Predictable Effects of Bottom-up Visual Salience in Experimental Decisions and Games

By XIAOMIN LI and COLIN F. CAMERER*

“Bottom-up” visual perception is fast, universal, and does not depend on personal goals or experience. Where exactly bottom-up attention is directed in a visual image—usually based on contrast, color, feature orientation, and centrality—can be accurately predicted by machine-learned algorithms. We test whether this type of salience can help explain economic decisions in four experimental analyses (three are new). When people pick between sets of valued fruits, salience can lead to mistakes (lower-value sets are sometimes chosen because they are salient). Two studies show strong and weak influence on choices, respectively, of salience in stock price charts and salient payoffs in game matrices. The central analysis is evidence from games in which choices are locations in images. When players are trying to cooperatively match locations, concentration of salience is associated with the success of matching ($r=.57$). In competitive hider-seeker location games, all players choose salient locations more often. This fact creates a “seeker’s advantage” (9% wins compared to the 7% predicted by unique equilibrium). The location game data are consistent with cognitive hierarchy and level-k models in which predicted salience influences nonstrategic level-0 choices. Model estimates from hider-seeker games can predict behavior in matching games despite their opposite strategic structures. Looking forward, bottom-up salience could be included as a predictable type of costless involuntary “pre-attention” in expanded models of Bayesian-rational inattention. Novel applications involving prices and taxes, disability, ethnic markers, and visually-influenced beliefs are suggested. (236 words)

JEL: C91 - Laboratory, Individual Behaviors, C72 - Noncooperative Games

* Wednesday 9th December, 2020, Li: Div HSS, Caltech, xli2@caltech.edu. Camerer (corresponding author): Div HSS and Computational and Neural Systems, Caltech, camerer@caltech.edu. 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 F90MH094258. 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, and Bocconi IGIER. We especially thank Elke Weber, Vince Crawford, Richard Thaler (for the Schelling list idea), Adam Sanjurjo (for “The Pearl” tip) and several anonymous referees for helpful comments, Anne Karing for a valuable image, Gidi Nave and Nina Solovnyeva (SURF 2020). Extra special thanks go to Eskil Forsell, Milica Moorman, Alec Smith whose energetic prior research on this topic laid the foundation for this paper, although their own work was never published due to poor project management by CFC.

1

Electronic copy available at: https://ssrn.com/abstract=3308886
I. Introduction

Attention is scarce. Salience is what grabs scarce attention. Bottom-up visual salience is what the human visual system notices most quickly, effortlessly, and universally. Visual salience is sufficiently well-understood that trained algorithms can be used to predict this type of bottom-up salience for any visual image.

We explore whether such one algorithm can improve predictions in experimental choices from object sets, investment options, and in games. The central question is whether salience prediction explains some of the variance in choices, which is not predicted by other theories. The follow-up question is how bottom-up visual salience could be useful in other ways for economics (we return to this question in the concluding discussion).

Bottom-up salience is one of two somewhat distinct types of salience.\(^1\) Bottom-up is perception of contrast, color, orientation, and other stimulus features (shapes and faces). Bottom-up perception is fast, automatic and, as far as we know, culturally universal. Bottom-up salience depends only on stimulus features, and not what most people expect, value, are familiar with, or are trying to do with what they see. Top-down attention is why the face of a popular BTS band member in a crowd of Korean boys will be highly salient for many K-pop fans, and not at all salient for others.

The driver of our empirics is a highly-trained bottom-up algorithm (called SAM\(^2\)). SAM takes any 2-D image as an input, and predicts what most people will look at in the first few seconds. Suppose that immediate bottom-up looking influences later choices somewhat involuntarily (which is an hypothesis that we will test). Then, you could create highly salient choices which people would pick more often just because they are salient, even if they are not valuable. While earlier experiments suggest that this can be done (Milosavljevic et al.,

\(^1\)Like most scientific dichotomies, there is not a sharp dividing line between bottom-up and top-down; see Awh et al. (2012); Gottlieb (2012) for discussion.
\(^2\)SAM is the first of several acronyms we use repeatedly. They are summarized in Appendix Table A7.
2012; Towal et al., 2013), our first experiment extends the earlier work by using SAM to measure bottom-up salience, and to see whether it sometimes leads to low-value choices. Visual salience should interest economists because there has been increased attention in economics to, and innovation about, both salience in general and specific theories of salience. The SAM model of salience is different in form than these discussions and methods. Because the differences are large, we defer the comparison of the SAM-based approach to other models to the conclusion section, after readers have had a chance to see the details of how SAM works in multiple choice settings.

#02 Before proceeding, we note that in many modern economic usages, the terms “attention” or “salience” are used to refer to an implicit decision weight which is placed on different pieces of information as revealed by choices (e.g. Taylor and Thompson 1982). We prefer to call this concept “inferred salience”, in contrast to saliency that can be directly measured by eye-tracking or by other real-time (“online”) sensory response to stimulus. For example, an investor might see a time series of prices, look at each one for the same amount of time—indicating equal salience—but then invest as if she is computing an expected return which weights recent returns more highly. In this example, memory (Bordalo et al., 2020; Wachter and Kahana, 2019), internal “rehearsal” or additional processing of the recent returns is inferred salience, which differs from the measured visual saliency. Inferred saliency is therefore a composite of different attentional, memory, and motor processing for human decisions of interest. In this paper, we are trying to conceptually and empirically trap and understand the first part—sensory-driven saliency—of that more complicated process.

II. The Saliency Attentive Model (SAM) algorithms

Algorithms which take visual images as inputs, and output predictions about where people will look, have been an active area of research in visual neuroscience since the 1990s. The algorithms are “trained” on large sets of data measuring
where neurotypical adult humans actually look when they are freely gazing at each image in a large set for a few seconds. These algorithms have made great strides rapidly. An Appendix A section describes the history of earlier models leading up to the one we use, called the Saliency Attentive Model (SAM).

The SAM algorithm is based on a machine learning method called a convolutional neural network (CNN) with a long short-term memory structure (LSTM, Cornia et al. 2018). One reason these algorithms have developed so quickly is that researchers can try out new ideas on four popular open-access saliency datasets (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, recorded using eyetracking and other high-quality methods measuring visual attention.

The reported performance of SAM on the website MIT-Saliency is 0.88 using the AUC-Judd measure (Riche et al., 2013). By this measure .50 is random and 1.0 is perfectly accurate. This accuracy significantly surpasses earlier algorithms, and approaches the accuracy of the best human-to-human benchmark (which is 0.92).

The SAM algorithm takes any image as an input. The output is a predicted saliency map. The saliency map is a saliency value from zero to one assigned to each pixel of an 2D image. The saliency map is typically shown as a “heatmap” in grayscale or in color where warmer (redder) colors indicate higher salience. Figure 1 shows an example of SAM outcome. We adopted the default parameters from the original approach and applied it to our image dataset. There are no additional free parameters.

---

3The best human benchmark indicates how strongly two large different sets of human fixation maps correlate for the same image. Each of the two sets contains many different individuals. The limitation is due to individual differences in their fixation patterns (Judd et al., 2012).

4We use the standard color protocol “jet” in Matlab for all the heatmaps in this paper.

5Note that this CNN model, or any simpler variations of it, could be retrained on new data to understand different kinds of salience. For example, two studies have coded abstract features of strategies in 2-person matrix games (e.g., minimax, equal payoffs, level-1) and fit ML models using those features to explain observed choices. Hartford et al. (2016) is a neural network and Fudenberg and Liang (2019) is a random forest.
To illustrate salience and SAM, look at the map drawn by Schelling (1960) in his famous discussion of focality and “psychological prominence” (see Figure 2a). 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, with a bridge located across 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? (p. 56)

Schelling said seven of the eight people who saw his map chose to rendezvous at the bridge.

In a larger incentivised experiment, N=61 UCLA students, who earned $1 if
they match, chose the bridge 59% of the time (see Figure 2b).6

The SAM algorithm predicts that the bridge area, and the upper left road fork, are most salient (Figure 2c). Note that SAM also predicts that the “x spot” is not salient, even though it was chosen by 25% of the subjects. Bottom-up SAM is wrong about “x” because it is trained on free gaze, so SAM does not learn the importance of either top-down cultural knowledge or the participants’ use of shared culture to achieve the top-down goal of matching. For bottom-up algorithms, “x” is just two diagonal lines that meet symmetrically in the middle. The algorithm is not trained to know that “x” is also a letter of the English alphabet which is familiarly known (to some UCLA subjects) to sometimes indicate locations of buried treasure on a map. Those specialized cultural meanings and value of “x” come from top-down attention, which is guided by the common goal of coordination via shared cultural salience.

To distinguish the effects of visual salience and shared meaning further, we did an online experiment in which 11 map locations are described in a verbal list, but there is no visual map. Matching the list choices of others gave a reward. The most popular choices were “X on the map” and “small house near the pond” (49% and 14%); only 5% chose “bridge” (details see Appendix A.A5). Thus, the popularity of the bridge choice depends on visual presentation.7

This simple example sets up the central empirical question in our paper: How well does bottom-up saliency— as predicted by SAM— predict actual choices in decisions and games? Does bottom-up saliency get partially or entirely inhibited when top-down, goal-directed attention takes over?

We describe three new experimental applications, (and one from a different study reported in detail elsewhere). They are:

1) Choices between visual images of two sets of fruits, which vary in induced values and in predicted salience. These data measure how strongly salience

---

6These data were collected in conjunction with Milica Moormann and Alec Smith.
7Rihn et al. (2019) finds a related effect, that visual attention to a logo rather than text description of a type of plant changes valuation.
competes with value, so that people sometimes pick lower-value sets because they are higher in salience.

2) Experimental investments in both actual and artificial stock price series (Bose et al., 2020). These data measure whether salience explains variance in investment decisions, beyond traditional moments including mean and stock returns. It is also an opportunity to interpret what SAM finds salient in terms of statistical features like price jumps (a process called variance of past “feature relevant” explanation).

3) Strategic choices of locations in visual images. In Schelling-style matching games, both players are rewarded if they choose the same location. In hider-seeker games the hider wants to mismatch and the seeker wants to match. These data measure whether a cognitive hierarchy structural model can fit data, and more ambitiously, make cross-game predictions from the hider-seeker game to the matching game.

4) Two-player 2x2 strategy matrix games. These data measure whether bottom-up saliency biases-- which tend to strongly favor looking at the top row, and weakly favor looking left compared to right--can explain strategy choices. This is a tough challenge for bottom-up theories because the experimental participants have a clear top-down goal, which is to choose payoff-
maximizing rows or columns, which usually requires inhibiting bottom-up saliency.

III. Decisions: Fruit displays and lifelike investment decisions

A. Study 1: Saliency and Induced Value in Visual Fruit Displays

#17 The first experiment measures the empirical importance of visual salience in a simple setting that is lifelike. Subjects are shown two fruit sets presented on the left and right parts of an image, as shown in Figure 3a. Each fruit type (e.g., apples or oranges) has a unique, pre-determined induced monetary value (Smith, 1976). The induced values artificially create value so that we can clearly judge if people are making mistakes, and also, we hope, swamp minor differences in intrinsic subjective value from personal or aesthetic preferences for fruits.

Subjects learn the assigned induced values of different fruits before the main session of 20 choices. 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 and has in mind a retail price for each fruit. The retail price of the fruit induces value to the vendor buying wholesale.

While the vendor should optimally be computing resale value, the visually salient properties of fruit (color, intensity, orientation, whether they “pop out” from a fruit basket) will influence bottom-up perception. The salience and value properties are independently controlled by our design. We constructed choice sets in which visual salience and fruit value are 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 lead to mistakes.

The main experiment includes 20 images similar to those in Figure 3. Choices are made with a 10s time limit. Trials are balanced between induced values, numbers of fruits in the two sets, and whether the more salient set is on the left.
Figure 3. Fruit Experiment

Note: a) illustrates the rule of this task. Each fruit worth a certain amount of dollar values. The value of a pile simply equals to the summation of all fruits in that pile. b) presents a sample image of a real trial in this task. c) presents the SAM result for the sample image in b). We can see clearly that the left pile is more salient than the right pile in this example. All images used in this task have the similar saliency distribution to this example that there is only one saliency peak of the entire image and the peak is in one of the two piles.

Saliency peak means the value of saliency is 1. Not only that, our samples satisfy that the difference of two peaks is large enough with the average difference being 0.24

or right (see appendix). Subjects earn money based on the induced value of the sets they choose in an incentive compatible design (a 10% chance of earning the value from one randomly-selected trial).

To measure saliency, the SAM algorithm was used on a large set of manually-created fruit displays. Images were chosen that had a clear difference in saliency between the left and right sets (see Figure 3). All images have only one maximum for saliency values, which is contained in only one of the two piles. The average difference between the two most salient peaks of the two piles is 0.24 in the 0-1 scale of saliency, which is a noticeable difference. More ambitious designs could covary the size of the salience difference with the size of value difference between the two sets. For our purpose, however, the goal was to just see if visual salience influences choice at all.

In half the trials SAM-salience and value are “congruent” (one set is higher on both values) and in the other half of the trials they are “incongruent” (the high-salience set has lower value or vice versa). The dependent variable is choice accuracy— did they choose the most highly-valued set?

With a 10-second time limit, choice accuracy drops from 85% to 79% between the congruent condition and the incongruent condition. This drop in accuracy,
### Table 1—Results of saliency in simple choice problem (fruit pile)

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correctness</td>
<td>0.44***</td>
<td>0.51***</td>
<td>0.47***</td>
<td>0.54***</td>
</tr>
<tr>
<td></td>
<td>(0.14)</td>
<td>(0.15)</td>
<td>(0.15)</td>
<td>(0.15)</td>
</tr>
<tr>
<td>Congruency</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>abs(valueDiff)</td>
<td>0.80***</td>
<td></td>
<td>0.79***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.11)</td>
<td></td>
<td>(0.11)</td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>1.32***</td>
<td>0.54***</td>
<td>1.56***</td>
<td>0.76*</td>
</tr>
<tr>
<td></td>
<td>(0.09)</td>
<td>(0.12)</td>
<td>(0.42)</td>
<td>(0.44)</td>
</tr>
<tr>
<td>Covariates</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Observations</td>
<td>1,382</td>
<td>1,382</td>
<td>1,307</td>
<td>1,307</td>
</tr>
<tr>
<td>Log Likelihood</td>
<td>-643.88</td>
<td>-591.12</td>
<td>-607.87</td>
<td>-557.47</td>
</tr>
<tr>
<td>Akaike Inf. Crit.</td>
<td>1,291.76</td>
<td>1,188.24</td>
<td>1,235.74</td>
<td>1,136.93</td>
</tr>
</tbody>
</table>

**Note:** *p<0.1; **p<0.05; ***p<0.01

**Note:** A trial is congruent if one option is both more salient and more valuable (so the other option is low on both dimensions). Otherwise, congruency = 0. abs (ValueDiff) is absolute terms of value differences. All standard deviations in this model are clustered on per subject level. “Covariates” denotes for whether the current model contains covariates or not. The model will contain additional controls: education, gender and income if Covariates is yes. The main effect estimates are not sensitive to these controls.
when saliency competes with valuation, is highly significant (p-value = 0.002, two-sided t-test). The statistical strength of these results is summarized in Table 1 which controls for individual-level clustering of standard errors. The Table results from a logistic regression model which regresses accuracy (=0-1 dummy variable) against the absolute choice-set induced value difference and the congruency (=0-1). Both variables significantly affect choice.

There are two boundary conditions in which the effect of visual salience disappears. These are expected because the SAM predictions are about rapid bottom-up perception which is usually replaced by effortful top-down attention.

One boundary condition which erases the effect of salience is when the induced value difference is large (using a median large-small split across the 20 images). When the value difference is large the accuracy is 0.94 for both congruent and incongruent conditions (p=0.91 for the test for a difference). When the value difference is small, the accuracies are lower and salience-value incongruence does have an effect (0.78 to 0.69, p = 0.01)

The second boundary condition is endogenous time allocation: When there is no time limit (N=25), participants in both conditions are near the ceiling of perfect accuracy (congruent =0.94 and incongruent=0.96). This no-pressure result merits an important qualification. As a result, in everyday economic choices, there is always an exogeneous time limit. In models such as rational inattention, and most theories of evidence accumulation (Sims, 2003, 2006; Caplin and Dean, 2013, 2015; Krajbich et al., 2010), if attention is costly then the goal is not to get the correct answer 100% of the time. In psychology this is called the speed-accuracy tradeoff; in economic terms it is a cost-accuracy tradeoff. If we think of bottom-up attention as having essentially zero cost, and top-down induced value attention as being scarce and costly, it should rarely be the case that the effect of initial bottom-up saliency is entirely wiped out by costly top-down attention. We cannot test this in our data because we do not have direct attention measures, but such data would be useful for testing these
ideas (cf. Towal et al. 2013).

A final remark in this section is that what is bottom-up salient may not be explicitly (i.e., consciously) known to people. To measure explicit knowledge of bottom-up salience, N=50 Prolific users were shown 18 of the basic raw fruit-pile images (as in the left of Figure 3) that were used study 1. They were told that “a computer algorithm designed to predict what people look at, in the first few seconds, predicted whether people would look left or right”. They were asked to predict whether left or right would be judged more salient by the algorithm, and they were incentivized for being accurate. Their accuracy averaged 59%, a little better than 50% chance \( p \leq .0001, \) interquartile range \([.50, .72]\)). So they have a modest understanding of what is salient.

\[ 8 \]

**B. Study 2: Saliency and Historical Returns Weighting in Equity Investment Experiments**

The second application of visual SAM salience is in investment experiments (Bose et al., 2020).\(^8\)

Participants see many different one-year time series of daily stock prices. For each time series, they decide what proportion of an endowment to invest to earn a return in the next year, after the historical year they have seen. Their experiment earnings are determined by those investment choices, and by the actual next-year returns.

Only the most basic features of this study are described here. The purpose is to understand whether SAM visual saliency is potentially useful in this domain of visual display and decision, which is different than the others we discuss.

The leftmost Figure 4a,c show grayscale heatmaps of the density of actual gazes, from N=57 human participants, which were recorded using eyetracking. The corresponding rightmost figures are the time series of actual prices over 60

\(^8\)Bazley et al. (2019) show experimentally that the highly salient color red influences risk-taking, expectations and trading. The influence is weaker in people who are color-blind and for Chinese for whom red has a special top-down meaning, connoting prosperity. \( \#33 \)
time periods. Blue-to-red heatmap colors in Figure III.Bb,d show the SAM predictions of where visual attention will be.

In the two examples, salience predicted by SAM appears to be correlated with actual eyetracking gaze density. The pixel-by-pixel Pearson correlation averaged across all time series has an average of .52 (compared to the original SAM-groundtruth correlation which is .88). Thus, participants are often looking where the SAM algorithm predicts they will look.

But looking is not choosing. One can intuitively think of visual perception as necessary, but not sufficient, as a basis for choice. To see whether salience seems to influence choice, the authors compute weights for price-to-price percentage returns based on visual SAM salience of adjacent prices. These saliency weights are then used to compute a saliency-weighted expected return. For example, in Figure 4(b) the SAM salience is highest in the center of the price time series, where there is a big positive return. If salience is guiding value perception, this stock will have a high saliency-weighted expected return, because attention is predicted to be strongly focused on the central large positive return. If their choices are influenced by salience, participants will invest heavily in it.

Indeed, there is a substantial correlation between saliency-weighted returns and investment, in three experiments. These correlations are above and beyond control variables which include the typical moments of historical returns (mean, standard deviation, and skewness), which also strongly influence investment decisions.

It is useful to ask whether visual salience corresponds to particular micro-features of the stock price time series, such as large “jumps”, short periods of rapid up-and-down negative autocorrelation (generating high visual contrast), primacy or recency of early or later returns, etc. The authors find that about 20% of the variance in saliency values can be explained by such statistics (see

9The eyetracking experiments used a temporally-coarsened series of 60 returns rather than around 250 daily returns, because of the limited spatial resolution of eyetracking. This experiment is just to see how well SAM and eyetracking are correlated, if at all.
Note: Two examples of the groundtruth density maps based on eyetracking (a,c), and the SAM-predicted heatmaps (b,d).

details in section below). That is, saliency, trained on free gaze of a large corpus of images (none of which are stock price charts– is associated with some of these statistics, but most of saliency is explained by something else. This analysis will be revisited below when we talk about “explainability” of black-box neural nets such as SAM.

IV. Games: Location and Matrix Games

A. Study 3: Matching and Hider-Seeker Location Games

The core of this paper is new experimental data from location games. In these games, two players see a common visual image and simultaneously choose a “location”– a pixel. (Schelling’s map game is one such location game.) We create a circle around the pixels (with a radius of 108 pixels). The circle is about 1/5 of the screen width, about the size of a US nickel. The baseline circle size is

10 The circles also wrap around the 2D boundaries as if the 2D image is a torus.
chosen so that if players are choosing pixels randomly, they will match 7.1% of the time (One experimental treatment below varies the circle size).

Before turning to data, we should point out why these classes of games are economically important. Matching games are models of natural socioeconomic situations with strategic complementarities. Important cases include network externalities in industries, strategy-structure complementarities within firms, and macroeconomic global games. There is no obvious role for visual salience in many such games. However, other matching game structures arise from the desire to conform. In conformity games, including fashion choices, cultural fads, or street protests, visual saliency might influence what everyone notices, and could act as an equilibrium selection device.

Hider-seeker games have a different structure in which seekers want to match and hiders want to mismatch. Interactions of this kind include predator-prey relations in nature and human examples such as choosing passwords to outwit hackers. Other examples are “coding” language and signals used in sports, gangs, and other rivalries to coordinate action with teammates and avoid detection by the other side, and industries with follower-leader dynamics (e.g., leaders want to “hide” by choosing unique new designs, and outsiders want to “seek” by matching those designs). Visual salience might conceivably play a role in some of these games.

Returning to the design, the experiment has three blocks of games: matching, the hider-seeker game as a 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 order randomized between subjects. During each block there is a “feedback” sequence in which the choice the other player made is revealed right after choice, by showing the circle around the other player’s pixel choice. In a “no feedback” sequence those results are not revealed.

The matching block was two sets of 20 images for the two feedback treatments

\[11\text{See e.g., Vives 2005, Milgrom et al. 1991; Milgrom and Roberts 1995, Morris and Shin 1998.}\]
(40 images) and the hider-seeker game used the same set of 19 images for each treatment (38 images, which subjects played once as hider and once as seeker). An additional short session of hider-seeker game follows right after (16 images) with 10x bonus payment than before. There was unlimited time to read instructions but only 6s to make a choice for one trial. They get zero if they didn’t respond before the known time limit (see the original instructions in appendix A.A8). The results shown to subjects in the feedback condition were drawn from previously tested actual subjects (using different “stranger” subjects for each image).

N=151 subjects participated, excluding a pilot dataset for power analysis. Of these 151 subjects, N=29 subjects (13 males, 16 females) participated in a lab, one at a time, in a small testing room where their eyetracking was recorded. N=15 subjects were from the Caltech community and N=14 from the neighboring community (there were no differences in results between the two groups). They will earn $0.2, $0.1, and $0.4 in matching, hiding, and seeking games differently, for each “win” per trial (image). They will earn the cumulative monetary amount in the end of the experiment.

The other N=122 subjects participated online through Amazon Mechanical Turk (“mTurk”).

There are some behavioral differences between the feedback and no-feedback conditions. The largest is that the matching rate is much higher with feedback (64% vs. 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 Appendix.

12#39 Online experiments have the same instructions and block orders as in lab version, except that now everything is shown in a web browser. This study was pre-registered on the Open Science Framework (https://osf.io/yuqjg/) during data collection and before analysis. The sample size was pre-determined before the data collection process, based on a pilot study (N= 29) carried out in March 2017.

13Both feedback and no-feedback blocks were included because each one answers a different question of interest. In ensuring subject comprehension, and especially in testing equilibrium concepts, the standard practice in experimental economics is to provide feedback after each trial. However, whether focality is influential even with no feedback about other players’ choices is an interesting question too (this no-feedback version is closer to what Schelling had in mind in his map example).
Two choices were considered a match if two circles centered at their pixel choice spots overlapped. Figure 5 shows examples of result screens that subjects saw during the experiment.

Figure 5. examples of trial outcomes showing circled pixel choices

Ninety-two (94) visual images were displayed on an eye tracking monitor (it is a normal computer screen in 1920x1036 resolution). Images were randomly selected from a large image pool (273) with five categories (abstract art, city, face, social, nature). The image set contains both images with only one obvious saliency spot and more complex images, which have multiple saliency centers (Judd et al., 2009).

B. Analysis and Results

Equilibrium analysis generates a statistical benchmark for what people might do.\textsuperscript{14} In location games, strategies are pixels in x-y space (and resulting circles).

For a matching coordination game, choices by the players of any two different

\textsuperscript{14}An important mathematical touchstone is “correlated equilibria” (Aumann, 1974) When both players receive a common public signal and a strategy is conditioned on different 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 which coordinate traffic and reduce accidents. In these terms, our study is about whether bottom-up visual salience of image locations works as a correlating device in matching games.
pixels which create overlapping circles constitutes a pure strategy Nash equilibrium. One image contains about two million (=1920×1080) pixels. Any pixel match is a pure equilibrium, so there are an enormous number of equilibria. In addition, there are many mixed equilibria. The standard game theory predicts anything can happen.

For the hider-seeker game, there is a unique Nash equilibrium in which all locations are chosen equally often.\(^\text{15}\)

The fact that equal randomization over all strategies is the unique hider-seeker equilibrium clarifies how logic conflicts with bio-logic. The last thing the brain is equipped to do is to ignore differences among many objects and choose them equally often. The human perceptual system evolved to efficiently filter a huge amount of information hitting the retina, estimated to be \(10^8\) bits of information, into only about 100 bits, by focusing attention on only most valuable information—the most salient. For the same reason we are so good at quickly noticing salient information, we are likely to be naturally bad at rapidly choosing what is unsalient, to hide, or to randomize equally by inhibiting salience entirely.\(^\text{16}\)

C. Matching games

To analyze the behavioral data, we first test whether subjects are playing an equal random mixture across all pixels and associated saliency levels. To compare results from different images, all saliency values in this section refer to the normalized levels, which are the rank percentiles of raw measures from the algo-

\(^{15}\)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 \(n\) locations) by an iterated logic. A special design that, if a circle touches any boundary, it wraps around from the opposite boundary, guarantees the equilibrium.

\(^{16}\)A similar conflict between logic and biology occurs in the games “rock, paper, scissors” and matching pennies (e.g., Crawford et al., 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).
algorithm, ranked within each image. We calculated the normalized saliency value for each click point (the choice saliency level) and then compared these values against the baseline of equal randomization independent of salience. Kolmogorov-Smirnov tests reject the hypotheses of randomness for all treatment conditions ($p < 10^{-4}$).

To get an example of how salience affects choices, we plot the choices from all the subjects on two specific images, shown in Figure 6. The saliency heat map

![Figure 6. Matching game image, saliency heatmap, and choices (red)](image)

*Note:* (Left column) The original images. (Middle column) The original images overlaid by the saliency maps. (Right column) The grayscale original image overlaid with the actual empirical choice distributions (more red indicates more frequent choice of an area).

is in the middle column. The right column shows, in redscale, the frequencies of subjects’ actual location choices. The predicted saliency in the middle column and the observed choice maps in the right column are highly overlapping.

Statistically, the mean of the saliency levels of actual points chosen in the coordination game is 0.87, which is far above the chance level of 0.5 ($p < 10^{-4}$).

**D. How predictable is the matching rate across images?**

Intuitively, the matching rate for an image should be affected by how dispersed saliency is. (This prediction also comes structurally from the SCH model described below). When saliency is highly concentrated then the rate of
choosing the same pixels, and matching, should increase if people are trying to match. And if there are many salient locations, then matching should be harder. Dispersion of saliency throughout an image is measured by the number of local saliency centers.\(^{17}\)

![Figure 7. Correlation across images between matching rate and number of saliency centers](image)

**Note:** (a) is an image with seven saliency centers (c) an image with one saliency center. (b,d) are corresponding maps (red dots) of actual choice data in a matching game. The choice map in (b) is more dispersed because the saliency centers in (a) are more numerous. (e) is the correlation between the number of saliency centers and the matching rate using both the feedback session and the no-feedback session to get a large image pool (image N = 40).

\(^{25}\) Figure 7 shows that indeed, the matching rate\(^{18}\) is negatively correlated

\(^{17}\) The typical raw saliency map has flat local maxima with many adjacent pixels with nearly-equal saliency. To detect saliency centers we first Gaussian blur (with [300pixel,300pixel] window size and standard deviation \(\sigma=75\) pixels) the entire image to smooth hyperlocal spikes in saliency. Gaussian blur uses a 2D convolution to smooth an image through a Gaussian function. Then we simply take the number of local maxima for the saliency distribution using the matlab function imregionalmax() with default settings. That function takes the local maximum inside each 3pixel*3pixel patch. If the original image has two local maxima that are close enough together, the Gaussian filter combines them.

\(^{18}\) This result is based on all images from both feedback session and no-feedback session using the
with the number of saliency centers (Pearson $r = -0.57$, $p < 10^{-4}$, df = 38). The actual and predicted matching rates span most of the range from perfect matching (100%) to random (7%). These results suggest that for any image, the matching rate can be predicted ex ante with substantial accuracy. Put the other way around, it is possible to find images with saliency 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 attract shared attention or undermine it.

**E. Hider-seeker games**

For the hider-seeker game, let’s start with an example image (Figure 8). Figure 8 shows that subjects’ choices (c is seeking and d is hiding) are much more spread out than in the matching game where saliency and choice strongly overlapped.

Figure 8. hider-seeker game image, saliency, and choice

*Note:* a: The original image. b: The original image overlaid by the saliency map. c,d: The grayscale original image overlaid with the actual empirical choice distributions (more red indicates more frequent choice of an area). c is for seekers’ choices and d is for hiders’ choices.

At a reader’s suggestion, we also calculated whether the number of saliency centers was correlated with the seeking win rate (across N=38 images). There is no correlation (Pearson $r = -0.10$, $p=0.23$).
In Figure 8d, the peak of the hider choice distribution no longer falls in the most salient area. However, in this example Figure 8d hiding game, the most salient locations are still chosen frequently by hiders; while the hidings as a group are moving away from salience, they still choose salient locations (especially the lower left, where the green tent is) more often than they should to hide successfully.

The direction of effects suggested by these two examples holds more generally. In the hiding game, the mean saliency level of user click points is 0.53, close to the chance level of .50.\textsuperscript{20} The mean saliency level chosen by seekers is higher, 0.61.\textsuperscript{21} A replication sample (\(N = 29\)) with payoffs ten times higher had very similar results, average saliency levels of 0.51 and 0.64 for hidings and seekers.\textsuperscript{22} A paired t-test showed this difference in choice saliency between hidings and seekers is highly significant (\(p < 10^{-4}\)), reflecting what is seen in the Figure 8 example. The no-feedback results have a similar difference (see Appendix A.A3).

**Seeker’s advantage:** Recall that the theoretical value for two circles to match from two random players is 0.071. Table 2 presents the realized matching probability in a specific game condition.\textsuperscript{23}

<table>
<thead>
<tr>
<th>Table 2—Realized matching rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nash mixed prediction</td>
</tr>
<tr>
<td>Matching game</td>
</tr>
<tr>
<td>Hider-seeker game</td>
</tr>
<tr>
<td>Hider-seeker high payoff (10x)</td>
</tr>
</tbody>
</table>

Note: 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.

\textsuperscript{20}p-value = 0.02, t-test CI: [0.51, 0.56]
\textsuperscript{21}mean = 0.61, p-value < 10^{-4}
\textsuperscript{22}Hiding: p-value for test against null of .50 saliency = 0.59, CI: [0.48,0.54], seeking: p-value < 10^{-4}
\textsuperscript{23}Tests to compare the matching rates with random baseline were carried out by bootstrapping a hiding data and a different person’s seeking data (or two data points from matching game) for 1000 batches (batch size is 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 sub-block, same image.
To check the robustness of these results, we also did the hider-seeker game online in a high-payoff condition with payments 10 times as large (N=29) and in a between-subjects condition where subjects played only one role (N=53). In both conditions, we get the same seeker win rate of 0.09, as in the main experiment.

This seeker advantage also doesn’t rely on time pressure. N=46 people from Mturk participated in the same hider and seeker experiment, but without a time limit. The seeker’s advantage remains significant at the level of 9% (p-value = 0.002). Subjects spend on average 3.14s, 4.61s, and 6.44s in matching, hiding and seeking conditions, respectively, when there is no time limit.\(^{24}\)

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 people. This null hypothesis supplies a lot of statistical power because it justifies pooling all observations together. So when independence is assumed the test has enough power to establish that 9% is significantly higher than 7%. A more conservative approach averages all data within an image and tests whether the image-wise matching rates are above 7% (N=19, p= 0.0005). A different conservative approach averages win rates for individuals and test whether the average individual seeker win rate is different than the Nash 7% (N = 29, p= 0.002).

A reader pointed out that the seeker’s advantage could depend on how large the circle is that is drawn to surround the chosen pixel. To explore this the circle size was enlarged to be 1.5x as large; then the chance and equilibrium matching rate are about twice as high, 16%. The seeker win rate was 18%, so there is still a small seeker’s advantage equal in absolute size (+2%) to the regular circle baseline (p = 0.003, N = 66).

This advantage should be due to some kind of correlation between the hiders’

\(^{24}\)The standard deviation are: 7.10s,15.54s and 19.49s for matching, hiding and seeking. These large standard deviations are not unusual for an online environment with unlimited time.
choices and the seekers’ choices. We have already shown that both of them prefer saliency at different levels. But how exactly do those biases lead to the seeker’s advantage? We show the conditional seeking win rates on different saliency level for hiders and seekers, separately, in Figure A7 (window size = 0.1 for saliency 0-1). The seeker’s advantage is mainly due to the concentration at the top 10% saliency level points.

V. A Saliency-perturbed Cognitive Hierarchy Model (SCH)

This section describes a parametric behavioral model meant to explain choices and their salience-sensitivity, closely following Crawford and Iriberri (2007a). It uses the level-k model of Stahl and Wilson (1994) and Nagel (1995), later extended by Camerer et al. (2004).

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

A. General Model description

The population consists of different levels of players starting from level zero. The proportion of level k players with frequency f(k) (with f(k) assumed to be Poisson distributed with parameter τ).

For all levels of players, we allow some degree of randomness which will be described using a conventional logit softmax function \( \frac{e^{\lambda x_n}}{\sum_m e^{\lambda x_m}} \) with parameter \( \lambda \). When \( \lambda \) equals zero, agents choose pure randomly without any response to differences in valuation. When \( \lambda \) approaches infinity, agents choose the best option.

In this SCH specification, the nonstrategic level zero players weakly prefer salient choices. The probability of choosing strategy/pixel n depends on the
direct saliency value $S_n$ from SAM according to:

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

If $\mu = 0$, salience is ignored and level 0 types choose randomly among all points. We assume that $\lambda$ and the saliency weight $\mu$ are common across subjects, although many heterogeneous versions are conceivable (e.g., Rogers et al. 2009).

Other than level zero players, all other levels of players 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 saliency only enters directly into the value calculations of level 0 players. This assumption tests whether a model in which saliency only enters $k \geq 1$ level players through beliefs is a good approximation (a la Mehta et al., 1994a in which “secondary salience” derived from primary salience).

**B. Model fitting results**

Besides the SCH above, there are many other ways to specify models of limited strategic thinking, which have been mixed and matched in previous research. We therefore fit six different model specifications to the hider-seeker data (see Appendix A3 #28).

#21 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

---

25 Just as before, the saliency values refer to the normalized ranking with respect to each image. This way, we can model data from different images and saliency distributions.
types (though see Wright and Leyton-Brown, 2019). Restricting \( f(0) = 0 \) in this way clearly degrades fit (Table A2). We therefore focus only on \( f(0) > 0 \).

A close relative of SCH is the “Level-k” model, which refers to level \( k \) types believing all others are level \( k-1 \) (rather than distributed from 0 to \( k-1 \)) (Crawford and Iriberri, 2007a,b). This is usually estimated non-parametrically, 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 equally well by the AIC criterion (although SCH is a little better by BIC). These games are not an ideal testing ground for comparing such differences. The goal, instead, is to see if SCH variants can explain both matching and hider-seeker games, which have different top-down attentional demands.

We first focus on the best specification of SCH. It has four free parameters—\( \mu \): the saliency weight parameter, \( \lambda \): the softmax parameter, and \( \tau_s \), \( \tau_h \): two role-specific parameters. Allowing role-specific \( \lambda \) and \( \mu \) parameters fits worse due to the large BIC penalty for extra parameters.

We used a standard training-testing separation to avoid over-fitting. Recall that each subject does two sessions. We use the first session as a training set to estimate parameters. The parameter values are then fixed and used to predict in the second session test set (see Appendix). The best fitting parameter values and measures of fit are shown in Table 3.

<table>
<thead>
<tr>
<th>Table 3—Estimation details, role-specific SCH</th>
</tr>
</thead>
<tbody>
<tr>
<td>Best fit parameters</td>
</tr>
<tr>
<td>----------------------</td>
</tr>
<tr>
<td>( \mu )</td>
</tr>
<tr>
<td>Number of observation</td>
</tr>
<tr>
<td>95% CI</td>
</tr>
</tbody>
</table>

Note: The parameters \( \mu \) and \( \lambda \) are constrained to be the same for both hiders and seekers. The confidence interval in the table is calculated using the bootstrap method with batch size 1096 for hider, 1090 for seeker and number of iterations 100.

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 9 shows the comparison between actual choice density (frequency) maps versus model predicted density maps for the hider-seeker game. Training data are shown in the top Figures 9ab and test data are shown on the bottom Figures 9cd. In the choice data, there is a sharp density increase starting around 0.9 saliency for both games (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 decreasing choice from the lowest saliency to higher saliency levels for hiders (but not for seekers), reflecting the fact that some hiders choose the lowest-salience locations. SCH can roughly fit these two major features of the data. The best-fit values of $\tau$, 0.4 and 0.1 for hiders and seekers, are much lower than typical estimates (e.g. Camerer, Ho and Chong, 2004): around $\tau = 1.5$. (see also Riche #44 Furthermore, Appendix Figure A7 shows that the hider-seeker match rates which result from the aggregate data in Figures 9cd are rather constant across saliency levels, with a jump only in the highest decile. 

27
et al., 2013, although Fudenberg and Liang (2019) find minimal prediction error in a large interval (0, 1.25).)

We think the low values of \( \tau \) estimated in our data for SCH results from the fact that the ability to identify \( \tau \) is limited in these visual choice games. A challenge is that level-1 hiders should choose low-salient locations but that behavior is not common. The model concludes that there are few level-1 hiders.

The Level-k model can lend better insight here about level frequencies.\textsuperscript{28} Compared to 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 those level 0 types. Level-k also estimates larger frequencies of level 2 and 3 types (\( f_{\text{seeker}}(3) = 0.66, f_{\text{hider}}(2) = 0.61 \)). While the overall Level-k fit is just a little less accurate than SCH, this type distribution is more consistent with other results than the SCH estimates of low \( \tau \) (see Appendix A.A4).

Thus, it is clear that a single-peaked SCH with Poisson \( f(k) \) does not meet the calibration challenge—level-1 hiders should anticipate saliency choice by level-0 seekers and move sharply to anti-salient locations. But level-1 hiders do not do that. The level-k distribution explains this, sensibly, by simply estimating few level-1 types. The parametric SCH model cannot do so because the frequencies of all types are tied together; so it does the best it can with an unusually low value of \( \tau \).

So while it is clear that both specifications fit the saliency-choice profiles adequately (as seen in the Figures), they suggest different evidence of level frequencies. These games were chosen to investigate the effect of predictable saliency, but not ideal to recover levels. Better methods can be developed.

\textsuperscript{28} A much better way to identify \( \tau \) is with games where different level types choose distinct strategies in a way that is carefully designed to separate them (such as matrix games pioneered by Stahl II and Wilson, 1994, and see Nagel, 1995; Ho et al., 1998; Costa-Gomes and Crawford, 2006; Kneeland, 2015; Fragiadakis et al., 2017.)
C. Cross-game predictive validation

To further test generalizability of SCH, parameters estimated from fitting the SCH model to hider-seeker data are next used to predict choice behavior in the matching game. There is no guarantee that this cross-game portability will work at all. Identification of the saliency weight $\mu$ in hider-seeker comes from the level 0’s and from a small portion who sometimes try to match the level 0 opponents and sometimes mismatch. In the matching games, all higher-level types are similarly guided by salience since 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 (see Hargreaves Heap et al., 2014). Furthermore, matching and hiding-seeking are completely opposite in strategic motives.

Figure 10 compare predictions of the saliency-frequency profile on the test set of matching game data. The left graphs 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”) the matching game results. The right graph shows predictions of matching test-set data using matching data for training (but using the cross-validation described above). Of course, training on the matching data and predicting matching test data should be more accurate than training on a different types of game, and it is ($LL = -1943$). However, training on hider-seeker is almost about 10% worse ($LL=-2176$). The Figures 10 show that the hider-seeker trained parameters do not predict how sharply matching-game data respond to the highest saliency.\(^{29}\)

\(^{29}\)We did not do the opposite analysis, predicting hider-seeker from matching, because matching is poorly suited to estimating different levels since all levels are driven by choosing salient strategies. This can be easily understood. In matching cases, regardless of level-0 assumption, every level leads to the same behavior, either saliency- preferred or the opposite. This causes severe identification problem to do the reverse inference.
Figure 10. The SCH model calibrated on hider-seeker game data can predict matching game choices.

Note: The comparison between the matching data distribution and the fitted matching game distribution.
(a) Parameter estimates from the hider-seeker game are used to predict matching game results. log-likelihood: -2176 (b) Parameter estimates from the matching game are used to predict matching game results. log-likelihood: -1943

D. Matrix games

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

First note that the SAM training set does not have any images which even remotely resemble matrices of payoffs. And subjects in matrix game experiments have a clear top-down attentional goal, which is to look at numbers in a matrix to make a high-payoff choice. These goals create complicated visual search to compute beliefs and implement decision rule. In fact, many studies using eye-tracking have shown these patterns of search which are indications of strategic thinking rather than bottom-up attention (Johnson et al., 2002; Polonio et al., 2015; Brocas et al., 2014; Arieli et al., 2011). Furthermore, most of the behavioral studies about coordination and hider-seeker games have aimed at establishing general principles of focality from top-down goals and set-theoretic properties of
strategies (see Appendix A.A2 for a review).

So it would not be surprising if bottom-up SAM saliency has no predictive power in matrix game.

The possible influence of visual salience is tested here using data from Polonio et al. (2015). In their experiment, N=56 people played 32 normal form games with different strategic structures. Eye-tracking was used to record visual attention. We can compare those actual gaze maps with the SAM predictions and with actual choices.30

Figure 11 shows an example and summary statistics. Figure 11a shows one example of the kind of matrix subjects see on their computer (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 11b is the average prediction of where people look from the SAM algorithm over all 32 games. There is an obvious bias toward looking at the top row, and a row-player payoff bias (even for column players). Figure 11c is the average measured attention map calculated from gaze data over different types of games (filtering out gazes which are away from payoffs). The comparison between Figures 11b and c suggest that the algorithm predicts the actual attention allocation during game plays rather well; this visual impression is supported by conventional statistics used to associate prediction and actual attention in visual science.31 Much to our surprise, the actual gaze data are quite similar for row players, who choose rows, and column players, who see the same column but choose columns (We quadruple-checked this result with the authors).

The next question is whether bottom-up visual attention, measured by SAM and also highly correlated with their recorded eyetracking, influences choices. To test for such an effect on choice we look at the 24 games which contain a

---

30See appendix A.A9 for more details.
31Two validation scores, AUC and CC are commonly used metrics to evaluate how closely saliency algorithm predictions are correlated with actual human gaze in the computer vision field. AUC: area under the receiver operating characteristics curve and CC: Pearson Correlation (see Kummerer et al. (2018). The Appendix Table A6 shows these statistics.
The key result is Figure 11d. It presents the different percentage of choices playing equilibrium strategies conditioning on whether the equilibrium strategy is in a more salient location or not. There are three groups of subjects, as grouped by different strategic levels of thinking from 0 to 2 (based on a sophisticated sequential classification procedure) in the original paper. When the equilibrium choices are in a more salient location, low level players (level 0 and 1) play it a little more often (p-value = 0.065 for a two-sided test and 0.032 for an one-sided test), while higher level players are not as affected (p-value = 0.57). Thus, there is a small influence of bottom-up salience on choices for the least strategic players.

VI. Discussion and Conclusion

Our study goes further than previous research in two ways:

1) Bottom-up salience is predicted by an underlying neuro-computational the-
ory (SAM) of which features of an image or information display most people look at first. SAM salience has a small effect in visualized binary set choices, in investment experiments where subjects see historical price charts, and in matrix games. The effects exist, but are not always strong because in all three cases there are clear goals that guide top-down attention and compete with bottom-up attention.

2) In the main set of experiments with location matching games, salience is a good predictor of where people choose and how often their choices match \( r = -0.57 \). In hider-seeker games, a saliency-perturbed cognitive hierarchy model (and a similar level-k model) can account for the slight seeker’s advantage in hider-seeker games. Parameters fit to hider-seeker data can also “portably” predict the saliency-choice relation in matching games, even though hider-seeker is strictly competitive and matching is cooperative.

A. Model comparison: SAM, rational inattention, and BGS salience theory

This section returns to the important business first discussed, then postponed, in the introduction: How does bottom-up salience relate to more standard economic approaches such as rational inattention and psychophysical approaches such as BGS salience?

Rational inattention is a modelling approach in which people are assumed to have a Bayesian prior about possible states of the world, and allocate attention to distinguish the states, to the extent that total net benefits can be improved taking into consideration that attention is costly (Sims, 2003, 2006; Caplin and Dean, 2015; Caplin et al., 2019; Kőszegi and Matějka, 2020; Caplin et al., 2020). This idea can be tidily expressed mathematically as maximizing some information criterion for how accurately subjective perception matches an objective percept, while subtracting attention cost or imposing a constraint.

A simple and biologically plausible way to extend the rational inattention approach is to treat bottom-up attention of the kind studied here as costless. It’s
costless in the sense that many sensory-motor processes are highly automated and involuntary (without special training). These include the startle reflex, breathing, homeostasis, sleep, and many emotional reactions. These processes are highly evolved specifically to \textit{not} require costly conscious attention.\footnote{The difficulty of inhibiting certain kinds of perception is illustrated by Steinbeck (2011). In this novella “The Pearl” the protagonist Kino has hidden a valuable pearl which everyone in the small town covets. An unscrupulous doctor comes to treat Kino’s baby. “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.”}

Precisely how costless bottom-up attention influences choice in an extended rational attention model clearly beyond the scope of this paper. An appealing approach is that bottom-up attention is fully allocated (because it is free) and creates prior beliefs, which are an input to rational inattention models which begin with priors about what might be perceived.\footnote{The interesting tradeoff is created when the resulting priors from bottom-up perception may systematically shift weight away from high-value percepts. An appealing way to model this is as a signal-extraction problem in which rapid perception is “upstream” and transmits information to a “downstream” system which cannot precisely tell whether the perception-biased prior is due to bottom-up cues or value, as in Cunningham, 2015. In this approach, bottom-up attention both biases priors and may also reduce the optimal allocation of endogeneous attention when bottom-up cues and value are sufficiently incongruent.}

Now we address the question of how SAM (and related algorithms) can be compared to more typical economic models and understood. What are its foundations?

To do so, it is useful to have a sharp comparison. So we will compare bottom-up SAM salience with the Bordalo, Gennaioli and Shleifer (BGS) saliency theory (Bordalo et al., 2012, 2013a,b).

We hope to convince the skeptical reader that while the style of the SAM algorithm is obviously different than the BGS specification, that there is plenty of room for both types of theories, and progress can be made.

In BGS, more extreme numbers are more salient.\footnote{Extremity-based saliency is also used in a computer science model by Zuckerman (2010).} In its parameterized form, the theory has two free parameters which control the size of salience adjustments. The theory has been applied widely, to finance, lottery choices, cross-game at-
tention, and agent-based modelling of drug interdiction (Cosemans and Frehen, 2020; Dertwinkel-Kalt and Köster, 2020; Avoyan and Schotter, 2020; Magliocca et al., 2019).

Defining what the theories are trying to do can help show what they have in common and how they differ.

First we show what is common. Both theories take [numbers] or [pixels] and assign a saliency value to each one, which depends on the entire set of [numbers] or [pixels] in a feature or a choice set. In both approaches, the features and choice sets of interest are economically interesting—prices, product qualities, game matrices, stock returns, etc.

The foundations of the BGS saliency approach are psychophysical principles which the simple functional form obeys. 35

Their 2013 specification for consumer product choice uses the following approach. The function \( \sigma(a_k, \bar{a}) \) is the salience of an attribute level ((\( a_k \) relative to the average in a choice set, \( \bar{a} \)). There is an ordering assumption that salience increases as the gap between \( a_k \) and \( \bar{a} \) widens; when \( a_k \) increases and \( \bar{a} \) decreases for \( a_k > \bar{a} \), and similarly for \( a_k < \bar{a} \) when \( a_k \) decreases and \( \bar{a} \) increases. Diminishing marginal sensitivity assumes that when a constant is added to both terms, salience falls. They note (p. 808) that these mathematical properties are also well-established empirical properties of basic perceptual psychophysics. Psychophysics is one foundation.

Another useful property is homogeneity of degree zero: Saliency \( \sigma(\alpha a_k, \alpha \bar{a}) = \sigma(a_k, \bar{a}) \) (for \( \alpha > 0 \)). This property is called range-adaptation in psychology. It is useful for economics because it is mathematically equivalent to simple property of price/quality ratios.

A saliency function which satisfies these properties is \( \sigma(a_k, \bar{a}) = \frac{|a_k - \bar{a}|}{a_k + \bar{a} + \theta} \)

The saliency function values are ranked from 1 to S (high to low) giving ranks

---

35Ellis and Masatlioglu 2019 provide a more general derivation and representation theorem. They are transparent and a consumer of the model can judge their plausibility.
Those ranks are transformed into normalized salience weights $\omega_{is}$ that sum to one according to 

$$\omega_{is} = \frac{\delta_{kis}}{\sum_{s'} \delta_{kis'} \pi_{s'}}$$

where $\delta \in (0, 1]$. The boundary $\delta = 1$ corresponds to no salience influence and lower $\delta$ values correspond to more salience influence.

Now the foundations of BGS salience have been presented. What is the most similar microfoundation of SAM? To most economist readers, algorithms like SAM are a “black box” because it is mysterious in how they are micro-founded and how they work intuitively. How do SAM and similar algorithm actually work?

The high-level answer about its microfoundations is that SAM is the latest in a series of algorithms (beginning around 1990s) that strive to reproduce in detail what the human visual system sees. Their foundation is explanation of human vision which reflects a cumulative history of neural evolutionary adaptations designed, across species, to flexibly guide extremely scarce attention to features of the world which have general (bottom-up) relevance. The evolutionary challenges were many, and so bottom-up visual saliency is too complicated to be expressed by psychophysical properties. The microfoundation of SAM is inference of a mathematically-expressible structure (explained in a neural network), that matches what people have evolved to see rapidly and universally.

That is a biologically serious answer about the foundations of SAM. But it is not an especially helpful one for connecting to economic modelling. The BGS saliency algorithm is transparent, in the sense that any student with a little algebra could take a set of numbers and compute their saliency weights from the BGS formula.

In comparison, SAM is unquestionably opaque. It is a neural network with 16 layers in one part (VGG-16) and 50 in another part. Black-box opacity of such neural networks is a common and valid criticism. Responding to this longstanding criticism, the field of “explainable AI” is an extremely active area of research trying to formulate principles and metrics of explainability (Belle and Papantonis, 2020; Hinton et al., 2015; Ras et al., 2018; Arrieta et al., 2020; Fan et al., 2020;
The practical goal of explainable AI is to create more transparent foundations for how neural-network algorithms work, so that users of models who are intelligent but not AI experts—such as judges, physicians, and police chiefs—know more about where predictions come from. This is essentially identical to our interest as economists in SAM foundations.

A popular method for creating explainability is called “feature relevance”. In this method, features of an image are “manually” coded—that is, some painstaking human judgment is required—indipendently of the saliency map. Then linear regression or a similar method measures how well the saliency map values are approximated by a linear or higher-order function of the features.

A concrete example of feature relevance comes from Bose et al. (2020) in their stock price experiment described earlier. They “manually-coded” 31 statistical features of subsets of prices in the stock price charts, taken from several scientific financial economics papers about “technical analysis” of price path features which can be localized in a price chart. These features include average returns over short or long periods, whether local prices are at a chart-high or chart-low, large-magnitude “jumps” in prices, etc.

Each chart is then reduced to a long list of feature values at each location. Because many features co-exist (e.g., positive returns in several periods will be associated with price-h highs at the end of those periods), principal component analysis was used to reduce the set of 31 coded features to 12 components. Then SAM saliency for each chart region is regressed against the 12 components. Table 4 shows partial results from just four components. The middle of the graph shows how heavily different features are weighted (factor loadings) in creating each multi-feature principal component. For example, component 5 is a weighted

---

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 explainability of the content of the resulting neural networks.

Consider feature-relevance in the fruit bundle experiment. One could code each bundle by the number of fruits, composition of types of fruits, the physical size, color, and brightness of fruits, how close each fruit or functions of the set are to central fixation (median, mean, max, min), etc. Then you regress saliency on those features. If the regression $R^2$ is high then SAM saliency is linearly approximated by the features, and their regression weights are their feature relevance.
Table 4—Correlation of relevant feature principal components with SAM saliency

<table>
<thead>
<tr>
<th>Statistical Feature</th>
<th>Component Comp1</th>
<th>Component Comp3</th>
<th>Component Comp10</th>
<th>Component Comp12</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eigenvalue</td>
<td>5.64</td>
<td>1.88</td>
<td>0.98</td>
<td>0.86</td>
</tr>
<tr>
<td>min price</td>
<td>0.47</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>average price</td>
<td>0.46</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>max price</td>
<td>0.45</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>distance from starting price</td>
<td>0.33</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>min price in % of path min price</td>
<td>0.30</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>loss domain</td>
<td>-0.31</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>max return in % of path max return</td>
<td>0.49</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>min return in % of path min return</td>
<td>0.53</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>std. dev. in % of path std. dev.</td>
<td>0.60</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>autocorrelation r, r_{t-1} in % of path autocorrelation</td>
<td>0.9979</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>average return in % of path average return</td>
<td></td>
<td></td>
<td>0.9996</td>
<td></td>
</tr>
<tr>
<td>SAM weight of components</td>
<td>-0.002</td>
<td>-0.010</td>
<td>0.007</td>
<td>0.013</td>
</tr>
<tr>
<td></td>
<td>(.0005)</td>
<td>(.0005)</td>
<td>(.0004)</td>
<td>(.0006)</td>
</tr>
</tbody>
</table>

**Note:**

The table reports rotated factor loadings of the 12 factors with Eigenvalues greater than 0.8. Eigenvalues are listed in the first row of the table. Loadings smaller than 0.25 are blanked out to enhance readability of the table. Statistical features are ordered by their loadings on the respective components, prioritizing components with a larger Eigenvalue. The factor loadings for variables with factor loadings <.25 are omitted.

average of areas of the price chart which have the highest minimum and maximum relative returns, and the highest relative standard deviation. All three factor loadings are around .5. The regression coefficient (standard error) of this factor on saliency is $-0.9 \times 10^{-2} (0.051 \times 10^{-2})$, so those features are associated with lower SAM saliency. Component 12 is a large weight on only one feature, when a graph location has an average return which is very high relative to the entire-graph average. This component is strongly positively correlated (.007, std error=.0004) with bottom-up saliency. The 12 components together explain 20% of the variance in saliency (see Bose et al. 2020). This feature-relevance analysis is one way to explain what the SAM neural net is encoding in price paths.

A related path to explainability is to go back to earlier saliency algorithms. The SAM algorithm was preceded by a series of initial breakthroughs which are more explainable. The first was the Itti-Koch algorithm (see Appendix A.A1). The IK algorithm is fully explainable because you can see from the flowchart how it combines distinct sets of features—color, intensity (brightness) and orientation, and center-surround differences in those three dimensions (e.g. contrast between
bright and dim probably marks an object boundary). A feature-relevance analysis would approximate Itti-Koch saliency well. That algorithm is therefore much more explainable than SAM. Later algorithms added special saliency for faces and centrality (Harel et al., 2007), which are two easily explainable features. Figure 12 shows how Itti-Koch saliency, an intermediate GBVS algorithm (Harel et al., 2007), and SAM vary for the same image.

![Figure 12. Comparisons between three saliency models](image)

*Note:* We show the result of different saliency models on the same image (a). Both the Itti-Koch model and GBVS are fully interpretable (shown in b and d). Also in this example GBVS (Harel et al., 2007) and SAM have very similar results. All three outputs are color-plot in the same standard colormap function in matlab (type: “jet”).

*This is also the standard heatmap code we used for all other heatmap plots in this paper.*

As has been stressed repeatedly, SAM is only intended to predict bottom-up salience which is independent (by the definition of “bottom-up” attention) of an observer’s experience and goals. But how could top-down saliency be modelled? Rather than a general algorithm, future research on this topic for economics is likely be more productive by focusing on one or more types of top-down causes. The reason is that data which mix together perception and behavior from different top-down causes will end up diluting each specific type of top-down attention,
much like swirling different colors together just ends up with black.

Suppose that value is a top-down cause of attention. If value is measured independently, at the individual, stimulus, or single trial level, it is easy to include as a predictor of attention. Indeed, an example of bottom-up attention vs value-driven attention conflict we have already seen is in hider-seeker games. Empirically, the hiders start by looking at high-SAM saliency image features, even though they know their top-down goal is to hide (i.e., low-salience locations are likely to have more value).

A more interesting type of top-down attention that can be analyzed in an economic model style is called “Bayesian surprise” (essentially, novelty). As the name suggests, Bayesian surprise is when a person has prior beliefs about what they expect to see, and a low-prior feature appears, so there is high information gain from prior to posterior updating of feature frequency. Experiments show that when prior beliefs about image content are trained in by exposure or instructed, then features in new images which have high Bayesian surprise grab attention (Itti and Baldi, 2009).

From an economic point of view, Bayesian surprise is an answer– how to model sensory novelty– in search of an economic question. When could this be useful? The computational neuroscience suggests people will top-down attend to information because of pure novelty even when it has low information value. When a fly Bayesian-surprised everybody by landing on a Vice Presidents’ head during a 2020 televised debate, attention was diverted from the important information the VP was conveying. Bayesian-surprise could be used in explaining changes in fashion, why big firms like McDonald’s change advertising slogan with fanfare every so often, “eye-catching” content of advertisements or media coverage, or certain reactions to earnings announcements.

And it was bottom-up salient too, because of the high contrast of black fly on white hair.
B. Where else in economics is visual salience useful?

Salience pops up increasingly often in studies of economic and social choices. In some of these cases, measures or even just conjectures about the details of how salience works visually— and psychobiologically— might improve the economics. Here are some examples. They are all speculative.

- **Tax and price saliency in consumer markets:** Price and value components that are presented to sensory systems, such as explicit price tags the eye can see, seem to receive more decision weight than components that need to be imagined and computed. This was first shown for unit-cost price tags by (Russo 1977) and has been shown in many recent studies (Ott and Andrus, 2000; Hossain and Morgan, 2006; Min Kim and Kachersky, 2006; Finkelstein, 2009; Taubinsky and Rees-Jones, 2017). All of these studies hypothesize differential salience but none actual measure visual salience independently of behavior. That could be done.

- **Disability salience in labor markets:** Laws and court cases in the US, beginning with the ADA in 1990, aimed to reduce discrimination in hiring of disabled people. But there are different types of disabilities. Button et al (2017) compared disabilities which they code as salient— missing limbs, blindness, limited mobility— and those that are covered by ADA but are not coded as salient, such as diabetes, arthritis and mental disorders. They found that an ADA amendment increased hiring, but only for non-salient disabilities. Inference about causality in this case is certainly complicated. However, one plausible conclusion is that the ADA substitutes for visual saliency of hidden disabilities, by enhancing decision-salience for those disabilities. An ideal empirical test would be to use saliency algorithms to predict visual salience of different classes of disabled people, and see if visual salience is associated with differential effects of law on hiring.

- **Nudges and design:** It is likely that at least some successful nudges de-
pend on good visual salience—i.e., salience could be, roughly speaking, a necessary but not sufficient condition for changing behavior. Many, many nudge experiments are being conducted currently. Their effects are often unpredictable (Mehr et al., 2020) and small (DellaVigna and Linos, 2020). Including salience predictions about what nudges people notice might improve results of these, and help design more effective ones.

- **Ethnicity:** Analyses of political economy and ethnic conflict often invoke salience of groups as a causal primitive (e.g., Esteban and Ray (2008). For example, Colussi et al. (2016) shows that heightened salience of minority Muslims during Ramadan influences polarization, voting, and attitudes in Germany 1980-2013. The causal effect of salience could be partly visual, and in principle could be measured and studied to understand causal mechanisms.

Salience is often undefined, but presumed to be identifiable by markers such as social class or ethnicity. Visual salience is likely to play a central empirical role in understanding these differences, in the form of visible differences in clothing, facial structure, names, and rituals, or in auditory differences in accents. Measures of visual, auditory, and other kinds of salience could improve predictive accuracy of these studies and connect them to evidence about how culture is manifested physically. The methods to do so already exist in cultural anthropology, sociology, linguistics, social psychology, and elsewhere.

- **#31 Beliefs:** Besides influencing choices, bottom-up saliency might influence what information is processed and what beliefs result. Padilla et al. (2017) showed a striking example of an effect on beliefs using geospatial plots of hurricane paths. The National Hurricane Center currently shows potential paths with a “cone of uncertainty”, like a 2D confidence interval forecasting a range of areas that a hurricane might conceivably reach several
days ahead. The cone is larger at locations corresponding to more distant days ahead because uncertainty is higher. An alternative visualization is an “ensemble plot” which shows many distinct possible paths and does not draw a cone around them (similar to “spaghetti plots”).

They find that the Itti et al (1998) algorithm predicts that cone plots focus bottom-up attention on the center and boundaries of the cone; those boundaries are further apart in the more distant future. This perception biases judgments of whether the hurricane will grow in storm size and intensity (e.g., wind speed) in the future. Their beliefs reflect a mistake: They think the growing size of the cone predicts that storms will grow in size and intensity. The ensemble plot has a different effect: Predicted saliency is highest at the location where different paths are clustered before they diverge. As a result, there is no size bias of expected hurricane growth in the future, in judgments after seeing the ensemble plot. Participants were more likely to say correctly that forecasters were more uncertain about the future if they saw the ensemble plot. Scientists studying cognitive visualization have many more examples where perception of visualizations create the wrong beliefs (e.g. Padilla et al. 2017, 2018; Itti et al. 1998). Familiar optical illusions work similarly.

C. What’s next?

Cognitive psychologists and neuroscientists know a lot about how attention and salience work (though almost exclusively at short time scales in lab experiments). In economics, by contrast, salience most often refers to hidden variation in likely salience inferred from observable choices (as nicely summarized by Gabaix, 2019). The BGS model is different, because it hypothesizes a particular way in which salience transforms informational inputs, and then predicts decision-salience effects.

Both reduced-form inferred salience approach and BGS predicted salience are
big steps. They will become more and more useful as the list of inferred salience estimates grows from new evidence. However, a deeper understanding of attentional mechanisms could surely be useful to understand differences in inferred salience too. For example, a clever experimental tax variation study (Taubinsky and Rees-Jones, 2017) estimated that inferred salience of undisplayed taxes averaged only 25% of full salience of the actual tax rate (with a very large amount of individual variation). Chetty et al. (2009) estimated the same parameter as 35% for groceries in an experiment and 6% for alcohol in field data. These estimates and others in Gabaix (2018 Table 1) vary a lot from zero to one.

Some variation in these revealed-attention estimates is surely due to differences in methods. But some of that variation could also be due to differences in how attention is actually deployed, which could be measured to try to explain inferred salience variation. It may be, for example, that the alcohol display was more visually salient than groceries, so attention is more distracted from taxes for alcohol than for groceries.

More knowledge of how attention actually works could then help us understand variation in inferred-salience estimates, make better predictions, and suggest causal experiments or inference.

REFERENCES


Crawford, V. P. (2014). 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.


Electronic copy available at: https://ssrn.com/abstract=3308886


Electronic copy available at: https://ssrn.com/abstract=3308886
Appendix

A1. History and details of saliency algorithms

The SAM algorithm takes one image as an input and outputs its predicted saliency map. The saliency map is a saliency value from zero to one (least salient to most salient) assigned to each pixel on an image. We adopted the default parameters from the original approach and applied it to our image dataset. As a result, the saliency predictions we are using to make predictions of location choices in games have no free parameters. Figure 1 in the text is a specific example of the SAM saliency map from one of the pictures we used in our experiments.

A little history of saliency mapping may be useful here to convey how well-founded these algorithms are (and to support further discussion below).

Inspired by a deep understanding of how the human visual system prioritizes attention, a series of progressively improving algorithms were developed to use visual images as inputs, and output predictions about where people will look in the first 1-2 sec of processing (Itti et al., 1998; Harel et al., 2007; Judd et al., 2009).

Figure A1 shows an early algorithm from Itti et al. (1998). These early algorithms used a combination of handcrafted features to extract information about contrast, color, and orientation. Dark-light contrast is special because it marks boundaries between objects. Color and orientation are also thought to have adaptive value in parsing images in ways that are ecologically useful.

These classes of features are called “bottom-up” because they are perceived rapidly, and do not use any abstract information about the meaning or value of what is perceived. “Top-down” perception, in contrast, depends on meaning, personal knowledge and task goals (Frintrop et al., 2010).39

39Like many scientific dichotomies, it is hard to draw a sharp line between bottom-up and top-down processes that guide attention (e.g., Awh et al., 2012). 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.
Consider the stick figure “I”. The bottom-up perception is a black vertical line of a certain length, with slightly extended top and bottom horizontal lines on top of the vertical line, surrounded by contrast with a white background.

Top-down perception adds meaning, making the perception more behaviorally useful. An English speaker will perceive “I” as a marker of first-person communication; a student just learning Roman numerals will perceive “I” as the number one; and an architecture aficionado may perceive it as an Iconic column, a part of a building. All the latter forms of salience use semantic knowledge— which is local, and acculturated— about the world to inform perception of what “I” means and what to do with that information. Top-down perception of one object rather than another tends to be influenced what which features are personally relevant, valued, familiar, and novel.40

40See studies on the effects on perception of recent choice history (Awh et al., 2012), familiarity and
Figure A2. SAM model

Note: presents the deep neural network model structure of the state-of-art SAM model framework.

The SAM algorithm (see Figure A2) we use was tuned using human free gaze data on a large number of images, without any special goals or incentives. The subject are just told to look. These algorithms were not designed to predict active choices in games with specific goals, such as matching and hider-seeker. The matching goal, for instance, is to choose a location another person is also likely to choose. This is a top-down influence on perception which is likely to produce visual fixations that are different from free viewing. Thus, the extent to which SAM can predict the influence of predicted salience is probably a lower bound on how well better models, incorporating top-down goals, will do.

A2. Focality in previous game experiments

There is a substantial, interesting series of experimental studies about focality in matching games. These studies are quite different from our approach but, for that very reason, are important to describe here for comparison.

There was a long lag between Schelling’s early 1960 discussion and later bursts of careful experimentation on focality.

#42 Mehta et al. (1994b) proposed an important contrast between “secondary novelty (Itti and Baldi, 2009), value for consumer goods (Towal et al., 2013), and self-reported “meaning” (Henderson and Hayes, 2017).
salience” and “Schelling salience”. Following Lewis (1969, pp. 24-36), they suggested that when players are not sure what to choose, they choose according to “primary salience”, which is “some (possibly stochastic) process that brings one of the labels to the player’s mind” (p. 660). Secondary salience is the belief about what creates primary salience for others. This process can obviously be iterated further.

Their experiments supported this distinction. In “picking” conditions people just picked an object from a choice set (e.g., a set of flowers). In “matching” conditions their choices were matched with randomly-chosen others and rewarded if they matched. The hypothesis is that picking measures primary salience and matching measures secondary salience. Indeed, the most common modal choice in the picking condition was usually chosen much more often when matching.

Note that this primary-secondary distinction is instantiated naturally in the SCH model (although that model was developed to explain behavior in a much wider range of games). In SCH the ‘process that brings one of the labels to the player’s mind’– its primary salience– is predicted \textit{ex ante} from the bottom-up SAM model. In the Mehta et al. (1994b) paradigms primary salience has to be measured by having people choose objects in the picking condition. Using SAM a primary salience prediction is delivered for all images; no new data or free parameters are needed.

In contrast to primary and secondary salience, an object has “Schelling salience” if it is unique or is chosen by a rule that leads to unambiguous results. Schelling salience need not arise from primary or secondary salience. For example, in a list of historical figures including Adolf Hitler, Hitler could be Schelling-salient even though few people would pick Hitler (primary salience) or think others would pick Hitler (secondary salience). Indeed, Mehta et al. (1994b,a) find evidence for both secondary and Schelling salience in their data.

More ambitiously, Bacharach (1993); Bacharach and Bernasconi (1997) proposed general principles underlying focality in matching choices from sets of ob-
jects, essentially trying to unpack Schelling Salience into component parts. Their idea was that if people know their goal is coordination, they will try to naturally categorize objects into subsets and chose from more distinctive—e.g., smaller—subsets. However, subjects’ actual choices were not always consistent with the most non-obvious of their principles. There experiments are elegant and careful. They were held back by the fact that a key element of the theory—“noticing” set-theoretic features—is measured only crudely (by self-report), whereas we now have eyetracking to measure noticing directly.

Focality is also likely to work differently in hider-seeker games (HS). Studies by Mehta et al. (1994b) Bacharach (1993); Bacharach and Bernasconi (1997) were focused on coordination; at that time in the research history, there was no ambition to create theories of focality that would span games of different competitive structures. Understanding matching was difficult enough.

In a separate strand of cumulated regularity, an early study by Rubinstein et al. (1997) (RTH) used a four-choice hider-seeker game. Their canonical example is a choice between four letters ordered from left-to-right, where one letter is a singleton subset, like so:

A B A A

#19 RTH hypothesized that the left and right A letters are avoided (because of “extremity-aversion”; cf. Bar-Hillel 2015). They hypothesize that the single B is clearly focal because it is both visually and semantically unique; and it will therefore be avoided by hiders. That leaves the second “interior” A from the right, which is least focal when compared to other choices (and therefore uniquely non-focal, giving it an ironic strategic focality).

In these early studies, extremity-aversion and B-focality are simply hypothesized intuitions; they were not guided by data or visual perception principles. On this basis, RTH predicted that the third A would be chosen most often. Indeed, in their experiments the third A is chosen most frequently both by hiders (40%) and seekers (45%). As a result, there is a “seeker advantage” because the seekers win
more often than Nash equilibrium prediction of 25%. However, our replications in Caltech and UCLA subjects found much lower rates of the choice of the third “inner” A, around 29%, closer to the Nash 25% prediction.

Falk et al. (2009) used visual hider-seeker games similar to the four letter choice. One game required choosing 3 cells out of the 25 locations in a 5x5 matrix. They observe both an edge aversion and a seeker advantage. There is a lot of other interesting data and psychology in their paper.

In the main text we noted that our modelling builds upon Crawford and Iriberri (2007a) (hereafter CI), they advanced a novel analysis of games like ABAA, based on level-k modeling. They hypothesized that behavior could be consistent with a level-k approach, in which level-0 behavior is influenced by salience. Salience is parameterized by the frequencies of the outer A’s and the central A’s. Specifically, CI assumed that level-k types only best respond to level k-1 types, and that the population didn’t contain any actual level zero types. Under this framework, they estimated both level zero players’ preferences towards different options (saliency biases) and population frequencies of level types. The general approach fits behavior well. Our paper expands on this approach by predicting saliency independently of choice, using no new data, in location games.

Hargreaves Heap et al. (2014) questioned the strength of the CI conclusions on the grounds that the salience of the extreme A’s and the unique B were estimated parametrically and not constrained across game structures. They created choice sets with a single “oddity” that is visually or semantically unique (e.g. a list of words which are all diseases plus the word “fitness”.) They test whether the oddity is equally salient for level 0 players in three types of games—coordination (matching), discoordination (players win if they both choose something different), and hider-seeker. They reject the hypothesis that level 0 salience is the same across games. Crawford (2014) comments on their paper.

Based on data reported in their paper, the seeking win rate in this experiment is 10.37% while the chance level is only 6.25%, implying a seeker advantage of +4.12%. These numbers are rather close to our own, which are about 7% and 9%, although the paradigms differ a lot.

See Stahl II and Wilson 1994; Nagel 1995 and see Crawford et al. (2013) for a thorough review.
We find better “portability” of salience across matching and hider-seeker games. Specifically, we are able to predict the saliency-choice correspondence in matching games from SCH hider-seeker estimation.

A3. Results from no-feedback trials

The realized matching rates when there is no feedback are as follows. Note that the hider-seeker rate (9%) is the same as trials with feedback:

<table>
<thead>
<tr>
<th></th>
<th>No feedback</th>
<th>Number of observations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nash mixed prediction</td>
<td>0.071</td>
<td></td>
</tr>
<tr>
<td>Matching game</td>
<td>0.35(0.0006)</td>
<td>550</td>
</tr>
<tr>
<td>Hider-seeker game</td>
<td>0.09(0.0002)</td>
<td>523(s)+527(h)</td>
</tr>
</tbody>
</table>

Note: 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).

Figure A3 below is a quantile to quantile plot, plotting the percentage rank of saliency for each location against the percentage rank of choice frequencies for those locations in matching games. To get the Q-Q plot, we first mapped all users’ choice data (not only click points, but all points which fell into the circle) onto a one-dimensional saliency value, normalized from zero to one (The highest saliency point in each entire image is one, and the lowest is zero). Then we ranked all these realized saliency values for all choices in the targeted sub-block. We also transformed the rank of the choice frequencies across all subjects into rank percentages. We plotted the normalized saliency value, which was also the percentage of saliency, against the percentage of points chosen with the same saliency ranking. The Q-Q plot below shows that all quantiles of choice data are above the same quantiles of saliency level, and hence above the diagonal dashed line that would result if people were choosing independently of saliency.

Figure A4 presents both Q-Q plots and density maps in the hiders-seeker game. Figures A4 a-b indicate that seekers’ choices are more biased towards salient locations than hiders’ choices are, and both are much less saliency-biased than
in the matching games (recall Figure A3). Keep in mind, however, that the hiders should be choosing locations as low in salience as they can perceive (i.e., a best-response Q-Q curve would be underneath the 45-degree identity line).

The density maps in Figures A4 c-d take every location in every game, and assign each one a saliency level (0-1 normalized within each image), and computes the frequency with which “strategies” (=locations) were chosen across all games and subjects. For hider-seeker games these should be flat horizontal lines in equilibrium (except for sampling error). However, there are a disproportionate number of choices of high-saliency locations (that is, the densities turn up sharply at the right end of the scale). Seekers choose the highest-saliency locations about three times as often, and hiders choose them about two times as often. There is a slightly disproportionate tendency to choose the lowest saliency locations (near zero at the left end of the scale), especially for hiders.

Figure A3. matching game q-q plot of choice frequency(x-axis) and saliency ranks (y-axis)

Note: The red-diamond point (0.05,0.4) indicates that only 5 percent of choice points were made at the locations at or below 40% salience. Equivalently, 95% of the points fall within the top 60% most salient points. Choices generated by chance would thus correspond to a diagonal line of this plot from (0,0) to (1,1). The maximal accuracy is the blue line: $y = 1$ for all $x > 0$, which would occur only if that all choices fall on exactly the most salient point.

A4. SCH Model comparison with different specifications

#28 In this subsection, we are going to compare four different sub-models. We choose the Bayesian information criterion (BIC) to be the criterion for model
Note: a, b: Q-Q plots for hiding role (a) and seeking role (b). c, d: kernel pdf density map of the choice frequency as a function of location saliency ranks. The x- axis is the rank of the saliency values and the y-axis is the probability density. Note: The kernel is Gaussian. The bandwidth is calculated using the formula: $\sigma \times \frac{1}{\sqrt{2 \pi N}}$}, in which is the standard deviation of the samples and $N$ is the number of observations.

selecting, since it balances the goodness of fit and the possibility of overfitting.

In all cases we restricted the softmax sensitivity parameter $\lambda$ from 0 to 100. Larger values carry little information since $\lambda = 100$ is close to best response. Constraining $\lambda$ also makes it easier to create a bootstrapped confidence interval, which is useful due to the non-smoothness of the target function (likelihood function).

Here are descriptions of models we are going to test (and see Table A.A4):

- **Model 1:** There are only two types of players: 1) naive players who play as level zero players described in the main text. 2) equilibrium players who do pure randomization. Both the proportion of naive players, $p_s$ and $p_h$ serve as free parameters.

- **Model 2:** There is no level zero player in the real population, but higher level types believe there is. The hiders and the seekers have different $\tau$s but the same $\mu$ and $\lambda$. 

Electronic copy available at: https://ssrn.com/abstract=3308886
- Model 3: Same as model 2, except that level zero players exist both in the belief structure and in the population.

- Model 4: This model fits hiding data and seeking data separately using two sets of parameters. Each game has three parameters: $\mu$, $\lambda$, and $\tau$. The best fit model of it dominates the best fit of model 3 since model 3 is a special case of model 4. However, model 4 allows more free parameters, which BIC value will penalize.

- Model 5: This model fits hiding data and seeking data using a common $\mu$, $\lambda$, but uses level-k framework rather than CH, assuming the population consists of players whose level ranging from one to four (no level .5).

- Model 6: This is the same setting as model 5, except it allows level zero types.

Table A2 lists the best fit results of each model. Both BIC and AIC indicate that model 3 is the best performing model. Model 2 performs worst for the reason that without level zero types, the model structure will over predict pure anti-salient hiders, which is the opposite compared to what is seen in the data.

The Level-k Model 6 is almost as accurate by AIC and BIC, and we commented on what can be learned from it in the text. Figure A5 plots predictions of that model and the data, for comparison to Figure 9.

A5. Word version of map coordination task

#35 One way to see how important visual saliency is for coordination is to test how people behave when they face the coordination problem in an non image environment. The SAM algorithm cannot help at such scenario. We tried a non-image version of Schelling’s location game, in which subjects were asked to coordinate on ten locations described in figure x but only in a word format. The

\footnote{see Nagel, 1995; Crawford and Iriberri, 2007a,b}
<table>
<thead>
<tr>
<th>Model</th>
<th>Description</th>
<th>Free parameters</th>
<th>AIC</th>
<th>BIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Level 0+equilibrium $p_s,p_h^{44}$</td>
<td>$[1,.3]$</td>
<td>12716</td>
<td>12728</td>
</tr>
<tr>
<td>2</td>
<td>Role-specific $\tau_x$, f(0)=0 $\mu,\lambda,\tau_s,\tau_h$</td>
<td>$[0.04, 99, .46, .002]$</td>
<td>12780</td>
<td>12803</td>
</tr>
<tr>
<td>3</td>
<td>Role-specific $\tau_x$, f(0)$\neq$0 $\mu,\lambda,\tau_s,\tau_h$</td>
<td>$[0.06, 100, .40, .10]$</td>
<td>12650</td>
<td>12673</td>
</tr>
<tr>
<td>4</td>
<td>Role-specific $\tau_x, \mu_x, \lambda_x$</td>
<td>$[0.01, 90, .40, .07, 90, .50]$</td>
<td>12640</td>
<td>12680</td>
</tr>
<tr>
<td>5</td>
<td>Level-k role-specific f(k), f(0)=0 $\mu,\lambda, f_s(1), f_s(2), f_s(3)$</td>
<td>$[1, 99, .22, 0, .78, .83, .05, .12]$</td>
<td>12680</td>
<td>12738</td>
</tr>
<tr>
<td>6</td>
<td>Level-k role-specific f(k), f(0)$\neq$0 $\mu,\lambda, f_s(0), f_s(1), f_s(2), f_s(3)$</td>
<td>$[0.18, 99, .29, .05, 0, .66, .17, .22, .61, 0]$</td>
<td>12652</td>
<td>12709</td>
</tr>
</tbody>
</table>

Note: Each model in the table is specified in the list before. BIC is defined as $-2\log L + \text{numParam} \times \log(\text{numObs})$ and AIC is $-2\log L + 2\times \text{numParam}$.

options were: house at the bottom of the map, bridge, small house near the pond, house at the top of the map, pond, two houses together, creek, fork in the road, X on the map, and Y on the map. The questions were presented in a randomized order. N=37 people participated the survey on prolific. Each of them only answered the question once. Most of them choose the option “x on the map” (49%) while none of them chooses “y on the map”. Only 5% people choose the bridge, which was the most popular option when the question was presented in an image format.

A6. Fruit Experiment -Data

N = 75 participants took part in this study on Prolific, a European online data collection platform, following a pre-registration process on the Open Science Foundation website (OSF). All the participants were pre-screened to have a prior approval rate of at least 70% based on their previous participation. Each subject
Figure A5. Level-k Model 6 training-testing comparison

Note: The x-axis is the saliency values of all click points. Each point on a graph indicated what percentage of choices were made for locations within images based on the saliency of those locations. (a): choice data and model prediction in the training dataset seeking condition. (b): choice data and model prediction in the training dataset hiding condition. (c): choice data and model prediction in the testing dataset seeking condition. (d): choice data and model prediction in the testing dataset hiding condition. Can be compared to Figure 9 in the text.

was only allowed to participate once for all types of batches (including pilot studies). Participation from mobiles and tablets were not allowed in order to control for attention effects.

A7. Stimuli Property and Selection Mechanism

We took 72 photos of different combinations of real fruits displayed on a dining table. For a better performance of SAM, we did not choose to display the stimuli on a more clean virtual background, like a whiteboard, because the image distributions of real photos are closer to the image distributions of the training sample. In text, Figure 3 showed examples of SAM predictions. Each image contains two piles of fruits and each pile contains three to five fruits. We flipped all the images in the horizontal direction so that we got another 72 images with
Table A3—The choice percentage of all choices in Figure 2

<table>
<thead>
<tr>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>X on the map</td>
</tr>
<tr>
<td>House at the bottom of the map</td>
</tr>
<tr>
<td>Bridge</td>
</tr>
<tr>
<td>Small house near the pond</td>
</tr>
<tr>
<td>House at the top of the map</td>
</tr>
<tr>
<td>Pond</td>
</tr>
<tr>
<td>Two houses together</td>
</tr>
<tr>
<td>Creek</td>
</tr>
<tr>
<td>Fork in the road</td>
</tr>
<tr>
<td>Y on the map</td>
</tr>
</tbody>
</table>

Note: This table represents the percentage of people (N=37) playing the map game we described in Figure 2. Each participant only did it once.

the same content, but with the pile locations flipped. These 144 images consist of our image pool.

We selected 20 images from the image pool and all of the selected images satisfy four conditions:

1) **One-side saliency centered** All of the selected images are strictly one-side saliency centered, which means that the most salient locations only appear in one fruit pile. Figure 3c represents an example of a one-side salient image, while Figure 3a shows an image that is not one-side saliency centered. Formally, consider two sets of pixels constituting the left pile and the right pile, \( P_l \) and \( P_r \). Function \( s \), the saliency model, maps the union of \( P_l \) and \( P_r \) to \([0, 1]\). The most salient location of an image consists of a set of pixels \( S_h : \{x | s(x) > 0.99\} \). An image is one-side saliency centered, if and only if exactly one of the two conditions hold true: \( S_h \cap P_l = \emptyset \) or \( S_h \cap P_r = \emptyset \).

2) **Balanced saliency center locations**: All of the selected images have saliency centers equally located on the left side or right side. Half of the

---

45This procedure is to avoid any left-right biases when taking images. It is done using a matlab function flipimg().

46Since saliency is a relative measure, there will always be at least one pixel with saliency value one.
images have saliency centers on the left and the other half have them on the right.

3) **Balanced valuation distribution**: There are only two types of fruits: oranges and apples. Each apple is worth 1.3 dollars and each orange is worth 2.2 dollars. The total value differences between two piles range from 0.4 dollars to 4 dollars. There are exact 50% of rounds with the more rewarding option located on the left and 50% of rounds with the more rewarding option on the right.

4) **Balanced congruences**: An image will be called “congruent” if the more rewarding option is also the more salient option. Among all images, there are 50% congruent images and 50% incongruent images. No image contains two piles with the same amount of values.

5) **Balanced number of fruits**: Among the total 20 images, in 18 images have the number of fruits only differ by one. The other two images differ by two. 11 images have more fruits on the left and 9 images have more fruits on the right.

### A8. Experimental procedures of location games

#### Screen 1:
You are now going to do a series of short games. In each one, you will see a series of pictures and you must choose a location on the picture by clicking with the mouse.

The rules of each game are slightly different, so read them carefully before you start!(You cannot go back and reread them.)

#### Screen 2:
You’ll start with a few practice items to help you get familiar with the basic set-up.

---

47 We did a pilot experiment with integer unit values. It turns out to be that integer values are too easy for the subjects that we don’t see any variation in choices.
Table A4—Summary of datasets

<table>
<thead>
<tr>
<th>Type of datasets</th>
<th>Platform</th>
<th>Whether no-feedback</th>
<th>N of subjects</th>
<th>Games</th>
<th>Time limit</th>
<th>Between vs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Main Batch</td>
<td>In lab</td>
<td>Yes</td>
<td>29</td>
<td>M,H,S</td>
<td>6s</td>
<td>Within</td>
</tr>
<tr>
<td>Main Batch Online</td>
<td>mTurk</td>
<td>Yes</td>
<td>38</td>
<td>M,H,S</td>
<td>6s</td>
<td>Within</td>
</tr>
<tr>
<td>Big Circle</td>
<td>mTurk</td>
<td>No</td>
<td>67</td>
<td>M,H,S</td>
<td>6s</td>
<td>Within</td>
</tr>
<tr>
<td>High Reward</td>
<td>mTurk</td>
<td>No</td>
<td>29</td>
<td>H,S</td>
<td>6s</td>
<td>Within</td>
</tr>
<tr>
<td>No time-limit</td>
<td>mTurk</td>
<td>No</td>
<td>49</td>
<td>M,H,S</td>
<td>Inf</td>
<td>Within</td>
</tr>
<tr>
<td>Between-Subject</td>
<td>mTurk</td>
<td>No</td>
<td>53</td>
<td>H,S</td>
<td>6s</td>
<td>Between</td>
</tr>
<tr>
<td>Time pressure</td>
<td>mTurk</td>
<td>Yes</td>
<td>31</td>
<td>M,H,S</td>
<td>2s</td>
<td>Within</td>
</tr>
</tbody>
</table>

Note: This table summarized seven different dataset collected in different time. Only the high reward group and the main batch group are the same group of participants. All other batches are completed by a new group of people. Repeated participation is not allowed in all other batches.

Table A5—Location games: dataset usage summary

<table>
<thead>
<tr>
<th>Analysis Names</th>
<th>Dataset Used</th>
<th>Observations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Seeking Win Rates</td>
<td>The main results: in-lab dataset.</td>
<td>M:559,H:529,S:531</td>
</tr>
<tr>
<td>(Seeker’s advantage)</td>
<td>Also reported this percentage for other robustness checks.</td>
<td>M:458,H:441,S:452</td>
</tr>
<tr>
<td>SCH model: training</td>
<td>In-lab dataset with both feedback group and no feedback group.</td>
<td>H:1096,S:1090</td>
</tr>
<tr>
<td>SCH model: testing</td>
<td>In-lab dataset, high reward group.</td>
<td>H=446,S=446</td>
</tr>
<tr>
<td>Choice saliency level analysis</td>
<td>In-lab dataset with both feedback and no feedback group (in footnote and appendix).</td>
<td>M:1139,H:1096,S:1090</td>
</tr>
<tr>
<td>Matching rate/Saliency center</td>
<td>In-lab dataset both feedback group and no-feedback group.</td>
<td>M:1139</td>
</tr>
</tbody>
</table>

Note: This table summarizes the dataset we used for each part of the analysis. We mainly and consistently use the dataset we collected in lab for all the analysis. For the seeker’s advantage part, we also tested different conditions for robustness checks. The “observation” column denotes the total number of observations under each game. M,S,H denotes for matching, seeking and hiding, separately. We omit all the missing data which happens rarely in the in lab sessions and more commonly in online sessions.
**Figure A6. Block Design of The Location Game Experiment**

*Note:* This figure shows the block design of the main experiment (location game). Each participant experienced matching game first, then hiding or seeking game in a randomized order. Under each game, there are two sub-block, the first one is always without feedback and the second one is with feedback.

Use the mouse to click a location anywhere on the picture. Notice that your selection is the entire area within the circle.

You will have 6 seconds to make your selection before the picture disappears. If you do not make a selection within 6 seconds, you will not get credit for that picture.

**Screen before each session depending on games:** Matching: Now you are playing a matching game with several other research participants like you.

For each image, you will play against a randomly selected opponent. If you and your opponent choose the same location in the picture, you both win $x.\text{49}$ If there is any intersection between your location and your opponent’s location, it will count as a “match”.

You won’t find out how much you won in this phase until the end of the game.

As before, you will only have 6 seconds to make your choice for each image.

---

\textsuperscript{49} x changes with games, we pay $0.2, $0.1, $0.4 for a success in matching, hiding and seeking.
A9. Matrix Game

N=56 people played 32 normal form games with different strategic structures: Dominant Solvable Self (DSS), Dominant Solvable Other (DSO), Prisoner’s Dilemma (PD), and Stag Hunt (SH).

In the 32 games, 24 games contain a unique equilibrium (SH has two). Each player is either assigned as a row player or a column player.

Table A6—Evaluation of SAM on price-path experiment and matrix game experiment

<table>
<thead>
<tr>
<th></th>
<th>AUC</th>
<th>CC</th>
</tr>
</thead>
<tbody>
<tr>
<td>SAM domain neutral</td>
<td>0.87</td>
<td>0.78</td>
</tr>
<tr>
<td>SAM vs fixations(games)</td>
<td>0.96</td>
<td>0.47</td>
</tr>
<tr>
<td>SAM vs fixations(price path)</td>
<td>0.81</td>
<td>0.52</td>
</tr>
<tr>
<td>Chance level</td>
<td>0.5</td>
<td>0</td>
</tr>
<tr>
<td>Range</td>
<td>(0,1)</td>
<td>(0,1)</td>
</tr>
</tbody>
</table>

Note: The table reports two common evaluation metrics on SAM algorithm for both the finance price-path experiment in Section III.B and the matrix game experiment in Section V.D. It reports area under the receiver operating characteristics (AUC) and Pearson Correlation Coefficient (CC) (Kummerer et al., 2018). We compare SAM’s performance on the domain-neutral images it was trained on against SAM’s performance against human eye-fixations for matrix games. The results on both metrics show that SAM predict human fixations far better than chance. Furthermore, AUC reaches a better performance than that tested on neutral images. We think it is due to two reasons: simple image sets and de-noise procedures of the data set.

A10. Additional analysis

Table A7—Summary of acronyms

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>SAM</td>
<td>The saliency model we used, Saliency Attentive Model</td>
</tr>
<tr>
<td>CI</td>
<td>Crawford and Irriberri (2007)</td>
</tr>
<tr>
<td>ABAA</td>
<td>Hider- seeker game on these four letters</td>
</tr>
<tr>
<td>CH</td>
<td>Cognitive hierarchy model</td>
</tr>
<tr>
<td>SCH</td>
<td>Saliency Cognitive hierarchy model</td>
</tr>
<tr>
<td>BGS</td>
<td>Bordalo, Gennaioli and Shleifer (BGS) saliency theory</td>
</tr>
</tbody>
</table>

Note: If readers have difficulty keeping track of all the acronyms, this table may help.
Figure A7. Seeking win rates as a function of different saliency levels

Note: 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.