.

25. Both feedback and no-feedback blocks were included because each one answers a different question of interest. To help ensure increased subject comprehension in learning-by-doing, and especially in testing equilibrium concepts, the standard practice in experimental economics is to provide feedback. However, whether salience is predictive even with no feedback is an interesting question, too. That is why we did both.

26. A game-theoretic idea which might help explain how salience influences choices is correlated equilibrium (Aumann 1974). When both players receive a common public signal and a strategy is conditioned on the signal values, a correlated equilibrium occurs when nobody wants to deviate from recommended strategies. Stop signs and green-yellow-red traffic lights, for example, act as correlating devices (also enforced by law) to create a commonly observed visual signal that coordinates traffic and reduces accidents. In these terms, our study is about whether the stimulus-driven visual salience of image locations works as a correlating device in matching games.

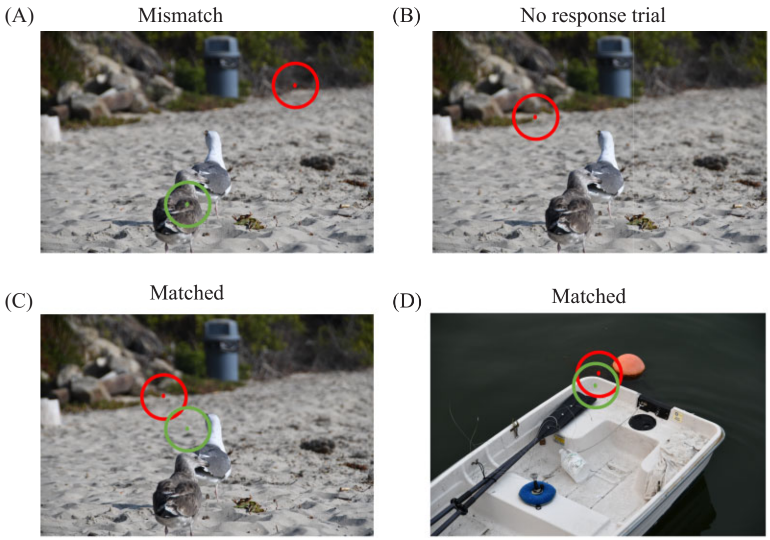

FIGURE IV

Examples of Trial Outcomes with Feedback, Showing Circled Pixel Choices

Source: Original photographs by Conor Wong Camerer (reproduced with permission).

pure strategy Nash equilibrium. One image contains about two million (1,920 × 1,080) pixels. Since any pixel match is a pure equilibrium, there are an enormous number of equilibria. There are also many mixed equilibria. So standard equilibrium theories do not rule out any of the location choices.[27]

For the hider-seeker game, there is a unique Nash equilibrium in which all locations are chosen equally often.[28] The fact that equal randomization over all strategies is the unique hider-seeker

27. Note that if players have a personal utility from picking a specific location or a type of image feature, such preferences might conceivably reduce the set of equilibria, particularly if a selection principle such as payoff-dominance is applied (see Bacharach (1993); Bacharach and Bernasconi (1997)). However, such results would likely be sensitive to whether such preferences were commonly known.

28. For those unfamiliar with game theory, intuition can be gained by a simplified example. Suppose there are just two locations and the hider chooses them with probabilities p and $1 - p$. If the seeker matches those probabilities, she has a $p^2 + (1 - p)^2$ chance of winning. This sum is always lower if the seeker chooses the most likely spot (i.e., the location with $p > 0.5$) because if $p > 0.5$, then $p > p^2 + (1 - p)^2$. To defend against this, the hider should mix equally, so $p = 0.5$. Every new location that is added should also have a $\frac{1}{n}$ chance of being chosen (if there are n lo-

equilibrium is an example of how game theory logic conflicts with the result of human biology. We are so good at quickly noticing salient information, while amateurs at rapidly choosing what is unsalient to hide.[29] The last thing the brain is equipped to do is to ignore salient differences among many objects and choose them equally often.[30]

IV.B. Matching Games

To analyze the behavioral data, we test whether subjects are playing an equal random mixture across all pixels and their associated salience levels. To compare results from different images, all salience values in this section refer to the normalized levels, which are the rank percentiles of raw measures from the algorithm, ranked in each image. We calculated the normalized salience value for each chosen pixel and then compared these salience values against the baseline of equal randomization independent of salience. Kolmogorov-Smirnov tests reject the hypothesis of equal randomization for all treatment conditions ($p < 10^{-4}$). Subjects' choices are not independent of salience.

To see examples of how salience affects choices, the choices from all the subjects are plotted on one specific image in Figure V.

The salience heat map is in the middle column. The right column shows, using red dots, the subjects' actual location choices. The predicted salience in the middle column and the observed choice maps in the right column are highly overlapping.

---

cations) by an iterated logic. A special design that if a circle touches any boundary, it wraps around from the opposite boundary, guarantees the equilibrium.

29. A similar conflict between logic and biology occurs in the games "rock, paper, scissors" (e.g., Crawford, Costa-Gomes, and Iriberri 2013). When players display the three choices with their hands, there is a slight tendency to match an opponent's choice (e.g., playing rock against rock) more often than predicted in equilibrium. The explanation is that imitation of another person's body movements is such a highly adapted automatic behavior, that the brain cannot inhibit the response, even though it reduces performance (e.g., you should play paper rather than imitating rock).

30. The difficulty of inhibiting certain kinds of perception is illustrated by Steinbeck (2011). In *The Pearl* the protagonist, Kino, has hidden a valuable pearl that everyone in the small town covets. An unscrupulous doctor comes to treat Kino's baby, hoping to find out about the pearl. "The doctor shrugged, and his wet eyes never left Kino's eyes. He knew the pearl would be buried in the house, and he thought Kino might look toward the place where it was buried. "It would be a shame to have it stolen before you could sell it," the doctor said, and he saw Kino's eyes flick involuntarily to the floor near the side post of the brush house."

FIGURE V

Two Matching Game Images, Salience Heatmaps, and Choices (red)

(Left) The original image. (Middle) The original image overlaid by the SAM salience maps. (Right) The grayscale original image overlaid with the actual empirical choice distributions (each red dot is one choice). This is a derivative of "Zachary Bedrosian" by Zachary Bedrosian and is licensed under the Creative Commons CC0 1.0 Universal Public Domain Dedication.

Statistically, the mean salience level of the pixel locations chosen in the coordination game is 0.87. This is far above the chance level of 0.5 ($p < 10^{-4}$).

IV.C. How Predictable Is the Matching Rate across Images?

Intuitively, the matching rate for an image should be affected by how dispersed salience is. When salience is highly concentrated, the rate of choosing the same pixels, and matching, should increase. If salience is not highly concentrated, the matching rate should be lower.

Dispersion of salience throughout an image can be measured by the number of local salience centers.[31]

Online Appendix Figure A1 gives two concrete examples of different numbers of salience centers and the corresponding game results.

Figure VI shows that, indeed, the matching rate[32] is strongly negatively correlated with the number of salience centers

31. The typical raw salience map has flat local maxima with many adjacent pixels with nearly equal salience. To detect salience centers we first apply a Gaussian smoothing (with [300 pixel, 300 pixel] window size and standard deviation $\sigma = 75$ pixels) to the entire image to smooth hyperlocal spikes in salience. Then we simply take the number of local maxima for the salience distribution using the Matlab function imregionalmax() with default settings. That function takes the local maximum inside each 3 pixel $\times$ 3 pixel patch. If the original image has two local maxima that are close enough together, the Gaussian filter combines them.

32. This result is based on all images from both the feedback session and the no-feedback session using the in-lab data set (image $N = 40$).

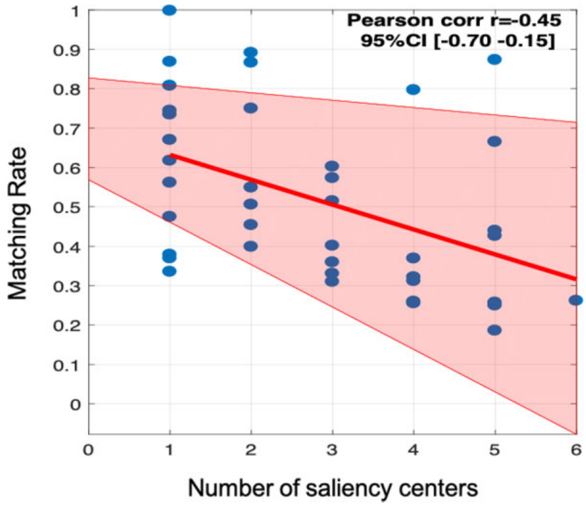

FIGURE VI

Correlation across Images between Matching Rate and Number of Salience Centers

The figure plots the correlation between the number of salience centers and the matching rate using both the feedback session and the no-feedback session to get a larger image pool ($N = 40$ images).

(Pearson $r = -0.45$, $p < 10^{-4}$, df $= 38$).[33] Online Appendix Figure A1 gives out two concrete examples of different numbers of salience centers and the corresponding game results.

The matching rates span a range from a high rate of about 75%, for one salience center, to just above random (20%) for seven salience centers. These results suggest that for any image, the matching rate could be predicted ex ante with substantial accuracy from the salience map, before any data are collected. Put the other way around, it is possible to find images with salience distributions that will predictably yield either near-perfect matching or near-random matching. This could be a useful tool for designers who are trying to either enhance shared attention or undermine it.

33. At a reader's suggestion, we also calculated whether the number of salience centers was correlated with the seeking win rate in hider-seeker games (across the $N = 38$ images). This is an interesting question because if there are many strategically naive hiders, the correlation will be positive. However, there is no correlation (Pearson $r = -0.10$, $p = .23$).

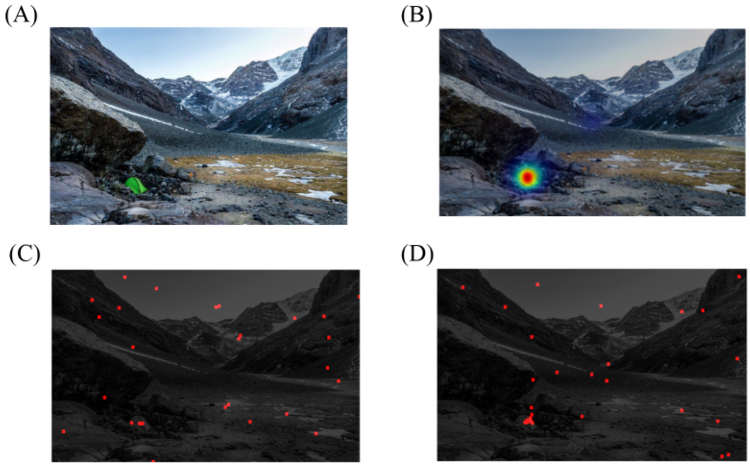

FIGURE VII

A Hider-Seeker Game Image, Salience Map, and Choices

Panel A: The original image. Panel B: The original image overlaid by the salience map. Panels C and D: The grayscale original image overlaid with the actual empirical choice distributions (each red dot represents an actual choice from one person). Panel C is for hider choices and Panel D is for seeker choices. This is a derivative of "Poniente Yerba Loca" by Pierre Bouillot and is licensed under the Creative Commons CC0 1.0 Universal Public Domain Dedication.

IV.D. Hider-Seeker Games

For the hider-seeker game, we start with an example image and data. Figure VII shows that subjects' choices are more spread out than in the matching game examples (Panel C shows hiding data and Panel D shows seeking data.) In Figure VII, Panel C, there is no distinct peak of the hider choice distribution, and few choices are in the most salient area.

The direction of effects suggested by this example holds more generally. The mean salience levels of hider and seeker click points were 0.53 and 0.61, close to the chance level of 0.50.[34] The same in-lab group ($N = 29$) with payoffs 10 times higher had very similar results, averaging salience levels of 0.51 and 0.64 for hiders and seekers.[35] A paired t-test showed this difference in choice salience between hiders and seekers is highly significant ($p < 10^{-4}$),

34. $p = .02$, t-test CI: [0.51,0.56].

35. Hiding: p-value for test against null of 0.50 salience = .59, CI: [0.48,0.54], seeking: p-value $< 10^{-4}$.

TABLE II
REALIZED MATCHING RATE

	Matching rate	N of observations
Nash mixed prediction	0.071	
Matching game	0.64 (0.006)	559
Hider-seeker game	0.09 (0.002)	1,060 (531(H), 529(S))
Hider-seeker game (between-subjects)	0.09 (0.002)	1,325 (600(H), 725(S))
Hider-seeker high payoff (10×)	0.09 (0.003)	892 (446(H), 446(S))

Notes. Statistical tests against the null hypothesis that the seeker win rate is the baseline level and choices are independently and identically distributed across subjects (which is the Nash benchmark prediction). The number in the bracket is the standard error of the seeking win-rate in each condition.

reflecting what is suggested by the Figure VII example. The no-feedback results had a similar difference (see Online Appendix D).

1. *Seeker's Advantage.* Recall that the theoretical frequency with which two randomly chosen location circles will match is 0.071. Table II presents the realized matching probability in each specific game condition.[36]

To check robustness, the hider-seeker game experiments were replicated in two other conditions: a high-payoff condition with payments 10 times as large ($N = 29$)[37] and a between-subjects condition where subjects played only one of the two hider or seeker roles across all their trials (53).[38] In both conditions, the seeker win rate was 9.0%, the same as in the baseline experiment. All differences from the equilibrium prediction of 7.1% were highly significant.[39]

36. Tests to compare the matching rates with random baseline were carried out by bootstrapping a person's hiding data and a different person's seeking data (or two data points from matching game) for 1,000 batches (batch size is a total number of different pairs). We get the empirical distribution for the matching rate and statistical significance against baseline 0.071 from that bootstrap. Specifically, each sample is drawn by matching two random users (different ones). The batch-seeking win rate is calculated accordingly. All values were calculated from the average of 500 iterations of randomly matching two data points from the data set if two subjects were in the same subblock, same image.

37. They did this session at the end of the in-lab group experimental session. See the full batch description in Online Appendix Table H1.

38. This was an mTurk separate sample; see Online Appendix Table H2.

39. The 9% win rate for seekers does not seem to be much larger than the equilibrium prediction of 7%. However, under the null hypothesis of Nash equilibrium, this win rate should be identically distributed for all images, and for all

To test whether the seeker advantage is only present under time pressure, 46 people from MTurk participated in the same hider and seeker experiment, but without a time limit. The seeker's win rate was again 9% ($p = .002$ for comparison with Nash benchmark 7.1%). Subjects spent an average of 3.14 seconds, 4.61 seconds, and 6.44 seconds in matching, hiding, and seeking conditions, respectively, when there was no time limit.[40]

The seeker's advantage could depend on the size of the circle that is drawn to surround the chosen pixel. To explore this possibility, in another experiment the circle size was enlarged to be 1.5 times as large as in the original experiments. Then the chance/equilibrium matching rate is about twice as high, 16%. The seeker win rate was 18%, so there is still a small seeker's advantage exactly equal in absolute size (+2%) to the benchmark circle results ($p = .003$, $N = 66$).

The seeker's advantage must be due to a correlation between the hiders' choices and the seekers' choices, which should not happen in equilibrium (except for sampling error).[41] We have already shown that hiders and seekers choose slightly higher salience locations, but at different frequencies. How exactly do those biases lead to the seeker's advantage? Figure VIII presents the seeking win rates conditional on different salience levels for hiders and seekers. The seeker's advantage is mainly due to the concentration of wins when both players choose locations that are in the top 10% in salience.

V. A Salience-Influenced Cognitive Hierarchy (SCH) Model

This section describes a parametric behavioral model meant to explain choices and their salience sensitivity, closely following

people. This null hypothesis supplies a lot of statistical power. A more conservative approach averages all data within an image and tests whether the image-wise matching rates are above 7% ($N = 19, p = .0005$). A different conservative approach averages win rates for individuals and tests whether the average individual seeker win rate is different than the Nash 7% ($N = 29$, $p = .002$).

40. The standard deviations were 7.10 seconds, 15.54 seconds, and 19.49 seconds for matching, hiding, and seeking. These large standard deviations are not unusual for an online experiment with unlimited time because some subjects take much longer time than others.

41. We know that people are capable of approximately equal randomization in these games because when they play a random computer opponent their choices are approximately equally random (Heinrich and Wolff 2012).

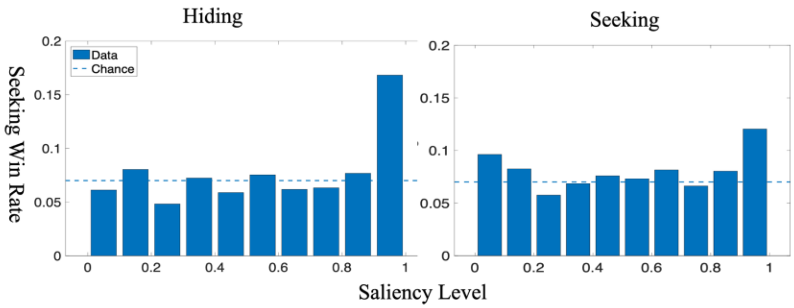

FIGURE VIII

Seeking Win Rates as a Function of Different Saliency Levels

This figure shows the average seeking win rate of hiders and seekers separately, at each saliency level bin from 0 to 1 (with bin size 0.1). This conditional seeking win rate looks a little different between hiders and seekers mainly because their choices are distributed differently across saliency levels.

Crawford and Iriberri (2007a). It uses the level-k model of Stahl and Wilson (1994) and Nagel (1995), later extended by Camerer, Ho, and Chong (2004).

The SCH model combines cognitive hierarchy levels, a quantal response function (softmax), and a salience-influenced level-0 assumption.

V.A. *General Model Description*

The population consists of different levels of players starting from level 0. The proportion of level-k players is $f(k)$, with $f(k)$ assumed to be Poisson distributed with parameter τ.

For all levels of players, there is randomness that will be described using a conventional logit softmax function $\frac{e^{\lambda x_n}}{\sum_m e^{\lambda x_m}}$ with parameter λ. Higher λ corresponds to more sensitivity to x_n.

In this SCH specification, the nonstrategic level-0 players weakly prefer salient choices. The probability of choosing strategy/pixel n depends on the direct salience value[42] S_n of that pixel from SAM according to:

$$P_{0n} = \frac{e^{\lambda(1+\mu S_n)}}{\sum_m e^{\lambda(1+\mu S_m)}}.$$

42. Just as before, the salience values refer to the normalized ranking with respect to each image. This way, we can use data from different images and salience distributions in a common specification.

If $\mu = 0$, salience is ignored and level-0 types choose randomly among all points. We assume that λ and the salience weight μ are common across subjects, although heterogeneous versions could be used (e.g., Rogers, Palfrey, and Camerer 2009).

All levels of players above 0 behave in the same way as in a standard cognitive hierarchy model. Level-k players assume that all other players are only of lower levels (0 to $k-1$), using normalized Poisson frequencies $f(k)$. A level-k player calculates the expected payoffs of choosing n, denoted as EU_{kn}. The probability of a level-k player i choosing option n is:

$$P_{kn} = \frac{e^{\lambda EU_{kn}}}{\sum_m e^{\lambda EU_{km}}}.$$

Note that salience only enters directly into the value calculations of level-0 players. This assumption tests whether a model in which salience only enters $k \geqslant 1$-level players through beliefs (and hence uses goal-directed attention) is a good approximation.[43]

V.B. Model Fitting Results

Besides the SCH, there are many other ways to specify models of limited strategic thinking, which have been mixed and matched in previous research. We fit six model specifications to the hider-seeker data (see Online Appendix E).

Some specifications restrict the frequency of actual level-0 types to be zero, $f(0) = 0$, as if level-0 players are only a figment of the imagination of higher-level types (though see Wright and Leyton-Brown 2019). Restricting $f(0) = 0$ in this way clearly degrades fit (Online Appendix Table A2). We therefore focus only on $f(0) > 0$.

A close relative of SCH is the level-k model, in which level-k types believe all others are level $k-1$ (rather than distributed from 0 to $k-1$ as in SCH) (Crawford and Iriberri 2007a, 2007b). Level-k is usually estimated nonparametrically, allowing all frequencies $f(k)$ (up to some maximum k) to be estimated separately.

Both SCH and level-k specifications with role-specific level frequencies fit the overall data about equally well by the Akaike information criterion (AIC) criterion (although SCH is a little better by the Bayesian information criterion, BIC). These games are

43. This is similar to Mehta, Starmer, and Sugden (1994) for matching games, in which secondary salience is derived from primary salience.

TABLE III
ESTIMATION DETAILS, ROLE-SPECIFIC SCH

	λ	μ	τ_h	τ_s
Best-fit parameters	100	0.06	0.4	0.1
Number of observations	1,096 for hiders and 1,090 for seekers			
95% CI	[72.3,100]	[0.05,0.08]	[0.32,0.47]	[0.08,0.13]

Notes. The parameters μ and λ are constrained to be the same for hiders and seekers. The confidence interval in the table is calculated using the bootstrap method with data batch size 1,096 for hider, 1,090 for seeker, and the number of iterations is 100.

not an ideal testing ground for comparing such differences. The goal, instead, is to see if either SCH or level-*k* variants can explain both matching and hider-seeker games, which have different goal-directed attentional demands.

We first focus on the preferred specification of SCH. It has four free parameters: μ, the salience weight parameter; λ, the softmax parameter; and two role-specific parameters τ_s and τ_h, which are the Poisson distribution parameters of strategic levels for hiders and seekers separately. (Allowing different λ and μ parameters for hiders and seekers fits worse due to the large BIC penalty for extra parameters).

We used a standard training-testing separation to avoid overfitting. Recall that each subject did two sessions.[44] We use the first session data as a training set to estimate parameters. The parameter values are then fixed and used to predict data from the second session test set (see Online Appendix). The best fitting parameter values and measures of fit are shown in Table III.

Figure IX compares the actual choice density (frequency) function and best-fit model predicted density functions for the hider-seeker game. Training data are shown in the top of Figure IX, Panels A and B, and test data are shown on the bottom of Figure IX, Panels C and D. In the choice data, there is a sharp density increase starting around 0.9 salience for both roles (although note that the *y*-axes are different, so the actual increase is about half as big for hiders as for seekers). There is also a smaller trend of slightly decreasing choice from the very lowest salience to medium salience levels for hiders (but not for seekers). This small dip reflects the fact that some hiders did manage to strategically

44. Two sessions contain different image sets. A first session of normal payment trials including feedback and no-feedback trials and a second session of high-payment trials.

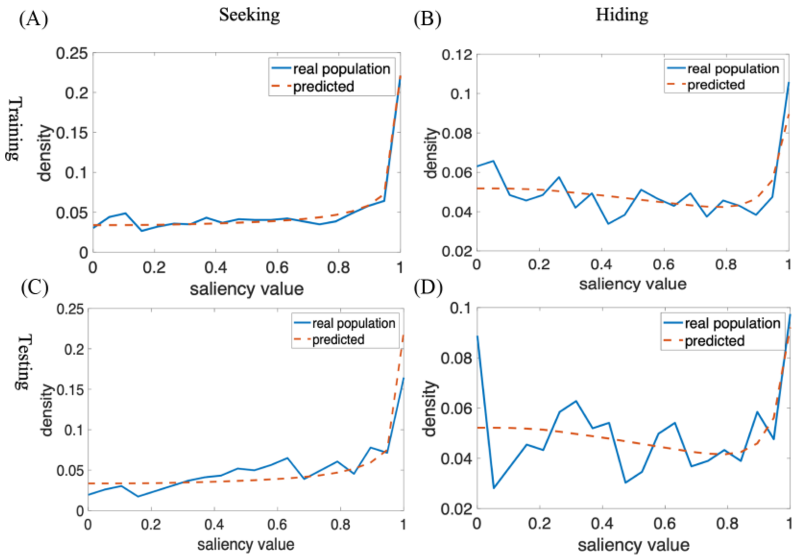

FIGURE IX

Frequency of Choice by Salience Level with Model Fitted Distributions

The graphs indicate what percentage of choices were made for locations with the salience of those locations on the x-axis. Panel A: Choice data and model prediction in the training data set seeking condition. Panel B: Choice data and model prediction in the training data set hiding condition. Panel C: Choice data and model prediction in the testing data set seeking condition. Panel D: Choice data and model prediction in the testing data set hiding condition.

choose the lowest-salience locations. SCH can roughly fit these two major features of the data.

However, the best-fit values of τ, 0.4 and 0.1 for hiders and seekers, are much lower than typical estimates around $\tau = 1.5$ (e.g., Camerer, Ho, and Chong, 2004; see also Riche et al. 2013, although Fudenberg and Liang 2019 find minimal prediction error in a large interval (0, 1.25) including low τ values).

The low values of τ estimated for SCH result from the fact that the ability to identify τ is limited in these visual choice games. A single-peaked SCH with Poisson $f(k)$ does not meet the calibration challenge well. Level-1 hiders should anticipate high-salience choices by level-0 seekers and move sharply to antisalient locations. But there are not that many low-salience choices in the hider data (as Figure IX, Panel D shows). The SCH distribution explains the infrequency of low-salience hiding the only way it can, by simply estimating few level-1 types through a low value of τ.

The level-k model gives better insight here about plausible level frequencies.[45] Compared with SCH, the best level-k specification estimates lower frequencies of level 0 ($\hat{f}_s(0) = 0.17$ and $\hat{f}_h(0) = 0.29$) for seekers and hiders, and a higher salience weight $\hat{\mu} = 0.18$ for level-0 types. Level-k also estimates larger frequencies of level-2 and -3 types ($\hat{f}_s(3) = 0.66$, $\hat{f}_h(2) = 0.61$). Although the overall level-k fit is just a little less accurate than SCH, this type distribution is more consistent with experimental results than the SCH estimates of low τ (see Online Appendix E). So while it is clear that both specifications fit the salience-choice profiles adequately (as seen in all the figures, including Online Appendix Figure D1), they suggest different evidence of level frequencies. These games were chosen to investigate the effect of predictable salience but were not ideal to recover levels accurately. Better methods can be developed.

V.C. Cross-Game Predictive Validation

To further test generalizability of SCH, parameters estimated from fitting the SCH model to hider-seeker data will now be used to predict choice behavior in the matching game. There is no guarantee that this cross-game portability will work at all (see Hargreaves Heap, Rojo Arjona, and Sugden 2014; but see Crawford 2014). Identification of the salience weight μ in hider-seeker games comes purely from the level-0's choices and from higher-level player beliefs and choices. In the matching games, all higher-level types are similarly guided by goal-directed attention because they are all trying to match the lower-level types. The strength of salience sensitivity that is estimated in the two cases could easily be different. Furthermore, matching and hiding are completely opposite in strategic motives.

Figure X compares predictions of the salience-frequency profile on the test set of matching game data. The left graph shows predictions based on using hider-seeker training, that is, the free parameters are trained on the hider-seeker data, then fixed and used to predict ("test on") the matching game results. The right

45. A better way to identify τ is by creating games in which different level types choose distinct strategies (such as in the matrix games pioneered by Stahl and Wilson 1994, and see Nagel 1995; Ho, Camerer, and Weigelt 1998; Costa-Gomes and Crawford 2006; Kneeland 2015).

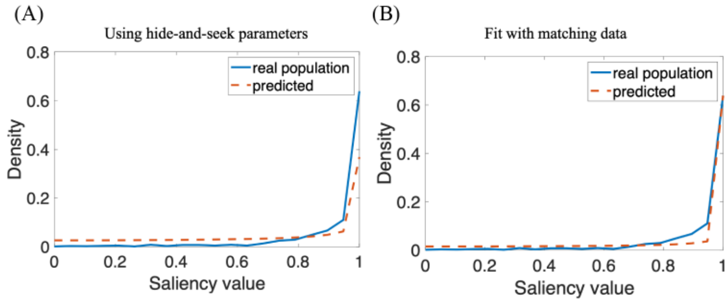

FIGURE X

The SCH Model Calibrated on Hider-Seeker Game Data Can Predict Matching Game Choices.

The comparison is between the matching data distribution and the two fitted matching game distributions. Panel A: Parameter estimates from the hider-seeker game are used to predict matching game results; log-likelihood: −2,176. Panel B: Parameter estimates from the training matching game data are used to predict test matching game data; log-likelihood: −1,943.

graph shows predictions of matching test-set data using matching data for training (i.e., using the two-session train-test cross-validation described already). Of course, training on the matching data and then predicting matching test data should be more accurate than training on a different type of game, and it is (LL = −1,943). However, training on the hider-seeker data and testing on matching is only about 10% worse (LL = −2,176). Comparing Figure X, Panels A and B shows that the main difference is that the hider-seeker trained parameters underestimate how sharply matching-game test data respond to the highest salience.[46]

The hider-seeker structure is a good example of how stimulus-driven and goal-directed salience can be combined. Level-0 players are only influenced by stimulus-driven salience (from the SAM algorithm) because they do not have a strategic goal. Higher-level types need to compute expected values of strategies, which requires goal-directed attention. But they also form beliefs about level-0's that requires simulating the stimulus-driven attention of level-0's. Therefore, both types of attention need to be

46. We did not do the opposite analysis, predicting hider-seeker data based on parameters estimated from a matching game. The meaning of doing this opposite analysis is limited because of the identification problem. Using matching game data only is not enough to identify the strategic level parameters because all level players are using similar strategies of choosing salient locations.

combined to make good choices. The fact that hiders lose more often than expected in equilibrium is associated (via the structural model) with the fact that they are choosing too many locations with stimulus-driven salience. Their goal of hiding, which should guide attention to low-salience locations, does not appear to sufficiently inhibit stimulus-driven salience.

VI. Study 3: Matrix Games

Location game experiments are unusual. Most game theory experiments, following visual conventions in textbook game theory, use normal-form games in a matrix format (or occasionally game trees). To establish boundaries of where visual salience is predictive and where it is not, it is therefore useful to ask whether SAM salience can help explain choices in the common matrix game format.

First, note that the SAM training set does not contain images that resemble matrices of payoffs. Subjects in matrix game experiments also have a clear attentional goal, which is to look at numbers in a matrix to make a high-payoff choice. These goals are likely to create a complicated visual search to compute beliefs and implement decision rules, which is different than the rapid stimulus-driven attention that SAM is designed to predict.

In fact, many studies using Mouselab and eye-tracking stretching back three decades have shown patterns of search consistent with goal-directed perception for strategic thinking (Camerer et al. 1993; Costa-Gomes, Crawford, and Broseta 2001; Johnson et al. 2002; Arieli, Ben-Ami, and Rubinstein 2011; Brocas et al. 2014; Polonio, Di Guida, and Coricelli 2015; Devetag, Di Guida, and Polonio 2016). Furthermore, most of the behavioral studies about coordination and hider-seeker games have aimed at establishing general principles of focality or psychological prominence from strategic goals and set-theoretic properties of strategies (see Online Appendix C for a review). So it is already known that goal-directed allocation of attention is evident in choices from matrix payoff games. An unanswered new question is whether stimulus-driven SAM salience has any additional predictive power.

The possible influence of visual salience is tested here using data from Polonio, Di Guida, and Coricelli (2015). In their experiment, 56 people played 32 normal-form games with different strategic structures. Eye-tracking was used to record visual

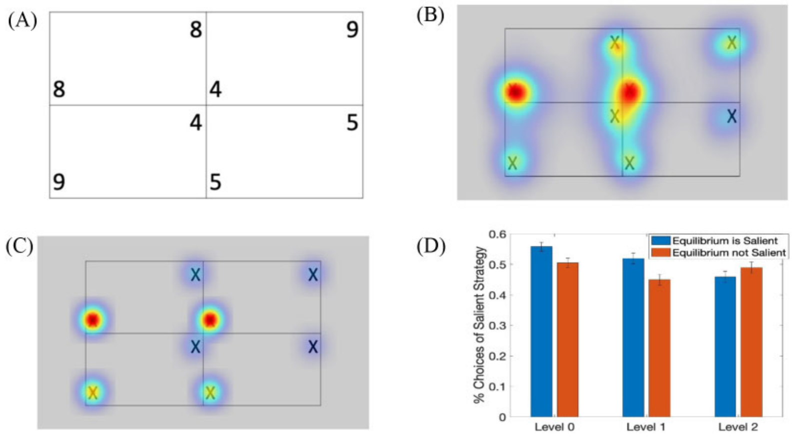

FIGURE XI

Saliency and Choices in Matrix Games

Panel A: One example (prisoner's dilemma) of the games used in the experiment. Panel B: The average SAM prediction of all games. Panel C: The ground truth gaze density map generated by gaze data. Panel D: Percentage of choices choosing the most SAM salient strategy grouped by levels (strategic thinking levels classified by gaze and behavior data by Polonio, Di Guida, and Coricelli 2015). (N: level 0 = 551, level 1 = 402, level 2 = 371) *Source:* Polonio, Di Guida, and Coricelli (2015).

attention. These data are especially useful because actual gaze maps can then be compared with SAM predictions and actual choices.[47]

Figure XI, Panel A shows one example of the type of matrix that subjects see on their computers (it's a prisoner's dilemma in structure). Row player payoffs are in the lower left of each matrix cell, and column player payoffs are in the upper right of each matrix cell.

Figure XI, Panel B is the average prediction from the SAM algorithm about where people look, averaged over all 32 games. There is a predicted bias toward looking more at the top row and the left column, as well as a row-player payoff bias (even for column players). Figure XI, Panel C is the average measured attention map calculated from eye-tracked gaze data over different types of games (filtering out gazes that are away from payoffs). The comparison between Figure XI, Panel B (algorithm) and Panel C (gaze data) suggest that the algorithm does predict the

47. See Online Appendix I for more details.

actual attention allocation during game play rather well. This visual impression is supported by conventional statistics used in visual science.[48] Much to our surprise, the actual human gaze data are also quite similar for row and column players (as is the SAM salience map, because it does not vary with player roles). This is surprising because higher-level strategic thinkers need to direct attention to different row and column payoffs.

The main question is whether there is a congruency effect (as in the fruits experiment 1): that is, does salience affect how often people choose the equilibrium strategy? We look at the 24 games that contain a unique equilibrium strategy for both players. We also use Polonio, Di Guida, and Coricelli (2015)'s classification of subjects into three groups based on strategic levels of thinking from 0 to 2, using the cognitive hierarchy model.[49]

Figure XI, Panel D shows that level-0 and -1 types do choose the salient strategy more often when it is an equilibrium, and level-2's go slightly in the opposite direction. This is consistent with the idea that level-0's are not using goal-directed attention, and level-1's and -2's use more goal-directed attention.[50]

Table IV tests whether the likelihood of choosing the equilibrium strategy depends on salience congruency. There is no general effect when all level types are pooled together (model 1). However, model 2 shows that there is a substantial effect of congruency, but only for level-0 players. (Note that level 2 is the omitted level category so that the congruency main effect estimates the level-2 effect, which is negative). However, the significance of the level-0 effect is only $p = .12$ when Bonferroni-corrected for multiple comparisons.

Thus, the evidence for an influence of stimulus-driven salience is suggestive but not statistically strong. It is also a surprise that the salience map and gaze data are so similar. Future

48. In the computer vision field, two validation scores, AUC and CC, are commonly used metrics to evaluate how closely salience algorithm predictions are correlated with actual human gazes. AUC: area under the receiver operating characteristics curve; CC: Pearson correlation (see Kummerer, Wallis, and Bethge 2018). Online Appendix Table I1 shows these statistics.

49. They classify players' types based on their gaze patterns on matrix games. The level-0's only focus on the payoff property itself (intra-cell). Level-1 players compare their own payoffs (own focused). Level-2 players also look at others' payoffs (distributed attention). This classification from gaze data was then correlated with predictions about what choices the three types should make.

50. Note that we did not preregister this prediction, so our conclusions should rightly be taken as exploratory and not a planned test of a hypothesis.

TABLE IV
The Effect of Salience-Equilibrium Congruency in Matrix Games

	Dependent variable: Whether the choice is an equilibrium strategy	
	(1)	(2)
Congruency (whether equilibrium strategy is salient)	0.008 (0.082)	− 0.208 (0.142)
Congruency*level-0		0.465** (0.189)
Congruency*level-1		0.073 (0.197)
Constant	0.240* (0.140)	0.348** (0.143)
Observations	1,323	1,323
Log likelihood	−910.061	−908.221
Akaike inf. crit.	1,834.122	1,834.443

Notes. The dependent variable is (0-1) whether the chosen strategy in a matrix game is the equilibrium strategy. (All games in the data set have a unique Nash strategy.) "Congruency" indicates whether the equilibrium option in that particular game is also more salient (which is the top row/left column). Covariates (coefficients not reported) are: game types (DSS, PD, DSO), role (row, column), levels (level-0, level-1, level-2). Standard errors are clustered at the individual level. * $p < .1$; ** $p < .05$; *** $p < .01$

experiments could explicitly manipulate salience (guided by SAM predictions) of particular payoffs to see if stronger effects can be created.

VII. Comparison with Other Salience and Attention Approaches

This section briefly reviews recent economic theories that have analyzed salience and attention and describes the relations of those theories to our approach.

VII.A. Salience Theory

Salience theory is a theory of salience that has been widely applied for the past 10 years in economics and finance and in other areas (Bordalo, Gennaioli, and Shleifer 2012b, 2013a, 2013b). It was the first economic theory to specify exactly how salience is generally derived from attributes, and affects choice, to make clear predictions testable from observable data. The goal of this section is to describe how salience is computed in that theory and compare it to stimulus-driven SAM algorithmic salience.

In salience theory, attribute values of choice objects that are relatively farther from a reference point (such as the average attribute value[51]) are judged to be more salient. We'll use the notation from analysis of multiattribute choice (Bordalo, Gennaioli, and Shleifer 2013b) to see how salience theory works. A choice k has attribute level a_k along a particular attribute dimension. The average level across the entire choice set is $\bar{a}$.

The salience function is defined by $\sigma(a_k, \bar{a})$. This function is assumed to obey two properties called ordering and diminishing sensitivity.

Ordering means that increasing the magnitude of the attribute level a_k by ϵ from $\bar{a}$, while decreasing the reference point in the opposite direction by ϵ', increases salience.[52] Kőszegi and Szeidl (2013) proposed a similar "focusing" model in which all values of an attribute are weighted more heavily when an attribute has more wide-ranging utilities (see Bordalo, Gennaioli, and Shleifer 2013b, 815–16 for comparison). Diminishing sensitivity means that increasing the level of both a_k and $\bar{a}$ by the same positive amount reduces the salience of a_k. Although ordering and diminishing sensitivity are enough for most of the applications to work, a more strict version further assumes homogeneity of degree zero (i.e., $\sigma(a_k, \bar{a}) = \sigma(\alpha a_k, \alpha\bar{a})$ for $\alpha > 0$). A simple salience function that satisfies all these properties is $\frac{|a_k - \bar{a}|}{|a_k| + |\bar{a}|}$. We now make two remarks about salience theory.

First, attributes such as product quality or endowment states do not have to be numbers to be judged as salient. They could be perfume aromas or restaurant noise levels. However, attributes are assumed to have subjective estimated values, so that salience

51. In some applications, it is plausible that an external reference point that is not part of a choice set influences salience. For example, the explanation of endowment effects works with goods that have two attributes, and the consideration set includes having nothing (0,0) (Bordalo, Gennaioli, and Shleifer 2012a). Including this null state makes the best quality of the initially endowed good salient, which creates a valuation that is inflated (compared with a no-salience benchmark). For example, in Thaler (1985), when people are asked about their willingness to pay for a beer on a hot day, most people will value hotel cans more than the cans from a normal corner shop, even though they are identical goods.

52. Formally, define a sign function by $\mu(a_k - \bar{a}) = 1$ iff $a_k - \bar{a} \geqslant 0$ and $\mu(a_k - \bar{a}) = -1$ iff $a_k - \bar{a} < 0$. Ordering is the property that $\sigma(a_k + \mu(a_k - \bar{a})\epsilon, \bar{a} - \mu(a_k - \bar{a})\epsilon') > \sigma(a_k, \bar{a})$, for $\epsilon, \epsilon' \geqslant 0$ and $\epsilon + \epsilon' > 0$.

can be computed and used to weight attributes in computing decision values. A salient thinker will overweight the salient attributes and underweight the unsalient ones.

Second, like our work, salience theory was clearly motivated by ideas and evidence in psychology and neuroscience. Ancestors of context-sensitivity and the ordering property are common in historical and modern psychology. (For example, we have repeatedly noted the importance of low-level contextual contrast in Itti, Koch, and Niebur 1998 and later algorithms.) Diminishing sensitivity is also a ubiquitous psychophysical (Weber-Fechner) principle of perception. William James's (1863) speculative list of things that engage "passive immediate sensorial attention" included "strange things" that can be translated as context-deviating attributes or objects. In modern neuroscience, salience is often defined as absolute magnitude (deviation from zero) and is known to be encoded in the brain (McCoy and Platt 2005; Armel, Beaumel, and Rangel 2008; Litt et al. 2011).

Recent perceptual judgment experiments (Kunar et al. 2017) illustrate one way that salience of extreme values affects judgment. Participants saw sequences of 12 two-digit numbers, presented rapidly ($<$ 100 milliseconds) one at a time. Judgments reflected more attention to the highest and lowest numbers in each stream (which are those with the highest Bordalo, Gennaioli, and Shleifer salience; see also Tsetsos, Chater, and Usher 2012).[53] Larger differences are also more salient when people are looking for one target object out of many (including "distractors"). The target is easier to find when it is more different than distractors on features—such as searching for an X in a group of O's rather than in a group of Y's. The target-distractor differences should be expressible as numbers similar to normalized values of $|a_k - \bar{a}|$ (Wolfe and Horowitz 2017, 2), as in salience theory, but we do not know of direct equivalences of this sort.[54]

53. Kunar et al. (2017) found that when people were instructed to report whether they saw a specific target number, they missed that number more often when it was preceded by the highest or lowest number in the sequence. This is consistent with the joint hypothesis that people were more attentive to the extreme numbers, and exhibit a typical "attentional blink" in which attention lapses a bit after the high attention paid to extreme numbers.

54. Wolfe and Horowitz (2017) compile a list of visual properties of features that robustly "guide" attention. In vision science jargon, a variable X guides attention if a target having property X increases the accuracy and speed of finding that target. Relative size and higher subjective value are two guiding variables in Wolfe and Horowitz (2017).

Because of this generality, salience theory has been used to explain or interpret phenomena and empirical evidence in finance, lottery choices (including drug trafficking), legal judgments, price-quality markets, and cross-game attention (Bordalo, Gennaioli, and Shleifer 2013a, 2015; Spitmaan, Chu, and Soltani 2019; Dertwinkel-Kalt and Köster 2020; Magliocca et al. 2019; Avoyan and Schotter 2020; Dertwinkel-Kalt and Köster 2020; Cosemans and Frehen 2021).

Salience theory and stimulus-driven salience (as defined and applied above) focus on different aspects of salience and their implications. In most applications, the two theories do not make competing predictions, without additional specialized assumptions. The experiment 1 fruit sets design is an example. SAM salience predicts visual salience of images, then investigates whether that special type of salience affects choices. In contrast, salience theory is about salience of valued attributes, regardless of how they are displayed or described, so it does not have a natural role for aspects of visual salience that are unrelated to attribute values.

Both theories are simplifications that have advantages and limits. Salience theory has the advantage of portability to many familiar microeconomic and social science applications. It benefits from the simplicity that comes from ignoring details of visual perception. Algorithmic SAM-type salience has the advantage of predicting rapid stimulus-driven visual attention for all possible images, but applying the theory to familiar domains such as price-quality competition is not straightforward (as noted in our discussion of explainable AI) and will be stimulus-constrained.

VII.B. Rational Inattention

Rational inattention (RI) models assume that people optimally trade off the benefits and costs of paying closer attention. In more technical terms, endogenously allocated attention creates a subjective perception of objective factors. More accurate subjective perception is more costly but also improves expected decision value.[55] These models are goal-directed because there is a clear goal—better perception is chosen to improve decision value.

RI models often start with a prior belief distribution μ over a set of states $\{\omega | \omega \in \Omega\}$. In our fruit experiment, each ω is a

55. For more detail, see Sims (2003, 2006); Caplin and Dean (2015); Caplin, Dean, and Leahy (2019); Caplin et al. (2020); Kőszegi and Matějka (2020); Mackowiak et al. (2020).

possible image. For each image, there is an optimal action $a \in \{L, R\}$ (left or right, depending on which has the higher induced value). Denote the optimal actions by $a^*(\omega)$. There is also a pair of numbers $S^L(\omega)$, $S^R(\omega)$ that are the predicted SAM saliencies in the L and R halves of an image ω.[56]

In RI, attention creates a set of latent "signals" $\gamma(\omega)$ from a mapping $\pi\colon \Omega \to \Delta(\Gamma)$ (Caplin and Dean 2015; Caplin, Dean, and Leahy 2019). In the fruit example, $\gamma(\omega)$ could be the subjective belief probability of image ω after all the learning processes. The "rationality" in RI comes from the assumption that the signal structure is chosen to maximize a gross decision value minus a cost of attention. The key term in the decision value is $\max_{a\in A}\sum_{\omega\in\Omega}\gamma(\omega)u(a, \omega)$. Because the saliencies $S^L(\omega)$, $S^R(\omega)$ do not enter the utility function $u(a, \omega)$ and do not provide information about the optimal action $a^*(\omega)$, an RI agent should ignore them.[57] However, the results from the fruit experiment show that stimulus-driven salience can interfere with goal-directed RI and moves decisions away from RI optimality.

VII.C. Dynamic Channeled Inattention and Bayesian Surprise

Some economic models seek to understand the dynamic effects of limited attention. This is different than our use of predicted salience to understand static choices.

Schwartzstein (2014) studies a problem of forecasting a binary variable y that depends on x and a subjectively encoded variable z. When z is expected to be important enough in forecasting y, with an expected value above a "busyness" threshold b, z is accurately encoded. Otherwise, z is ignored, and if z is ignored, no missing value is imputed.

Gagnon-Bartsch, Rabin, and Schwartzstein (2018) proposed a similar idea of channeled attention during learning in which people do not always recognize the results of their inattention. For example, a person who often forgets to take her medicines but does not have a strong prior belief that she might forget, does not

56. To be clear, the fruits experiment is not an ideal proper test of RI. To do so would require controlling the set Ω more carefully and assuming, measuring, or inducing a prior belief that salience and induced value are uncorrelated, which was not done.

57. It would be useful to figure out precisely how to integrate the effect of stimulus-driven salience into RI, to explain examples like the fruits experiment. Li (2020, 81) provides a saliency-sensitive state separation that can explain the saliency effect in simple choices.

notice or keep track of her forgetting. She won't pay for a reminder technology. They refer to these missed data as "statistical gorillas" (from the famous attention-blindness experiment of Simons and Chabris 1999). They derive dynamic conditions under which statistical gorillas will be noticed or not.

In dynamic image sequences, such as movies, one property of images that is known (from eye-tracking) to grab attention strongly is called Bayesian surprise. This concept begins with a prior belief over "models" in model space $\mathcal{M}$. Itti and Baldi (2009) used an example in which a person turns on her TV, not knowing what channel was last watched and will pop up first. $\mathcal{M}$ is the set of possible TV channels. $P(D|M)$ are the likelihoods of perceptual data D conditional on a model M (a TV channel). For example, if blonde women are more common on $M = \{\text{Fox News}\}$ than other channels, then $P(\text{blonde women}|\{\text{Fox News}\}) > P(\text{blonde women}|\mathcal{M})$.

"Surprise" for a given (D, M) combination is defined as $S(D, M) \equiv log \frac{P(M)}{P(M|D)}$.[58] A person might be greatly surprised, for example, by seeing a blonde woman on the sports channel ESPN if $P(ESPN) \gg P(ESPN|\text{blonde woman})$. The ratio $\frac{P(ESPN)}{P(ESPN|\text{blonde woman})}$ and its logarithm will then be much greater than one, measuring how surprising that data-model combination is. Experienced surprise from data D, averaged over model posteriors, is a measure of overall experienced surprise from perceptual data D:

$$\sum_M P(M|D)S(D, M).$$

Note that Bayesian surprise does not fit into the stimulus-driven versus goal-directed dichotomy. It depends on a perceiver's prior beliefs, so it is not purely stimulus-driven. But surprise detection is also highly general and is therefore not typically

58. There is a loose relation between the ratio $\frac{P(M)}{P(M|D)}$ and a concept of representativeness as relative likelihood $\frac{P(D|M_1)}{P(D|M_2)}$ (see Tenenbaum and Griffiths 2001 and Bordalo et al. 2016 for stereotypes, where D is a social type and models M are groups). The surprise ratio for a particular M is a measure of how unrepresentative or anomalous D is, and the summation adds up the total degree of unrepresentativeness of D for all models M.

considered a perceptual goal like, say, searching for a familiar face in a crowd or for a high resale value fruit.[59]

Bayesian surprise is not used in the types of experiments in this study because the presented images were not deliberately linked in a dynamic sequence (as in a movie). However, in typical experiments, prior perceptual beliefs are induced by short exposures to each of a large number of images, so that what is surprising in a subsequent image (relative to those priors) can be quantified. This could easily be done in the fruits experiment. For example, if many images in a row included no apples, then in a new image with an apple, the apple would be Bayesian-surprising and is predicted to be salient and attract attention.

The Bayesian surprise model is well supported experimentally (Itti and Baldi 2009) and has the advantage that some analytical results are available for the class of conjugate priors (Baldi and Itti 2010). Potential economic applications include a sequential visual presentation of price changes in a time series, or testing for salience from a new advertising campaign, product design, or logo change.[60]

VII.D. *Relative Attention m(x)*

It is useful to have a simple measure of inattention, as revealed by choices, to compare across domains. A good one is summarized by Gabaix (2019). Define both rational and behavioral actions, as a function of a perceived normative variable x (such as a price), by $a^r(x) = argmax_a u(a, x)$ (rational) and $a^b(x) = a^r(mx)$ (behavioral). The behavioral model is assumed to maximize but underperceives or underweights the true variable value x,

59. Pierre Baldi said in a personal communication by email (6/4/2021) that "Bayesian surprise is agnostic with respect to any bottom-up or top-down considerations."

60. Note that there is an apparent opposition between ignoring statistical gorillas in Gagnon-Bartsch, Rabin, and Schwartzstein (2018) and Bayesian surprise. A gorilla on a basketball court is typically very high in Bayesian surprise and hence predicted to be quite salient; then why don't people notice the gorilla? The answer is that scarce attention is focused on one mentally taxing goal—counting basketball passes (the instructed goal in the seminal study)—so that a Bayesian-surprising object is ignored. Magic tricks work the same general way: Skillful "misdirection" draws attention away from the sneaky sleight of hand (Macknik et al. 2008; Wiseman and Nakano 2016). In economic settings, Bayesian surprise and other goal-directed attention will be productive substitutes, a hypothesis that can be tested by phenomena like timing and reaction to unusual corporate earnings announcements (e.g., DeHaan, Shevlin, and Thornock 2015).

shrinking it toward zero to a degree measured by a parameter $m < 1$. (A canonical example is paying too little attention to a hidden component of price, such as taxes, where $m < 1$ measures the degree of tax underweighting.) Gabaix (2019) shows that $m(x)$ can be recovered from the ratio of marginal effects of the x variable on actions in different attention treatments, $\frac{a_x^b}{a_x^r}$. His table 1 summarizes numerical estimates from several field experiments and datasets.

A version of $m(x)$ can also be computed from the fruit experiment data based on an ad hoc assumption. Suppose choice under time pressure is designated as the "behavioral" condition and choice with unlimited time is designated as the "rational" condition. The intuition is that in the behavioral condition, stimulus-driven salience is not fully inhibited (even though it is irrelevant), which reduces the influence of goal-directed attention ($m < 1$) to the induced value x variable. From Table I the marginal effect of the normative variable (the induced value difference) on choice accuracy is $\hat{a}_x^b = 0.795$. The value of $\hat{a}_x^r$ can be computed from the same regression as in Table I, using data from the unlimited time treatment. That value turns out to be $\hat{a}_x^r = 2.249$. The ratio of the behavioral and rational coefficient estimates is therefore $\frac{\hat{a}_x^b}{\hat{a}_x^r} = \frac{0.795}{2.249}$, which is 0.35. This figure is close to the mean $m(x) = 0.44$ reported in Gabaix (2019) table 1. This numerical exercise shows how the effect of stimulus-driven salience as a behavioral condition can be compared numerically to other kinds of limited attention.

VIII. Discussion and Conclusion

Our study leads to two new conclusions:

i. Stimulus-driven salience can be predicted by an underlying neuro-computational theory (SAM) of which features of an image or information display most people look at first. SAM-estimated salience has a small but significant effect in visualized binary set choices (fruit sets) and in matrix games. These effects are not always strong because in both cases stimulus-driven attention competes with goal-directed attention in a way that SAM the algorithm does not attempt to predict.
ii. In the main set of experiments with location matching games, salience is a good predictor of which location

people choose, and how often their choices match ($r = -0.45$). In hider-seeker games, a salience-influenced cognitive hierarchy model (and a similar level-k model) can account for the small but robust seeker's advantage in hider-seeker games. Parameters fit to hider-seeker data can also "portably" predict the salience-choice relation in matching games, even though the hider-seeker game is strictly competitive and matching is cooperative.

VIII.A. *Where Else in Economics Could Salience Be Useful?*

Before proceeding to further visual salience speculation, note that vision is only one of five senses; other sensory systems have salience structures, too. Auditory (sound) attention is also driven by both goal-directed and stimulus-driven processes. One can attend to an important conversation to achieve a social goal while tuning out background noises at a party. But the stimulus-driven system will hijack attention if a champagne glass shatters with a loud crash.[61] Research parallel to ours could explore auditory salience in domains like advertising, business communication, security analyst earnings calls, open-outcry auctions, negotiations, and so on.

This last section speculates how an empirical understanding of stimulus-driven salience might improve other economic studies.

- *Behavioral IO:* The fruits experiment is a paradigm that invites thinking and future exploration about the supply-side response to consumer psychology, a subfield called behavioral industrial organization (Heidhues and Kőszegi 2018, and others). A central concept in behavioral IO is whether product attributes are "shrouded" (Gabaix and Laibson 2006)—that is, deliberately hidden by sellers. Measuring whether attributes are low in stimulus-driven salience is one scientific measure of shrouding, which is perhaps useful for consumer policy regulators. By understanding stimulus-driven salience, a retailer could create a product display with the goal of maximizing profit

61. A general example is the stimulus-driven salience of human screams (which have an unpleasant power spectrum quality called "roughness"). Screams are rated more quickly as fear-inducing, are more accurately localized, and activate the amygdala and primary auditory cortex more strongly (Arnal et al. 2015). Kaya and Elhilali (2014) proposed an auditory salience map based on five features (envelope, harmonicity, spectrogram, bandwidth, and modulation) and tested it.

margin. High-margin items would be displayed to maximize their stimulus-driven salience. An open and interesting question is whether consumers can recognize and ignore such supply-side salience manipulations.

- *Tax and price salience in consumer markets:* Price and value components that are presented to sensory systems, such as explicit price tags that the eye can see, seem to receive more decision weight than equivalent components that need to be imagined and computed. This effect was first shown for unit-cost price tags by Russo (1977) and has been shown carefully in many recent studies (Ott and Andrus 2000; Hossain and Morgan 2006; Min Kim and Kachersky 2006; Finkelstein 2009; Taubinsky and Rees-Jones 2017). In principle, SAM could be applied to visual images of store price tags or e-commerce websites, as was done in the fruit-valuation study 1, to guess the visual salience of explicit and hidden prices. These measures could be compared with salience as measured from behavior in these papers and as summarized in the $m(x)$ measure in Gabaix's (2019) table 1.
- *Nudges and design:* Nudges are changes in design and choice architecture, which do not drastically change information content or incentives, but can make information processing simpler and improve decisions.[62] Many nudge experiments have been done and are ongoing. But their effects are often unpredictable (DellaVigna and Linos 2022; Milkman et al. 2021). Predictions about what nudges are visually salient might help us understand what has worked and create better designs. If a financial regulator is trying to design a form to nudge goal-directed attention toward particular information, for example, their design will probably work better if the targeted information also has stimulus-driven salience (e.g., Hilchey, Osborne, and Soman 2021).
- *Beliefs*: Besides influencing choices, visual salience can influence what information is processed and what beliefs result.[63] Padilla, Ruginski, and Creem-Regehr (2017)

62. See Goldin (2015); Thaler and Sunstein (2009); and Luo, Soman, and Zhao (2021).

63. See Padilla et al. (2018); Itti, Koch, and Niebur (1998); and Mackowiak et al. (2020).

showed a striking example of an effect of stimulus-driven salience on beliefs about hurricanes. The National Hurricane Center currently shows potential geospatial paths with a cone of uncertainty, a 2-D confidence interval forecasting a range of areas that a hurricane might conceivably reach. The cone becomes wider, spreading out geographically, for forecasts projecting more days ahead (which are typically more uncertain). An alternative visualization is an ensemble plot, which shows many distinct possible individual paths and does not draw a cone around them (they are called "spaghetti plots"). Padilla, Ruginski, and Creem-Regehr (2017) apply the Itti, Koch, and Niebur (1998) algorithm (a precursor to SAM) to these two different visualizations. The algorithm predicts that cone plots will focus attention on the center and on the furthest boundaries of the cone, where the cone is widest. This perception biases actual human judgments of whether the hurricane will grow in storm size and intensity (e.g., wind speed) in the future. These subjective beliefs reflect a cognitive mistake: People think the growing size of the cone predicts that the size of the storms and their intensity will grow. The ensemble plot has different predicted salience and a different effect on beliefs. Predicted salience is highest at the location where different paths are clustered before they diverge into different paths. Attention is widely dispersed over the ending points of the different trajectories (rather than concentrated at the cone plot boundary). As a result, judgments about future storm size and intensity are not infected by a size bias (as they are from cone plots). Thus, the cone plot leads to mistaken beliefs and the ensemble plot does not. The salience algorithm accurately predicted the direction of that effect. An economic example of a similar kind is the visualization of regression discontinuity effects. Korting et al. (2021) show that axis-scaling, x-axis bin width, and spacing all influence the perceptions people have about causal effects when shown different graphs based on the same data. SAM or other salience algorithms could be applied to these data, to learn more about how stimulus-driven processes affect what scientific consumers think a graph is telling them.

California Institute Of Technology, United States

Supplementary Material

An Online Appendix for this article can be found online at the Harvard Dataverse, https://doi.org/10.7910/DVN/9LCYKG.

Data Availability

Data and code replicating tables and figures in this article can be found in Li and Camerer (2022) in the Harvard Dataverse, https://doi.org/10.7910/DVN/9LCYKG.

References

Arieli, Amos, Yaniv Ben-Ami, and Ariel Rubinstein, "Tracking Decision Makers under Uncertainty," *American Economic Journal: Microeconomics*, 3 (2011), 68–76.

Armel, Carrie, Aurelie Beaumel, and Antonio Rangel, "Biasing Simple Choices by Manipulating Relative Visual Attention," *Judgment and Decision Making*, 3 (2008), 396–403.

Arnal, Luc H., Adeen Flinker, Andreas Kleinschmidt, Anne-Lise Giraud, and David Poeppel, "Human Screams Occupy a Privileged Niche in the Communication Soundscape," *Current Biology*, 25 (2015), 2051–2056.

Arrieta, Alejandro Barredo, Natalia Díaz-Rodríguez, Javier Del Ser, Adrien Bennetot, Siham Tabik, Alberto Barbado, Salvador García, Sergio Gil-López, Daniel Molina, and Richard Benjamins et al., "Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opportunities and Challenges toward Responsible AI," *Information Fusion*, 58 (2020), 82–115.

Aumann, Robert J., "Subjectivity and Correlation in Randomized Strategies," *Journal of Mathematical Economics*, 1 (1974), 67–96.

Avoyan, Ala, and Andrew Schotter, "Attention in Games: An Experimental Study," *European Economic Review*, 124 (2020), 103410.

Awh, Edward, Artem V. Belopolsky, and Jan Theeuwes, "Top-Down versus Bottom-Up Attentional Control: A Failed Theoretical Dichotomy," *Trends in Cognitive Sciences*, 16 (2012), 437–443.

Bacharach, Michael, "Variable Universe Games," in *Frontiers of Game Theory*, Ken Binmore, Alan Kirman, and Piero Tani, eds. (Cambridge, MA: MIT Press, 1993), 255–276.

Bacharach, Michael, and Michele Bernasconi, "The Variable Frame Theory of Focal Points: An Experimental Study," *Games and Economic Behavior*, 19 (1997), 1–45.

Baldi, Pierre, and Laurent Itti, "Of Bits and Wows: A Bayesian Theory of Surprise with Applications to Attention," *Neural Networks*, 23 (2010), 649–666.

Baluch, Farhan, and Laurent Itti, "Mechanisms of Top-Down Attention," *Trends in Neurosciences*, 34 (2011), 210–224.

Belle, Vaishak, and Ioannis Papantonis, "Principles and Practice of Explainable Machine Learning," arXiv preprint arXiv:2009.11698, 2020.

Bordalo, Pedro, Katherine Coffman, Nicola Gennaioli, and Andrei Shleifer, "Stereotypes," *Quarterly Journal of Economics*, 131 (2016), 1753–1794.

Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer, "Salience in Experimental Tests of the Endowment Effect," *American Economic Review*, 102 (2012a), 47–52.

———, “Salience Theory of Choice Under Risk,” *Quarterly Journal of Economics*, 127 (2012b), 1243–1285.
———, “Salience and Asset Prices,” *American Economic Review*, 103 (2013a), 623–628.
———, “Salience and Consumer Choice,” *Journal of Political Economy*, 121 (2013b), 803–843.
———, “Salience Theory of Judicial Decisions,” *Journal of Legal Studies*, 44 (2015), S7–S33.
Bornstein, Robert F., “Exposure and Affect: Overview and Meta-Analysis of Research, 1968–1987,” *Psychological Bulletin*, 106 (1989), 265–289.
Brocas, Isabelle, Juan D. Carrillo, Stephanie W. Wang, and Colin F. Camerer, “Imperfect Choice or Imperfect Attention? Understanding Strategic Thinking in Private Information Games,” *Review of Economic Studies*, 81 (2014), 944–970.
Camerer, Colin F., Teck-Hua Ho, and Juin-Kuan Chong, “A Cognitive Hierarchy Model of Games,” *Quarterly Journal of Economics*, 119 (2004), 861–898.
Camerer, Colin F., Eric Johnson, Talia Rymon, and Sankar Sen, “Cognition and Framing in Sequential Bargaining for Gains and Losses,” *Frontiers of Game Theory*, 104 (1993), 27–47.
Caplin, Andrew, Dániel Csaba, John Leahy, and Oded Nov, “Rational Inattention, Competitive Supply, and Psychometrics,” *Quarterly Journal of Economics*, 135 (2020), 1681–1724.
Caplin, Andrew, and Mark Dean, “Revealed Preference, Rational Inattention, and Costly Information Acquisition,” *American Economic Review*, 105 (2015), 2183–2203.
Caplin, Andrew, Mark Dean, and John Leahy, “Rational Inattention, Optimal Consideration Sets, and Stochastic Choice,” *Review of Economic Studies*, 86 (2019), 1061–1094.
Cerf, Moran, Jonathan Harel, Wolfgang Einhäuser, and Christof Koch, “Predicting Human Gaze Using Low-Level Saliency Combined with Face Detection,” in *Advances in Neural Information Processing Systems,* 20 (Cambridge, MA: MIT Press, 2007), 241–248.
Chen, Cheng, Xilin Zhang, Yizhou Wang, Tiangang Zhou, and Fang Fang, “Neural Activities in V1 Create the Bottom-Up Saliency Map of Natural Scenes,” *Experimental Brain Research*, 234 (2016), 1769–1780.
Chun, Marvin M., Julie D. Golomb, and Nicholas B. Turk-Browne, “A Taxonomy of External and Internal Attention,” *Annual Review of Psychology*, 62 (2011), 73–101.
Cornia, Marcella, Lorenzo Baraldi, Giuseppe Serra, and Rita Cucchiara, “Predicting Human Eye Fixations via an LSTM-Based Saliency Attentive Model,” *IEEE Transactions on Image Processing*, 27 (2018), 5142–5154.
Cosemans, Mathijs, and Rik Frehen, “Salience Theory and Stock Prices: Empirical Evidence,” *Journal of Financial Economics*, 140 (2021), 460–483.
Costa-Gomes, Miguel A., and Vincent P. Crawford, “Cognition and Behavior in Two-Person Guessing Games: An Experimental Study,” *American Economic Review*, 96 (2006), 1737–1768.
Costa-Gomes, Miguel, Vincent P. Crawford, and Bruno Broseta, “Cognition and Behavior in Normal-Form Games: An Experimental Study,” *Econometrica*, 69 (2001), 1193–1235.
Crawford, Vincent P., “A Comment on ‘How Portable Is Level-0 Behavior? A Test of Level-k Theory in Games with Non-neutral Frames’ by Heap, Rojo-Arjona, and Sugden,” Manuscript, 2014.
Crawford, Vincent P., Miguel A. Costa-Gomes, and Nagore Iriberri, “Structural Models of Nonequilibrium Strategic Thinking: Theory, Evidence, and Applications,” *Journal of Economic Literature*, 51 (2013), 5–62.
Crawford, Vincent P., and Nagore Iriberri, “Fatal Attraction: Salience, Naivete, and Sophistication in Experimental ‘Hide-and-Seek’ Games,” *American Economic Review*, 97 (2007a), 1731–1750.

———, "Level-k Auctions: Can a Nonequilibrium Model of Strategic Thinking Explain the Winner's Curse and Overbidding in Private-Value Auctions?," *Econometrica*, 75 (2007b), 1721–1770.

Cunningham, Thomas, "Biases and Implicit Knowledge," Paris School of Economics Working Paper, 2013.

DeHaan, Ed, Terry Shevlin, and Jacob Thornock, "Market (in) Attention and the Strategic Scheduling and Timing of Earnings Announcements," *Journal of Accounting and Economics*, 60 (2015), 36–55.

DellaVigna, Stefano, and Elizabeth Linos, "RCTs to Scale: Comprehensive Evidence from Two Nudge Units," *Econometrica*, 90 (2022), 81–116.

Dertwinkel-Kalt, Markus, Katrin Köhler, Mirjam R. J. Lange, and Tobias Wenzel, "Demand Shifts due to Salience Effects: Experimental Evidence," *Journal of the European Economic Association*, 15 (2017), 626–653.

Dertwinkel-Kalt, Markus, and Mats Köster, "Salience and Skewness Preferences," *Journal of the European Economic Association*, 18 (2020), 2057–2107.

Devetag, Giovanna, Sibilla Di Guida, and Luca Polonio, "An Eye-Tracking Study of Feature-Based Choice in One-Shot Games," *Experimental Economics*, 19 (2016), 177–201.

Fan, Fenglei, Jinjun Xiong, and Ge Wang, "On Interpretability of Artificial Neural Networks," arXiv preprint arXiv:2001.02522, 2020.

Finkelstein, Amy, "E-ztax: Tax Salience and Tax Rates," *Quarterly Journal of Economics*, 124 (2009), 969–1010.

Fudenberg, Drew, and Annie Liang, "Predicting and Understanding Initial Play," *American Economic Review*, 109 (2019), 4112–4141.

Fudenberg, Drew, Philipp Strack, and Tomasz Strzalecki, "Speed, Accuracy, and the Optimal Timing of Choices," *American Economic Review*, 108 (2018), 3651–3684.

Gabaix, Xavier, "Behavioral Inattention," in *Handbook of Behavioral Economics: Foundations and Applications 2*, B. Douglas Bernheim, Stefano DellaVigna, and David Laibson, eds. (Amsterdam: North-Holland, 2019), 261–343.

Gabaix, Xavier, and David Laibson, "Shrouded Attributes, Consumer Myopia, and Information Suppression in Competitive Markets," *Quarterly Journal of Economics*, 121 (2006), 505–540.

Gagnon-Bartsch, Tristan, Matthew Rabin, and Joshua Schwartzstein, "Channeled Attention and Stable Errors," Harvard Business School Working Paper 18–108, 2018.

Goldin, Jacob, "Which Way to Nudge: Uncovering Preferences in the Behavioral Age," *Yale Law Journal*, 125 (2015), 226.

Haji-Abolhassani, Amin, and James J. Clark, "An Inverse Yarbus Process: Predicting Observers' Task from Eye Movement Patterns," *Vision Research*, 103 (2014), 127–142.

Hargreaves Heap, Shaun, David Rojo Arjona, and Robert Sugden, "How Portable Is Level-0 Behavior? A Test of Level-k Theory in Games with Non–Neutral Frames," *Econometrica*, 82 (2014), 1133–1151.

Hartford, Jason S., James R. Wright, and Kevin Leyton-Brown, "Deep Learning for Predicting Human Strategic Behavior," in *Advances in Neural Information Processing Systems* (Cambridge, MA: MIT Press, 2016), 2424–2432.

He, Sen, Hamed R. Tavakoli, Ali Borji, Yang Mi, and Nicolas Pugeault, "Understanding and Visualizing Deep Visual Saliency Models," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2019), 10206–10215.

Heidhues, Paul, and Botond Kőszegi, "Behavioral Industrial Organization," *Handbook of Behavioral Economics: Applications and Foundations 1*, 1 (2018), 517–612.

Heinrich, Timo, and Irenaeus Wolff, "Strategic Reasoning in Hide-and-Seek Games: A Note," Working paper #74, Thurgau Institute of Economics and Department of Economics at the University of Konstanz, 2012.

Hilchey, Matthew D., Matthew Osborne, and Dilip Soman, "Does the Visual Salience of Credit Card Features Affect Choice?", *Behavioural Public Policy* (2021), 1–18.
Hinton, Geoffrey, Oriol Vinyals, and Jeff Dean, "Distilling the Knowledge in a Neural Network," arXiv preprint arXiv:1503.02531, 2015.
Ho, Teck-Hua, Colin Camerer, and Keith Weigelt, "Iterated Dominance and Iterated Best Response in Experimental 'p-beauty Contests'," *American Economic Review*, 88 (1998), 947–969.
Hossain, Tanjim, and John Morgan, "... Plus Shipping and Handling: Revenue (Non) Equivalence in Field Experiments on Ebay," *Advances in Economic Analysis & Policy*, 6 (2006), 1429.
Igami, Mitsuru, "Artificial Intelligence as Structural Estimation: Deep Blue, Bonanza, and AlphaGo," *Econometrics Journal*, 23 (2020), S1–S24.
Itti, Laurent, and Pierre Baldi, "Bayesian Surprise Attracts Human Attention," *Vision Research*, 49 (2009), 1295–1306.
Itti, Laurent, Christof Koch, and Ernst Niebur, "A Model of Saliency-Based Visual Attention for Rapid Scene Analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20 (1998), 1254–1259.
James, William, *The Principles of Psychology* (New York: Henry Holt, 1863).
Johnson, Eric J., Colin Camerer, Sankar Sen, and Talia Rymon, "Detecting Failures of Backward Induction: Monitoring Information Search in Sequential Bargaining," *Journal of Economic Theory*, 104 (2002), 16–47.
Judd, Tilke, Frédo Durand, and Antonio Torralba, "A Benchmark of Computational Models of Saliency to Predict Human Fixations," MIT-CSAIL-TR-2012-001 technical report, 2012.
Judd, Tilke, Krista Ehinger, Frédo Durand, and Antonio Torralba, "Learning to Predict Where Humans Look," in *2009 IEEE 12th International Conference on Computer Vision* (IEEE, 2009) 2106–2113.
Kaya, Emine Merve, and Mounya Elhilali, "Investigating Bottom-Up Auditory Attention," *Frontiers in Human Neuroscience*, 8 (2014), 327.
Kneeland, Terri, "Identifying Higher-Order Rationality," *Econometrica*, 83 (2015), 2065–2079.
Korting, Christina, Carl Lieberman, Jordan Matsudaira, Zhuan Pei, and Yi Shen, "Visual Inference and Graphical Representation in Regression Discontinuity Designs," IZA Discussion Paper No. 14923, 2021.
Kőszegi, Botond, and Filip Matějka, "Choice Simplification: A Theory of Mental Budgeting and Naive Diversification," *Quarterly Journal of Economics*, 135 (2020), 1153–1207.
Kőszegi, Botond, and Adam Szeidl, "A Model of Focusing in Economic Choice," *Quarterly Journal of Economics*, 128 (2013), 53–104.
Krasovskaya, Sofia, and W. Joseph MacInnes, "Salience Models: A Computational Cognitive Neuroscience Review," *Vision*, 3 (2019), 56.
Kummerer, Matthias, Thomas S. A. Wallis, and Matthias Bethge, "Saliency Benchmarking Made Easy: Separating Models, Maps and Metrics," in *Proceedings of the European Conference on Computer Vision (ECCV),* (2018), 770–787.
Kunar, Melina A., Derrick G. Watson, Konstantinos Tsetsos, and Nick Chater, "The Influence of Attention on Value Integration," *Attention, Perception, & Psychophysics*, 79 (2017), 1615–1627.
Lewis, David, *Convention: A Philosophical Study* (Hoboken, NJ: John Wiley & Sons, 2008).
Li, Xiaomin, "Attention, Strategy, and the Human Mind," Ph.D. diss., Caltech, 2020.
Li, Xiaomin, and Colin F. Camerer, "Replication Data for: 'Predictable Effects of Visual Salience in Experimental Decisions and Games'," (2022), Harvard Dataverse, https://doi.org/10.7910/DVN/9LCYKG.
Lipton, Zachary C., "The Mythos of Model Interpretability," *Queue*, 16 (2018), 31–57.
Litt, Ab, Hilke Plassmann, Baba Shiv, and Antonio Rangel, "Dissociating Valuation and Saliency Signals During Decision-Making," *Cerebral Cortex*, 21 (2011), 95–102.

Luo, Yu, Dilip Soman, and Jiaying Zhao, "A Meta-Analytic Cognitive Framework of Nudge and Sludge." (2021).
Macknik, Stephen L., Mac King, James Randi, Apollo Robbins, John Thompson, and Susana Martinez-Conde et al., "Attention and Awareness in Stage Magic: Turning Tricks into Research," *Nature Reviews Neuroscience*, 9 (2008), 871–879.
Mackowiak, Bartosz, Filip Matejka, and Mirko Wiederholt et al., "Rational Inattention: A Review," CEPR Discussion Papers, (2020).
Magliocca, Nicholas R., Kendra McSweeney, Steven E. Sesnie, Elizabeth Tellman, Jennifer A. Devine, Erik A. Nielsen, Zoe Pearson, and David J. Wrathall, "Modeling Cocaine Traffickers and Counterdrug Interdiction Forces as a Complex Adaptive System," *Proceedings of the National Academy of Sciences*, 116 (2019), 7784–7792.
McCoy, Allison N., and Michael L. Platt, "Risk-Sensitive Neurons in Macaque Posterior Cingulate Cortex," *Nature Neuroscience*, 8 (2005), 1220–1227.
Mehta, Judith, Chris Starmer, and Robert Sugden, "Focal Points in Pure Coordination Games: An Experimental Investigation," *Theory and Decision*, 36 (1994), 163–185.
Milkman, K.L., D. Gromet, H. Ho, J. Kay, T. Lee, P. Pandiloski, Y. Park, and A. Rai et al., "Megastudies Improve the Impact of Applied Behavioral Science," *Nature*, 600 (2021), 478–483.
Milosavljevic, Milica, Vidhya Navalpakkam, Christof Koch, and Antonio Rangel, "Relative Visual Saliency Differences Induce Sizable Bias in Consumer Choice," *Journal of Consumer Psychology*, 22 (2012), 67–74.
Min, Kim Hyeong, and Luke Kachersky, "Dimensions of Price Salience: A Conceptual Framework for Perceptions of Multi-Dimensional Prices," *Journal of Product & Brand Management*, 15 (2006), 139–147.
Mormann, Milica, and Russo J. Edward, "Does Attention Increase the Value of Choice Alternatives?" *Trends in Cognitive Sciences*, 25 (2021), 305–315.
Nagel, Rosemarie, "Unraveling in Guessing Games: An Experimental Study," *American Economic Review*, 85 (1995), 1313–1326.
Ott, Richard L., and David M. Andrus, "The Effect of Personal Property Taxes on Consumer Vehicle-Purchasing Decisions: A Partitioned Price/Mental Accounting Theory Analysis," *Public Finance Review*, 28 (2000), 134–152.
Pachur, Thorsten, Michael Schulte-Mecklenbeck, Ryan O. Murphy, and Ralph Hertwig, "Prospect Theory Reflects Selective Allocation of Attention," *Journal of Experimental Psychology: General*, 147 (2018), 147–169.
Padilla, Lace M., Sarah H. Creem-Regehr, Mary Hegarty, and Jeanine K. Stefanucci, "Decision Making with Visualizations: A Cognitive Framework across Disciplines," *Cognitive Research: Principles and Implications*, 3 (2018), 29.
Padilla, Lace M., Ian T. Ruginski, and Sarah H. Creem-Regehr, "Effects of Ensemble and Summary Displays on Interpretations of Geospatial Uncertainty Data," *Cognitive Research: Principles and Implications*, 2 (2017), 1–16.
Polonio, Luca, Sibilla Di Guida, and Giorgio Coricelli, "Strategic Sophistication and Attention in Games: An Eye-Tracking Study," *Games and Economic Behavior*, 94 (2015), 80–96.
Ras, Gabriëlle, Marcel van Gerven, and Pim Haselager, "Explanation Methods in Deep Learning: Users, Values, Concerns and Challenges," in *Explainable and Interpretable Models in Computer Vision and Machine Learning,* Hugo Jair Escalante, Sergio Escalera, Isabelle Guyon, Xavier Baró, Yagmur Güçlütürk, Umit Güçlü, and Marcel van Gerven, eds. (Cham, Switzerland: Springer, 2018), 19–36.
Ratcliff, Roger, "A Theory of Memory Retrieval," *Psychological Review*, 85 (1978), 59.
Ratcliff, Roger, Philip L. Smith, Scott D. Brown, and Gail McKoon, "Diffusion Decision Model: Current Issues and History," *Trends in Cognitive Sciences*, 20 (2016), 260–281.

Riche, Nicolas, Matthieu Duvinage, Matei Mancas, Bernard Gosselin, and Thierry Dutoit, "Saliency and Human Fixations: State-of-the-Art and Study of Comparison Metrics," in 2013 *IEEE International Conference on Computer Vision* (IEEE, 2013) 1153–1160.
Rihn, Alicia, Xuan Wei, and Hayk Khachatryan, "Text vs. Logo: Does Eco-Label Format Influence Consumers' Visual Attention and Willingness-to-Pay for Fruit Plants? An Experimental Auction Approach," *Journal of Behavioral and Experimental Economics*, 82 (2019), 101452.
Rogers, Brian W., Thomas R. Palfrey, and Colin F. Camerer, "Heterogeneous Quantal Response Equilibrium and Cognitive Hierarchies," *Journal of Economic Theory*, 144 (2009), 1440–1467.
Russo, J. Edward, "The Value of Unit Price Information," *Journal of Marketing Research*, 14 (1977), 193–201.
Schelling, Thomas C., *The Strategy of Conflict* (Cambridge, MA: Harvard University Press, 1960).
Schwartzstein, Joshua, "Selective Attention and Learning," *Journal of the European Economic Association*, 12 (2014), 1423–1452.
Shimojo, Shinsuke, Claudiu Simion, Eiko Shimojo, and Christian Scheier, "Gaze Bias Both Reflects and Influences Preference," *Nature Neuroscience*, 6 (2003), 1317–1322.
Simons, Daniel J., and Christopher F. Chabris, "Gorillas in Our Midst: Sustained Inattentional Blindness for Dynamic Events," *Perception*, 28 (1999), 1059–1074.
Sims, Christopher A., "Implications of Rational Inattention," *Journal of Monetary Economics*, 50 (2003), 665–690.
———, "Rational Inattention: Beyond the Linear-Quadratic Case," *American Economic Review*, 96 (2006), 158–163.
Smith, Vernon L., "Experimental Economics: Induced Value Theory," *American Economic Review*, 66 (1976), 274–279.
Spitmaan, Mehran, Emily Chu, and Alireza Soltani, "Salience-Driven Value Construction for Adaptive Choice under Risk," *Journal of Neuroscience*, 39 (2019), 5195–5209.
Stahl, Dale O. II, and Paul W. Wilson, "Experimental Evidence on Players' Models of Other Players," *Journal of Economic Behavior & Organization*, 25 (1994), 309–327.
Steinbeck, John, *The Pearl* (London: Penguin, 2011).
Taubinsky, Dmitry, and Alex Rees-Jones, "Attention Variation and Welfare: Theory and Evidence from A Tax Salience Experiment," *Review of Economic Studies*, 85 (2017), 2462–2496.
Tenenbaum, Joshua B., and Thomas L. Griffiths, "The Rational Basis of Representativeness," in *Proceedings of the 23rd Annual Conference of the Cognitive Science Society* (2001), 1036–1041.
Thaler, Richard, "Mental Accounting and Consumer Choice," *Marketing Science*, 4 (1985), 199–214.
Thaler, Richard H., and Cass R. Sunstein, *Nudge: Improving Decisions about Health, Wealth, and Happiness* (London: Penguin, 2009).
Towal, R. Blythe, Milica Mormann, and Christof Koch, "Simultaneous Modeling of Visual Saliency and Value Computation Improves Predictions of Economic Choice," *Proceedings of the National Academy of Sciences*, 110 (2013), E3858–E3867.
Tsetsos, Konstantinos, Nick Chater, and Marius Usher, "Salience Driven Value Integration Explains Decision Biases and Preference Reversal," *Proceedings of the National Academy of Sciences*, 109 (2012), 9659–9664.
Veale, Richard, Ziad M. Hafed, and Masatoshi Yoshida, "How Is Visual Salience Computed in the Brain? Insights from Behaviour, Neurobiology and Modelling," *Philosophical Transactions of the Royal Society B: Biological Sciences*, 372 (2017), 20160113.
Vig, Eleonora, Michael Dorr, and David Cox, "Large-Scale Optimization of Hierarchical Features for Saliency Prediction in Natural Images," in *Proceedings*

of the IEEE Conference on Computer Vision and Pattern Recognition, (2014), 2798–2805.

Wiseman, Richard J., and Tamami Nakano, "Blink and You'll Miss it: The Role of Blinking in the Perception of Magic Tricks," *PeerJ*, 4 (2016), e1873.

Wolfe, Jeremy M., and Todd S. Horowitz, "Five Factors that Guide Attention in Visual Search," *Nature Human Behaviour*, 1 (2017), 1–8.

Wright, James R., and Kevin Leyton-Brown, "Level-0 Models for Predicting Human Behavior in Games," *Journal of Artificial Intelligence Research*, 64 (2019), 357–383.

Yarbus, Alfred L., *Eye Movements and Vision* (New York: Springer, 2013).

Zajonc, Robert B., "Attitudinal Effects of Mere Exposure," *Journal of Personality and Social Psychology*, 9 (1968), 1–27.