

## Using visual salience in empirical game theory

Xiaomin Li<sup>1</sup> Colin Camerer<sup>1,2</sup>

*Coordination games often have salient “focal points”. In games where choices are locations in images, we test for the effect of salience, predicted a priori using a neuroscience-based algorithm, Concentration of salience is correlated with the rate of matching when players are trying to match ( $r=.64$ ). In hider-seeker games, all players choose salient locations more often, creating a “seeker’s advantage” (seekers win 9% of games). Salience-choice relations are explained by a salience-enhanced cognitive hierarchy model. The novel prediction that time pressure will increase seeker’s advantage, by biasing choices toward salience, is confirmed. Other links to salience in economics are suggested.*

31 December 2018 <sup>1</sup>Div HSS, Caltech. <sup>2</sup>Computational and Neural Systems, Caltech. Support was provided by the Behavioral and Neuroeconomics Discovery Fund (PI Camerer) and NIMH Conte Center P50MH094258. Thanks to audiences at the Caltech Graduate Proseminar, Sloan/NOMIS Conference on Decision and Cognition, IAREP/SABE (Middlesex), Columbia University, Peking University, Elke Weber and Vince Crawford for helpful comments, to Anne Karling for a valuable image, and to Eskil Forsell, Milica Moorman, Gidi Nave and Alec Smith for prior research on this topic.

## 1. Introduction

At the US-Mexico border, “coyotes” traffic people illegally through the desert for money, hoping they can reach “safe houses” and eventually work in the US. The US Customs and Border Patrol (CBP) officers try to intercept coyotes (Urrea 2008). The coyotes and CBP are playing a hider-seeker game.

“In-ground sensors are buried in places known only to the Border Patrol. These sensors are known as Oscars. A Coyote [trafficker] would give his teeth to get hold of this information.... Sometimes the sensors are very cleverly placed—their little antennas stick up in the middle of creosote bushes. Cutters [coyotes] know that saguaros, the signature big cactus of the region, always grow among sheltering shrubs. So a stately old saguaro will not only serve as a signpost for the walkers [immigrants entering illegally], but a landmark for the cutters [coyotes], and the landmark has a scribble of handy bushes around it to hide the wire.”

The tall saguaro cactuses are good landmarks for coyotes remembering a safe route, because they are easily seen from a distance. But the saguaros only grow next to prickly creosote bushes with ground cover, a perfect place to hide Oscar sensors and make them *unsalient*.

In the game between coyotes and CBP, how strategic choices are physically perceived in the environment is an essential part of the game. If the strategy choices (locations) were described verbally or labeled  $S_i$ , and a payoff-isomorphic game was played using only those labels, it is likely that behavior would be different, because visual perception is erased by the relabellings.<sup>1</sup>

---

<sup>1</sup> Another natural hider-seeker game is password choice. Cybersecurity experts all know, and bemoan, the fact that user-assigned passwords are often easy to hack. People use familiar characters such as names, nicknames, initials, or birth dates, presumably to aid their memory (Zviran and Haga 1999). Websites often require at least one digit or special character to make passwords less hackable. However, one study showed that when ending their passwords with one or two digits, 24% of people chose to end in a single digit, 1. Such a tendency creates a strong “seeker’s advantage”. In this case, of course, the seeker’s advantage may exist in equilibrium if users with poor memories know they are choosing hackable passwords.

Schelling (1960) famously promoted the idea that how strategies are labeled—whether by visual salience, language, cultural importance, precedent, or in many other “psychologically prominent” ways— was necessary to predict what people would actually do in certain types of games, especially those involving coordination. Simply knowing an abstract, unlabeled assignment of strategy vectors to valued outcomes, Schelling conjectured, is not enough to predict accurately—and also gives no normative advice about what to do. Furthermore, when the goal is coordination, focality of strategies depends on a *shared* concept of salience. Focal strategies must stand out, and people must know they stand out to game partners.

This paper is about visual salience as a source of focality in games. In general, salience is defined as what makes a feature of a stimulus grab attention quickly. There are many scientific advantages to specifically studying salience of *visual images*. A lot is known about what visual features are immediately visually salient, based on observing steps of visual processing in the brain. It is also easy to do experiments with visual stimuli, and to alter images electronically to change their salience. And because visual perception is fast and requires almost no instruction, people can play hundreds of game-trials in experimental sessions. Visual displays can also be used in a wider variety of experiments on human populations, including children and illiterate adults, and in non-human animals, than semantic descriptions or matrix forms.

We study the effect of visual salience in both coordination games and competitive games. The coordination games are “matching” games. In matching games, players try to choose the same strategy chosen by their partners. If they match, they earn a fixed reward (which is independent of *which* strategy matched.) In the competitive “hider-seeker” game, one player — the “hider”-- earns a payoff if they mismatch by choosing different strategies. The “seeker” earns a payoff if they match strategies.

Our approach to focality in games is novel because we use a precise concept of saliency of image locations from a Saliency Attentive Model (SAM) algorithm developed in computer vision science. We also use eyetracking to examine associations between what people look at, what the SAM algorithm predicts they will look at, and what they choose.

The main goal of our study is to see if there is a generalizable (or “portable”) concept of salience that can be applied to games with different coordination and competition payoff structures. For example, does the way that salience influences coordinated matching help predict how salience influences competitive hide-seeker games? Heap et al (2014) argue that the answer is No, in their data (from a different class of multi-object games). Crawford (2007) offers a rebuttal.

A related portability question is whether salience is useful as a *general* foundation for level-0 behavior. Different level-0 specifications have been used in matrix games, in games with numerical strategies or labeled strategies, and in private information games. Can they all be considered a product of some sort of salience? We’ll return to this question in the Conclusion section 7. We also discuss whether salience used in economic choice domains (such as tax incidence and pricing) might be understood using visual salience concepts.

The next section (2) discusses some classic experimental studies of focality in games. Then in section 3 we introduce the visual image paradigm and algorithms that predict what features are salient. Section 4 describes our experimental design and basic results. In section 5 we present a salience-enhanced cognitive hierarchy model. We measure how well it can explain data, and whether it can portably do cross-game prediction. Section 6 tests whether time pressure increases the influence of salience. We also show how eyetracking data and choice are tied together in a hidden Markov model (HMM). Section 7 concludes by discussing salience more generally in recent game theory and economic applications, and pointing to future research.

## 2. Background

### *a. Bottom-up and top-down salience*

In psychology, attention-getting features that are salient are usefully divided into two categories: Bottom-up and top-down (Frintrop, Rome, and Christensen 2010).

Bottom-up features are based on sensory physical features such as color, contrast, and orientation. Bottom-up feature salience is devoid of personal meaning and independent of task goals. To the extent that human visual systems work the same way all around the world, the attention-grabbing salience of bottom-up features should be universal. As a result, attentional data from many diverse subject pools can establish what is bottom-up.

For example, the stick figure “I” is perceived, bottom-up, as 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. Primates, babies, and adult humans all have the same immediate perception. From the bottom-up view the stimulus “I” is *not* made more or less salient because it is a capital letter, a pronoun for “myself” in English, the Roman numeral one, or reminds you of an iconic column your interior decorator suggested for the front porch of your house. All the latter forms of salience are “top-down”: they use semantic knowledge about the (local, acculturated) world, include what is personally relevant, valued, familiar, and novel.

Bottom-up attention seems to emerge from rapid visual neural processing which is essentially common to humans (excepting disorders) and many other species. Inspired by a deep understanding of visual cortex, a series of progressively improving algorithms have been developed to use visual images as inputs, and output predictions about where people will look in the first 1000msec of

processing (Harel, Koch, and Perona 2007; Itti, Koch, and Niebur 1998; Judd et al. 2009).

However, as the “I” example suggests, many types of top-down meaning and value have been shown to be salient because they also grab attention. Variables that increase top-down salience in perception and choice include prior beliefs, recent choice history (Awh, Belopolsky, and Theeuwes 2012), familiarity and novelty (Itti and Baldi 2009), value for consumer goods (Towal, Mormann, and Koch 2013), and self-reported “meaning” (Henderson and Hayes 2017). Evidence also suggests that top-down attention allocation occurs more slowly than bottom-up attention. The SAM algorithm we will use is tuned using human free gaze data. Since the humans have no special goal to achieve during free viewing, the salience that is being captured algorithmically is probably a combination of bottom-up features, and top-down features (such as familiarity, novelty, special attention to faces, etc.) which grab the attention of the particular subject population.

#### *b. Previous evidence of focality in games*

We begin with Schelling’s (1960) map example. He writes:

“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)

The map (Figure 1a) shows small 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 centrally located across the river. Schelling reports that seven of eight people, whom he asked informally, chose the bridge as the place to rendezvous.

In a proper experiment, N=61 UCLA students incentivized to match chose the bridge 59% of the time (see Figure 1b). Indeed, the SAM saliency map predicts the bridge is highly salient (Figure 1c). However, while SAM does not predict the “x spot” is especially salient, it was chosen by 25% of the subjects.

This example illustrates the value of bottom-up saliency, which explains why the bridge stands out-- it has the most dark/light contrast, and is close to central. The example also illustrates the value of top-down saliency in games: The x is meaningful for those familiar with cultural knowledge about maps like these (“x marks the spot”). But its basic features—two diagonal lines crossing—are nothing special to a bottom-up-guided eye.

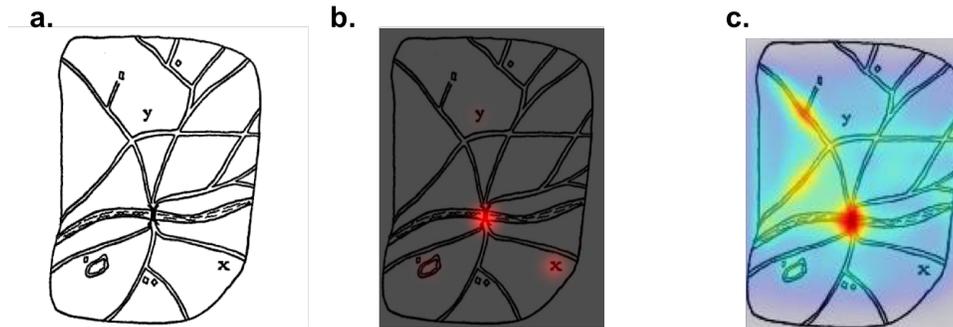


FIGURE 1 SCHELLING'S MAP REVISITED

(a) original map) (b) choice frequencies heatmap, the redness indicates choice frequency (c) SAM algorithm predicted saliency heatmap

There was a long lag between Schelling's early discussion of focality and later bursts of more careful experimentation. (Mehta, Starmer, and Sugden 1994; Bacharach and Bernasconi 1997)

Bacharach and Bernasconi (1997) proposed general principles underlying focality. Their domain was coordination games in which people choose one of

several displayed objects (stars, triangles, etc.) which are not exactly the same but can be categorized into groups. They offer principles governing when objects lumped in a group will be chosen, depending on whether differences are noticed. The empirical problem with their approach is that subjects almost always exhibited the capacity to deem one of the ‘identical’ objects in a group as more focal than the others.<sup>2</sup> Thus, their ambitious theory was missing a detailed method for predictably identifying what makes different choice objects more or less focal.

Focality is interesting in a different way in hider-seeker games (HS). In HS games a hider seeks to mismatch strategy choices and a seeker chooses to match strategy choices. In the simplest games, which we study, payoffs are either zero or a fixed unit prize, which does not vary with the location matches. In these games there is a unique Nash equilibrium: All agents should choose all locations equally often, regardless of whether they are hiding or seeking. However, a bias toward focal locations could lead hidens to get caught more often (a “seeker’s advantage”), or could lead seekers to locations that hidens know to avoid, so that that they catch hidens less frequently (a “hider’s advantage”).

An influential study (Rubinstein, Tversky, and Heller 1997) (RTH) used a four-choice hider-seeker game. Their canonical example is a choice between four letters ordered sequentially from left-to-right:

A    B    A    A

RTH hypothesize that the left and right A’s are avoided (because of “extremity-aversion”; cf. (Bar-Hillel 2015)). They hypothesize that the single B is clearly focal because it is visually and semantically unique; and it will therefore be

---

<sup>2</sup> In a pilot experiment with the five-object example (In an N=62 pilot experiment conducted with Mili Moorman and Alec Smith, 73% chose the triangle and 24% the trapezoid. Only 3% chose the circle-- as Bacharach and Bernasconi’s principle that the singleton polygons are equally focal and hence disqualified-- predicted.)

avoided by hidiers. That leaves the third “interior” A from the left, which is least focal by compared to other choices (and therefore uniquely non-focal). They predicted that the third A would be chosen most often. Indeed, in their experiments the third A is chosen most frequently both by hidiers (40%) and seekers (45%). As a result, there is a “seeker advantage” because the seekers win more often than predicted by Nash equilibrium (which predicts a 25% seeker win rate from random matching).<sup>3</sup>

An important paper by Falk, Falk, and Ayton (2009) collected a lot of data in paradigms similar to the four letter choice. One paradigm consisted of choosing 3 cells out of the 25 locations in a 5x5 matrix. They observe an edge aversion and a very high seeker advantage.<sup>4</sup> In another paradigm people choose strings of five letters, each either A or B. The strings AAAA, BBBB, ABABA (and opposite strings switching A for B) were most popular in matching. The string ABBAB was most popular in hider-seeker. They note that the latter string is similar to one posited as most “subjectively representative” by Kahneman and Tversky (1972), because it nearly balances A and B proportions, and does not have too many long or short streaks.

Crawford and Iriberri (2007a) (hereafter CI) advanced a different analysis of games like ABAA, based on level-k modelling. They hypothesized that behavior could be consistent with a level-k approach in which level-0 behavior is influenced by salience (Stahl II and Wilson 1994; Nagel 1995; Crawford, Costa-Gomes, and Iriberri 2013). Specifically, CI assumed that level k type only best responded to level k-1 types starting from level zero, and that the population didn’t contain any actual level zero types. Under this framework, they estimated

---

<sup>3</sup> Camerer, Moorman and Smith collected pilot data from Caltech and UCLA subjects on a similar AABA game. They observe a slightly lower rate of choice of the inner A (28.6%; 22% for N=59 hidiers and 36% for N=53 seekers) than in most other studies.

<sup>4</sup> 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%, a seeker advantage of +4.12%. These numbers are very close to our own although the paradigms differ dramatically.

both level zero players' preferences towards different options (saliency biases) and population type frequencies. The general approach fits behavioral well. (Our paper expands on this approach by predicting saliency independently of choice, using neuroscientifically-inspired computational algorithms, and with choices of locations in visual images as strategies.)

Hargreaves Heap, Rojo Arjona, and Sugden (2014) questioned the strength of the CI conclusions by testing the portability of level-0 assumptions in the type of level-k theory CI used across different games. They argued that the existing tests for level-0 behavior was not sufficiently helpful because most of them used an *ex post* model-fitting approach. Heap et al used *ex ante* hypothesis testing in games with natural choice sets where one choice is an “oddy” (i.e., semantically distinctive). One set is a series of six black dots and one white dot. Another set is a list of disease words and the word “fitness”. In matching, players strongly coordinate on the oddity. Subjects also played games with coordination, discoordination (players both win if they both choose something different) and hide-seeker payoff structure. Using the oddity as the level 0 specification, they conclude that deterministic level-k predictions do not explain differences in oddity choice rates across the other four games and roles. Crawford (2014) comments on their paper. We did not reach the same conclusion as Heap et al because structural predictions derived from one game type do generalize to a different payoff structure game (see section 6), although we did not study discoordination games as they did.

### *c. What we do that's new*

Our study goes further than previous research in five ways:

- (i) salience is predicted by an underlying computational theory of which features of an information display are salient (the SAM algorithm);
- (ii) salience is also measured independently of both choice and the SAM model (using eye tracking);
- (iii) the salience hypothesized by (i) and measured by (ii) is then correlated with strategic choices;
- (iv) a salience-enhanced cognitive hierarchy (SCH) model is estimated and used to portably explain data in one game based on estimates from another game with different payoff structure;
- (v) differences in the influence of salience are experimentally manipulated (using time pressure) to see whether time pressure causally changes choices. (It does.)

### 3. The Saliency Attentive Model (SAM) algorithm

Visual saliency algorithms are being actively researched. The goal of these algorithms is to predict where neurotypical adult humans will look when they are freely gazing at images. Note, by the way, that these algorithms were **not** designed to predict active choices in matching and hider-seeker games. Such choices are likely to be guided by shared salience (matching) and strategic thinking (HS). The payoff structures create top-down task goals which are *not* part of subject instruction in the data created and used to train the algorithms.

We used a recent state-of-the-art algorithm, Saliency Attentive Model (SAM) to obtain a normalized saliency map for each image (Cornia et al. 2016). The SAM

algorithm is based on a recent machine learning method called a convolutional long short-term memory network (LSTM). The network is trained on the four most popular saliency datasets (SALICON, MIT1003, MIT300, CAT2000) and is trained to best predict actual human gaze patterns recorded from eyetracking. To justify its accuracy, the reported performance of SAM on the website MIT-Saliency is 0.88 using the AUC-Judd measure (Riche et al. 2013), where .5 is chance and 1.0 is perfectly accurate. This accuracy significantly surpasses earlier algorithms, and approaches the accuracy of the best human-to-human benchmark<sup>5</sup> (0.9).

The SAM algorithm takes one image as input and outputs its predicted saliency map. The algorithm assigns a saliency value from zero to one (least salient to most salient) 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 making have no free parameters at all. (However, other parameters will be introduced below to map the link from saliency levels to strategic choice.) Figure 2 is a specific example of the SAM saliency map from one of the pictures we used in our experiments.

---

<sup>5</sup> The best human benchmark indicates how much do two large different sets of human fixation maps correlate. Each set contains many different individuals. Thus, this limitation is due to individual differences of their fixation patterns (Judd, Durand, and Torralba 2012)

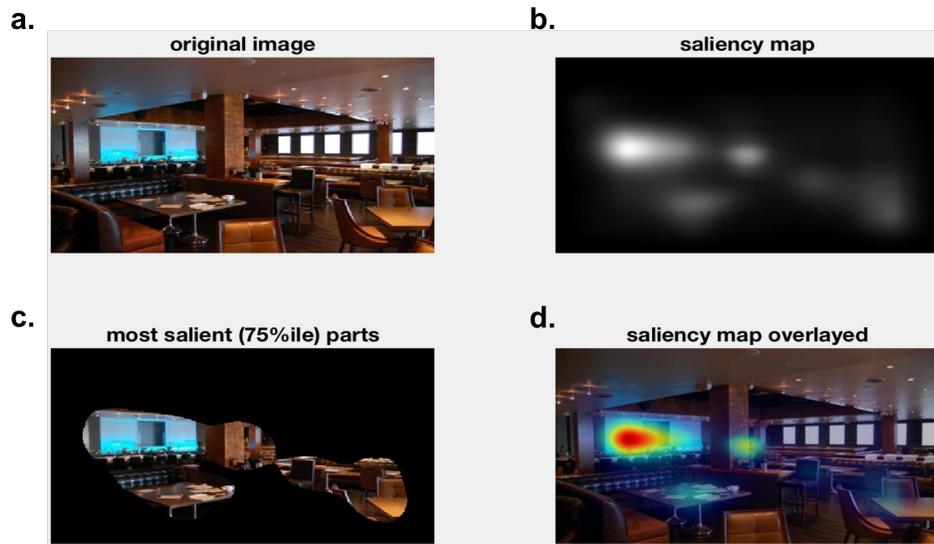


FIGURE 2: A SALIENCY ALGORITHM EXAMPLE.

a: An original image. b: The SAM saliency map, in which the brightness indicates salient level. (the whitest point is the most salient spot). c: The area of original image which is 75% most salient. This area is generated from ranking all saliency values of each pixel and take the ones that fall in first 75%. The other two plots on this figure use the output directly from the SAM algorithm without taking the ranks. d: the original image with a saliency heatmap overlaid onto it (“warmer” red colors indicate higher saliency).

#### 4. Experimental Procedure

The strategy space of this task consists of all the pixels in each image. This large range gives us large variation in saliency, and saliency can be studied directly with eye-tracking.

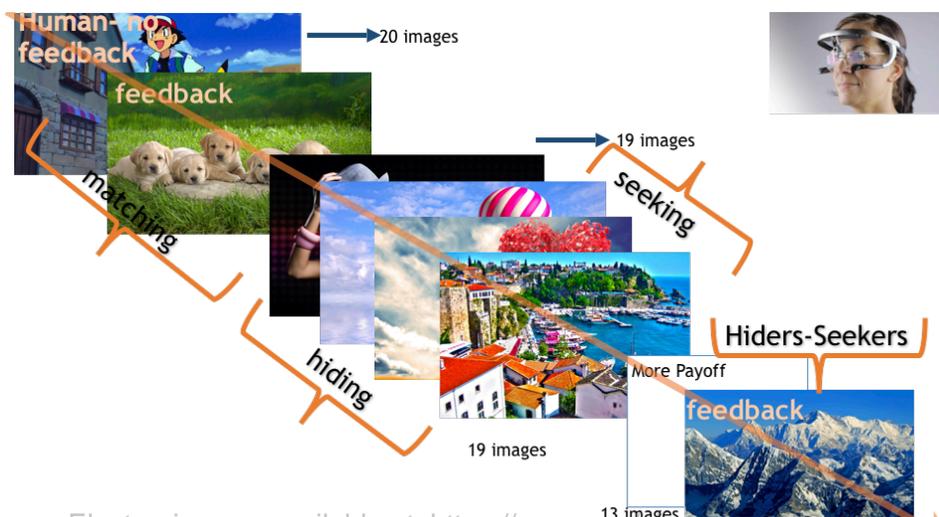
**Participants:** This study was pre-registered on the Open Science Framework (<https://osf.io/yuqjg/>) during data collection and before analysis process. The sample size was pre-determined before the data collection process, based on a pilot study carried out in March 2017. N= 29 subjects participated (13 males, 16 females) one at a time, in a small testing room. An additional N=69 subjects participated online through Amazon mechanical turk (“mTurk”). Among people who participated in the lab, N = 24 had their gaze data recorded. Fifteen of those

subjects belonged to Caltech community and 14 were from the neighboring community. None of them had a college major or work experience within economics or computer vision field. These subjects were tested individually in our lab testing room, and all gave informed consent to participate in research.

### *Experimental design*

The experiment consisted of three blocks of games: matching, hide-and-seek as seeker, and hide-and-seek as hider (Figure 3). The matching block always came first, followed by the hide and seek blocks in randomized order. During each block (game), subjects experienced two sub-blocks: “feedback” sub-block in which the results are revealed right after and “no feedback” sub-block in which no results are revealed. Matching game consisted of 20 images for each sub-block and the hider-seeker game shared the same set of 19 images for each sub-block. There was unlimited time to read instructions and break but only 6s to make a choice for one trial. The results shown in the feedback condition were drawn from previously tested actual subjects (different subjects for each image shown, of which participants were informed).

Since the image fills the entire screen, if the circle hits any boundary of the screen, the rest of the circle extended from (i.e., wrapped around) the other side of the screen to avoid any boundary advantage.



One image here stands for 20 or 19 shuffled images depending on games in each sub-block. In each block, the order of sub-blocks was also shuffled. The matching block always came first and the other two blocks came in randomized order. Images within each block were presented in randomized order.

Two choices were considered a match if two circles centered at their choice spots (with radius of 108 pixels) overlapped, regardless of the extent of overlap. (We use the word “match” for all three blocks, obviously corresponding to winning for seeker and losing for hider in the hider-seeker games.) Figure 4 shows examples of result screens that subjects saw during the experiment. The local group and online groups both participated in the same experimental design.

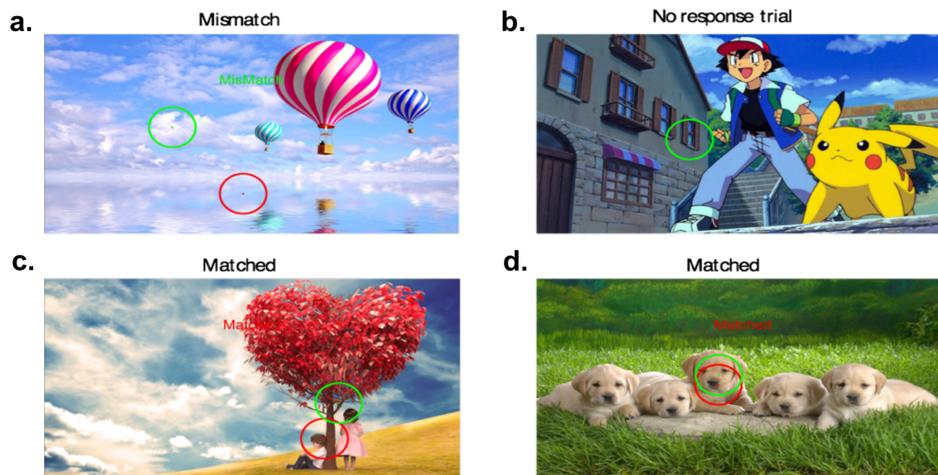


FIGURE 4: EXAMPLES OF TRIAL OUTCOMES.

(a) an example of mismatch (the green and red circles are not overlapping). (b) a trial in the no-feedback condition, with the green circle showing the participant's choice (c,d) two examples in which choices were successfully matched (because the red and green circles overlapped).

### *Stimuli*

Ninety-two (92) colored visual images were displayed on an eye tracking monitor (1920x1080 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 high entropy images and low entropy<sup>6</sup> images. High entropy images tend to have only one central object while images in low entropy images often have many different salient locations (Judd, et. al 2009).

### *Eye Tracking Procedure*

Subjects sat comfortably about 50cm in front of the screen while their eye movements were monitored with a Tobii TX300 eye-tracker with a sampling rate of 300Hz. Presentation of the game used the Matlab Psychophysics tool box. A nine-point calibration was performed in the beginning of the experiment. Before each sub-block, there was a drift correction phase in which subjects need to focus on a dot in the center of the screen. The experiment only proceeded, after drift correction, if the gaze detection was still accurate. Otherwise, the eye-tracker would automatically enter a second calibration session then continue. Between presentation of sequential images, a fixation cross was presented on a gray screen for one second to wash out any remaining gaze trace from the last image.

## **5. Analysis and Results**

### *a. Equilibrium analysis*

Equilibrium analysis generates a statistical benchmark for what people might do.<sup>7</sup> The structure of our games deviates from previous matching games, because

---

<sup>6</sup> Entropy is defined by  $-\sum p \times \log_2 p$ , where  $p$  is the (0-1) normalized salience of each pixel, and the summation is taken over all pixels on an image. Higher numbers indicate more concentration of salience (entropy is minimized when all values of  $p$  are equal). It can be used to characterize the complexity of the image. E.g., a picture with some content on it has less entropy than a white blank paper.

<sup>7</sup> An important mathematical touchstone is “correlated equilibria”. (Aumann 1974). A correlated equilibrium occurs when both players see private signals—which may be correlated between the players—and condition their strategies on their signals and their knowledge of correlation. Stop signs and green-yellow-red traffic lights, for example, act as

locations in x-y space are being chosen. Players click on a pixel but their choice circle is a 180-pixel circle around their chosen location (an area about the size of a nickel coin). Matching is defined by circle overlap.<sup>8</sup>

For a matching coordination game, it is useful to think about both pure strategy and mixed equilibria. Any two pixels (response clicks) chosen by the two players, that fall within the diameter of the circle constitutes a pure strategy Nash equilibrium. One image contains about two million (1920\*1080 pixels) points. Any match is a pure equilibrium, so there are an enormous number of equilibria. In addition, there are many equilibria in which players mix equally across a subset of different locations. For example, if both players choose equally randomly among all pixels, that mixture is also an equilibrium.

For the hider-seeker game, there is a unique Nash equilibrium in which all locations are chosen equally often. 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  $p>.5$ ) 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  $1/n$  chance of being chosen (if there are  $n$  locations) by an iterated logic.

The proof that equal randomization over all strategies is the unique hider-seeker equilibrium puts logic ahead of bio-logic. The last thing the brain wants to do is to choose many objects that are different equally often. The human perceptual

---

correlating devices to create a commonly-observed visual signal which coordinate traffic and reduce accidents. In these terms, our study is about whether visual salience of image locations works as a correlating device.

<sup>8</sup> The mapping from pixels to choice circles means that the size of the circle around the chosen pixel will influence random matching rates, increasing and decreasing them for larger and smaller circles. This is a feature that could be exploited in future research.

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 may be naturally bad at choosing what is *unsalient*.<sup>9</sup>

### *b. Matching games*

To analyze the behavioral data, we first tested 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 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 hypothesis of randomness for all treatment condition sub-blocks ( $p < 10^{-4}$ ).

To get a direct view of how salience affects choices, we plot the choices from all the subjects on two different specific images, shown in the top and bottom rows of the left column of Figure 5. The saliency heat map (middle column) uses redder colors to show more salient locations from the SAM algorithm. The right column shows, in redscales, frequencies of subjects' location choices in matching “human feedback” (top) and “human no-feedback” (bottom) sub-blocks on a

---

<sup>9</sup> A similar conflict between logic and biology occurs in the games “rock, paper, scissors” and matching pennies (e.g., Belot, Crawford, and Heyes 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).

single grayscale image. The predicted saliency in the middle column and the observed choice maps in the right column are highly overlapping.

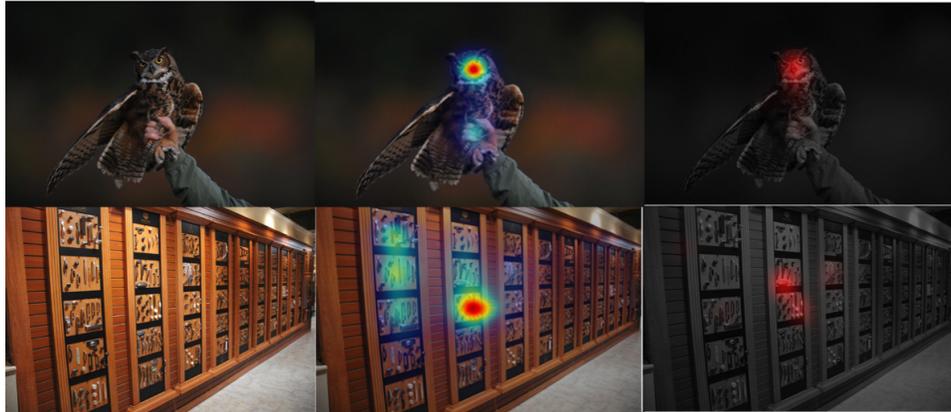


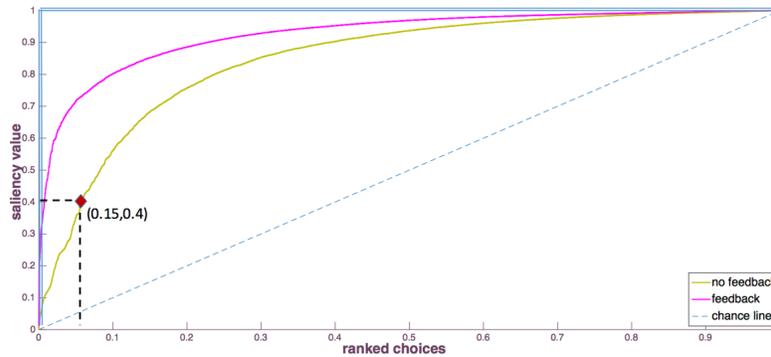
FIGURE 5: MATCHING GAME EXAMPLE DATA.

(Left column) Two original images in top and bottom rows. (Middle column) The original images overlaid by SAM saliency heat map. (Right column) grayscale original image overlaid with the actual empirical choice distributions. The upper image is a sample from “feedback sub-session”, the lower image is a sample from “no-feedback sub-session”. The redness of the locations indicates more choice clicks around that location.

Statistically, the mean of the choice saliency level in the coordination game “no-feedback” sub-block is 0.87, which is far above chance (0.5) ( $p$ -value  $< 10^{-4}$ , 95% confidence interval (CI): (0.86,0.88),  $t$ -test). In the “feedback” sub-block, this level increases further to 0.95 ( $p$ -value  $< 10^{-4}$ , one-tailed paired- $t$  test against “no-feedback” sub-block).

Figure 6 below is a Q-Q plot, plotting the percentage rank of saliency for each location against the percentage rank of choice frequencies for those locations. 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 6:** MATCHING GAME Q-Q PLOT OF QUANTILES OF CHOICE FREQUENCY (X-AXIS) AND QUANTILES OF SALIENCY RANKS (Y-AXIS).

The red-diamond point (0.15,0.4) indicates that only 15 percent of choice points were made at the locations at or below 40% saliency. Equivalently, 85% 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 means that all choices fall on exactly the most salient point.

### *c. Hider-seeker games*

For the hider-seeker game, we present two panels of plots for hidiers (Figure 7) and seekers (Figure 8) separately for two specific example images, using the same display as in Figure 5. Figures 7-8 show that subjects' choices (right column) are much more spread out in both hiding and seeking compared to the concentrated saliency maps (middle column), than in the matching coordination games where saliency and choice strongly overlapped.

In Figure 7 the peak of the choice distribution no longer falls in the most salient area. However, in this example Figure 8 seeking game, the most salient locations are still chosen frequently by seekers (suggesting that, in this example, seekers are not being very strategic).

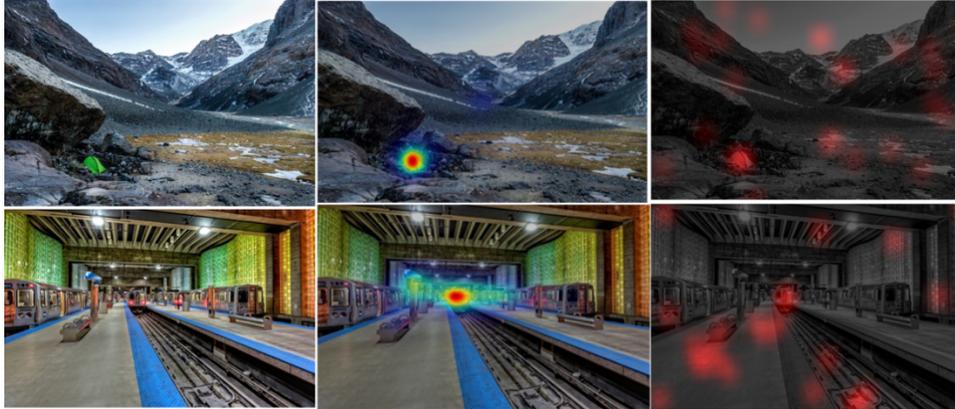


FIGURE 7: HIDING GAME IMAGES, SALIENCY, AND CHOICE

Left column: original images. Middle column: original image overlaid by saliency heat map. Right column: grayscale original image overlaid by user choice empirical distribution (the upper image is a sample from “feedback sub-session”, the lower image is a sample from “no-feedback sub-session”).

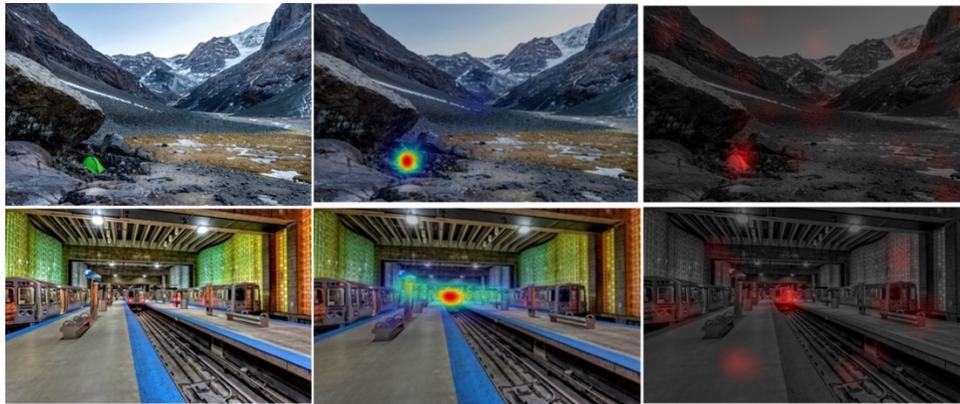


FIGURE 8: SEEKING GAME IMAGES, SALIENCY, AND CHOICE

Left column: original images. Middle column: original image overlaid by saliency heat map. Right column: grayscale original image overlaid by user choice empirical distribution (the upper image is a sample from “feedback sub-session”, the lower image is a sample from “no-feedback sub-session”).

The general statistics are consistent with the direction of effects in these two examples. In the hiding game, the mean saliency level of user click points in no-

feedback and feedback is 0.48<sup>10</sup> and 0.53, close to chance level<sup>11</sup>. The corresponding statistics in seeking game are 0.62<sup>12</sup> and 0.61<sup>13</sup>. A replication sample (N = 29) with ten times higher payoffs yields very similar results, 0.51 and 0.64 for hiders and seekers (feedback only)<sup>14</sup>.

Comparing the results between hiders and seekers, it is statistically significant that people chose salient locations more in the seeking game (any sub-block) than in the hiding game (any sub-block) (one-tailed paired t-test,  $p < 10^{-4}$ ), which reflects what showed in figure 7 and figure 8.

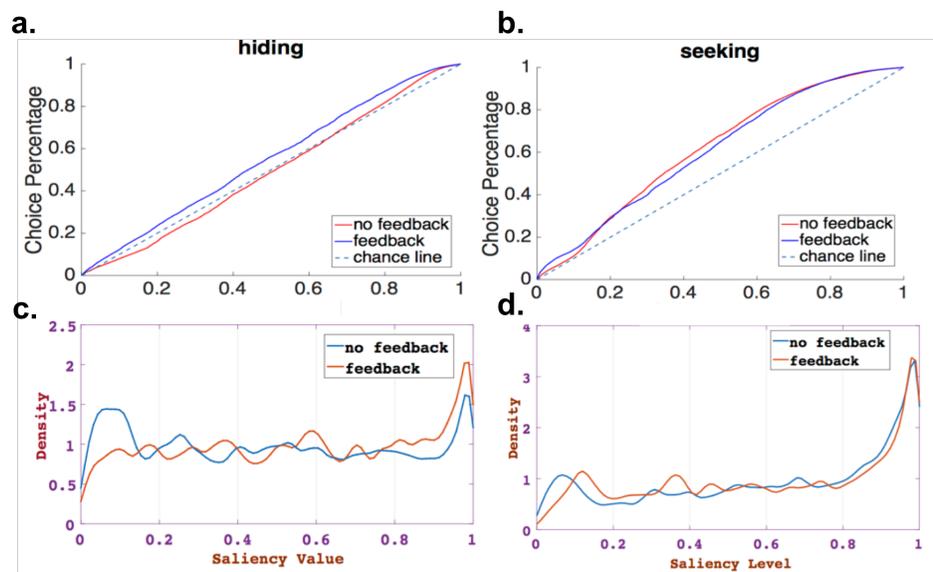


FIGURE 9: STRATEGIES SUBJECTS USED FOR HIDE AND SEEK GAME

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 (plotted

<sup>10</sup> p-value = 0.1, one-sample t-test, CI: (0.46,0.50)

<sup>11</sup> p-value = 0.02, one-sample t-test CI: (0.51,0.56)

<sup>12</sup> p-value  $< 10^{-4}$ , one-sample t-test, CI: (0.59,0.64)

<sup>13</sup> mean = 0.61, p-value = 0.43, paired t-test against no-feedback sub-block

<sup>14</sup> Hiding: p-value = 0.59, CI: (0.48,0.54), seeking: p-value  $< 10^{-4}$

separately for feedback and no feedback sessions). Note: The kernel is Gaussian. The bandwidth is calculated using the formula:  $\sigma \times \frac{4}{3N}^{0.2}$ , in which  $\sigma$  is the standard deviation of the samples and  $N$  is the number of observations.

Figure 9 presents both Q-Q plots and density maps in the hiders-seeker game. Figures 9 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 6). 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 9 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 (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.

### *c. Seekers' advantage*

In this section, we report the realized matching rate in all conditions and compare it to the chance outcome. The theoretical value for two circles to match from two random players is four times the area (distance within a larger circle of two times the diameter of the original circle) of the circle used in the experiment over the entire area (1920\*1080). That frequency is 0.071.

Table 1 presents the realized matching probability in a specific sub-block. All values were calculated from the average of 500 iterations of randomly matching two data points from the dataset if two subjects were in the same sub-block, same image.

TABLE 1: REALIZED MATCHING RATES

	Number of Observations	No feedback	Feedback
Nash mixed prediction		0.071	0.071
Matching game	1147	0.35 (4e-4)	0.64 (4e-4)
Hider-seeker game	1090	0.09 (3e-4)	0.09 (3e-4)
Hider-seeker game high payoff	462	----	0.09 (6e-4)

Statistical tests against null hypothesis that seeking win rate is the baseline level and choices are independently and identically distributed across subjects (which is the Nash benchmark prediction)<sup>15</sup>.

In the higher payoff session, subjects received 10 times the payment received in the previous session hider-seeker game.

As in many other studies of matching, there is quite successful coordination in the matching game based on salience: A successful match happened five to ten times more frequently than in the random-matching benchmark. In the hider-seeker game, there is a significant seeker's advantage across all conditions we tested (as seen in most other studies e.g. RTH, 1994, Crawford and Iriberri, 2007). This advantage does not disappear or diminish in the high-payoff session when subjects could earn ten times more.

Note that while a 9% win-rate for seekers may not seem much larger than 7%, the test assumes the null hypothesis that 7% is the matching rate which should, in the Nash equilibrium, be identically distributed for all periods and all people. Since there are very many observations when pooled together, when

---

<sup>15</sup> 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.

independence is assumed the test has enough power to establish that 9% is significantly higher than 7%. Two more conservative approaches are to compare the actual and predicted hider-seeker match rates across images (averaging all data within-image) and across people. These tests allow dependence of trials within an image, and within a person. The results indicate that seeker's advantage still remain significant using either pooling methods<sup>16</sup>.

## 6. A Saliency-perturbed Cognitive Hierarchy Model (SCH)

In this section we describe a parametric behavioral model meant to explain the main patterns in choices and their salience-sensitivity. Following Crawford and Iriberri (2007a), it uses step-level thinking. The level- $k$  model was introduced by Stahl and Wilson (1994) and Nagel (1995). Camerer, Ho, and Chong (2004) later proposed a modified cognitive hierarchy (CH) version with a slightly different specification about what each level type believes about others. See Crawford, Costa-Gomes, and Iriberri (2013) for a thorough review.

To explain the behavioral findings, we developed a model that combines cognitive hierarchy, a quantal response function (softmax) and a saliency-perturbed level 0 assumption.

### *Model description*

The population consists of different levels of players starting from level zero and with level  $k$  players with frequency  $f(k)$  (assuming to be Poisson distributed with parameter  $\tau$ ).

---

<sup>16</sup> Pooling across individuals (N=29) gives matching statistics for each of the individual image and tests those against the .071 benchmark. The p-values are 0.002 and 0.0005 for no-feedback and feedback conditions. Pooling across images (N=19) for each subject gives p-values of 0.002 for both cases.

For all levels of players, we allow some degree of randomness and mistakes due to the large space of choices, which will be described using a normal 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 specification the nonstrategic level zero players weakly prefer salient choices. Preference is incorporated by a kind of subjective value assigned to saliency of choices. Formally, the probability of choosing point  $n$  depends on the saliency value<sup>17</sup>  $S_n$  in a softmax format with a common softmax parameter<sup>18</sup>  $\lambda$  and saliency perturbation rate  $\mu$  :

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

Other than level zero players, all other levels of players behave in the same way as in CHC. Level  $k$  players assume that all other players are only of lower levels (0 to  $k - 1$ ) and softmax respond to the expected behaviors of such an imagined opponent group. Therefore, the probability for player  $i$  choosing option  $n$  is:

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

In the equation above,  $EU_{kn}$  stands for the standard expected monetary payoff of the level  $k$  players by choosing the point  $n$ :

---

<sup>17</sup> We will use the direct saliency measure from the algorithm (raw value) for the modeling input instead of normalized ones (saliency ranks). The difference is that saliency ranks distributed linearly among points while the raw value distributed closer to exponential distribution. In the Q-Q plot, normalized values makes the results comparable with chance diagonal line since rank value is linearly distributed.

<sup>18</sup> This parameter will be used for the softmax function for all level players, instead of using level-specific ones. Many variants of these models are possible (e.g., Rogers, Palfrey, and Camerer 2009)

$$EU_{kn}^s = R^s P_{\{0,1\dots k-1\}n}^h, EU_{kn}^h = R^h (1 - P_{\{0,1\dots k-1\}n}^s)$$

The superscripts s and h stands for roles in the hider-seeker game. R is the monetary reward value for each game.  $P_{\{0,1\dots k-1\}n}^h$  and  $P_{\{0,1\dots k-1\}n}^s$  is a mixed probability of level zero to k-1 choosing choice n calculated by linear combining their strategies using a normalized Poisson distribution.

Note well that saliency *only* enters into the responses of level-0 players. The challenge is to see whether saliency which only *directly* influences the lowest-level 0 players then has an amplified effect on higher-level thinkers, through their beliefs and then through their choices (a la Mehta et al's (1994a) "secondary salience" concept).

This model has three free parameters  $\mu, \lambda, \tau$  to fit. The parameters capture the preference for saliency of level zero players  $\mu$ , the degree of response smoothing in softmax  $\lambda$ , and the average levels of strategic thinking in the population  $\tau$ , respectively. We constrained the softmax parameter  $\lambda$ <sup>19</sup> and the saliency weight  $\mu$  to be the same between hiding and seeking but allowed a difference in  $\tau$ , to see whether average thinking levels might differ according to roles. We fitted the model with the data in the hider-seeker game, using the hiding game and seeking game together for both sub-sessions ("feedback", "no-feedback"), using maximum likelihood. To have enough data for training, we used the entire first sessions as our training sets and the second sessions as validation test sets. The best fitted parameters and measures of fit are shown in Table 2.

---

<sup>19</sup> The range of  $\lambda$  is allowed from 0 to 100. Larger  $\lambda$  does not carry more information since the exponential function makes the predicted action approaching best response fast. While allowing  $\lambda$  to be free more globally makes it impossible to do the bootstrapped confidence interval due to the non-smoothness feature of the target function. More discussions about the feature of the target function are included in Appendix.

TABLE 2: BEST-FIT SET OF PARAMETERS

	$\lambda = 100$	$\mu = 0.06$	Hider $\tau_h = 0.4$	Seeker $\tau_s = 0.1$
Best fitting				
Number of observation			1096	1090
Likelihood training			-5.78	
Likelihood testing			-5.86	
AIC			11.57	
BIC			42.04	
95% Confidence interval	[72.3,100.0]	[0.05,0.08]	[0.32,0.47]	[0.08,0.13]

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

The best-fit values of  $\tau$ 's, 0.4 and 0.1, indicate that most people did not do many iterative thinking steps when they made their choice. These low values of  $\tau$  match with some previous findings (Fudenberg and Liang 2018). The estimation of  $\tau$  in Camerer, Ho and Chong (2004) and many other papers is often higher, around  $\tau = 1.5$  (see also Crawford et al 2013).

We think the low values of  $\tau$  estimated in our data might results from a tradeoff between the simplicity of visual choice games and the ability to identify all parameters in an SCH model. The best way to identify  $\tau$ , for example, is with games where different level types choose distinct strategies in a way that is carefully designed to separate them (see Stahl and Wilson 1995; Nagel 1995; Costa-Gomes and Crawford 2006; Kneeland 2015; Fragiadiakis, Kovaliukaite, and Arjona 2017; Costa-Gomes and Crawford 2006). In our visual choice games there are usually a small number of salient locations; most players choose those or (if hiding, for example) approximately randomize over other locations. This makes it hard to precisely identify the weight on salience, the response sensitivity, and the levels separately. However, it is notable that the overall likelihood is -5.78 in the training set and only slightly worse, -5.86, in the test set. This indicates that

we are not overfitting by very much (since otherwise the decrease from training to test fit would be much larger).

Figure 10 shows the comparison between real density maps versus model predicted density maps for the hider-seeker game in training (Figs 10ab) and out-of-sample test data (Figs 10cd). From the data distributions (blue lines), it can be seen that there was a sharp density increase starting around 0.9 saliency level for both games as well as a minor decreasing trend from the lowest saliency to 0.9 saliency level for hidiers but increasing trend for seekers. The CH model was able to capture both of these features of the data.

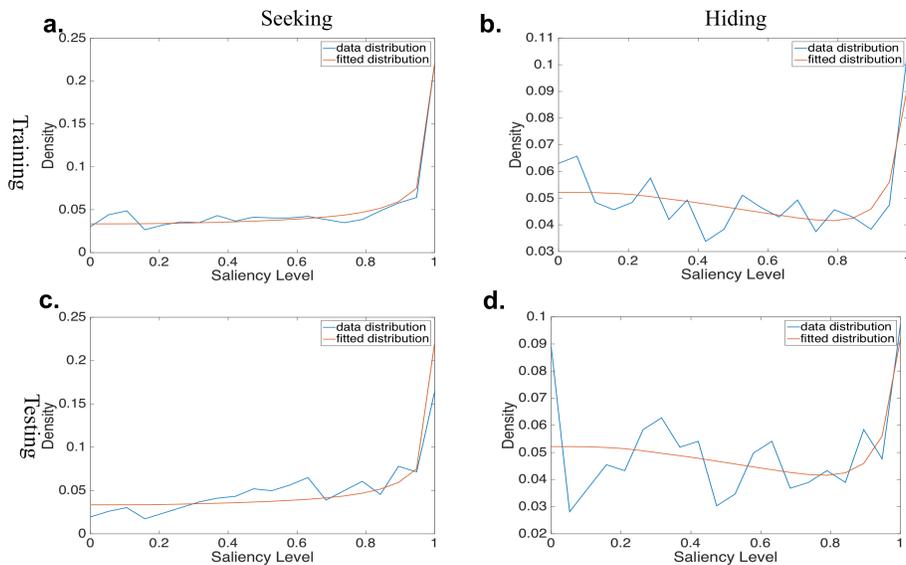


FIGURE 10: COMPARING FREQUENCY OF CHOICE BY SALIENCY LEVEL WITH MODEL FITTED DISTRIBUTIONS.

The x-axis is the saliency values of all click points (normalized to range from 0 to 1 in each image). 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.

According to the best-fit  $\tau$ 's, there existed a large percentage of level zero population for both hidiers and seekers, whose behaviors were only dominated by

saliency value. Besides the actual level zero players' effect on population behavior, such saliency biases from level zero players also affected higher level players' strategies when they were forming beliefs on lower levels.

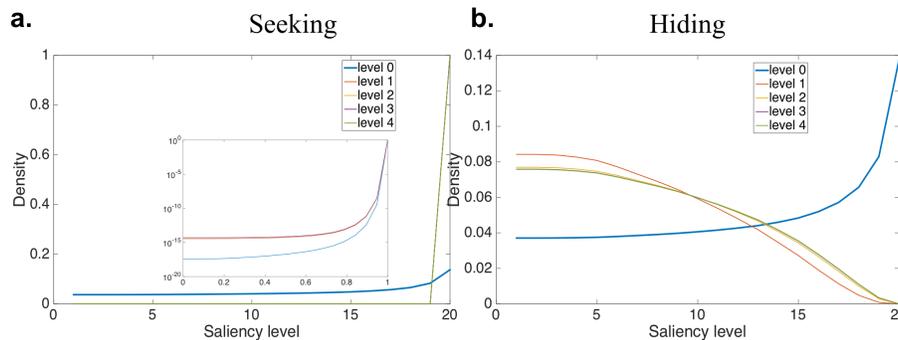


FIGURE 11: STRATEGY DISTRIBUTIONS OF DIFFERENT LEVEL PLAYERS PREDICTED BY THE MODEL

The blue curves on both figures correspond to level zero strategies. In the seeker's case (a), players at level 1 and above highly prefer salient locations. In the larger graph, all levels 1-5 are plotted as green. The inset graph shows there is some variation in sensitivity of level 1 types and above, which can only be seen with a finer-grained y-axis of density. In the hider case (b), level 0 players choose strategies with increase with saliency, and all higher-level types choose less salient locations more often.

Based on the best-fit saliency weight value  $\mu$  (0.06), the perturbation towards saliency is not that large. It only changes the probability of choosing in the top 5% salient area to 0.14 from the even chance level 0.05 (as blue curves show in Figure 11). Though the hiders and the seekers share the same level zero strategy (by assumption), the strategies of higher levels ( $k \geq 1$ ) are affected in different ways: hiders prefer unsalient places while seekers choose the most salient locations more often. Such mixed choice distributions of both level zero and higher-level players are able to explain the trend differences in Figure 10 —the seekers' aggregated choice distribution is monotonically increasing in saliency while the hiders' aggregated choice distribution first has a small decrease, then has a sharp increase at the maximal saliency.

*Cross-game predictive validation:*

To further test the generalizability of the model, we use parameters estimated from fitting the SCH model to hider-seeker data to then predict behavior in the matching game. There is no guarantee that this cross-game portability will work well. Identification of the saliency weight  $\mu$  in hider-seeker comes from the level 0's and from higher-level types who sometimes try to match the level 0 opponents and sometimes mismatch. In the matching games, all higher-level types are guided by salience since they are trying to match the level 0's. The strength of salience-sensitivity that is estimated in the two cases could easily be different (which is precisely Heap et al's (2014) point).

The left panel in Figure 13 shows the comparison between the experimental data distribution from matching games, and the model-predicted distribution using the best-fit parameters from the hider-seeker game. The fit is visually reasonable, except that the prediction clearly underestimates the big jump in empirical choice frequency at the highest saliency level.

The right panel of Figure 12 presents the fitting result by letting  $\tau$  be a free parameter maximizing fit in the matching data while fixing the other two parameters from hider-seeker result. The best-fit  $\tau$  is 0.87, which is higher than the values estimated in hider-seeker game.

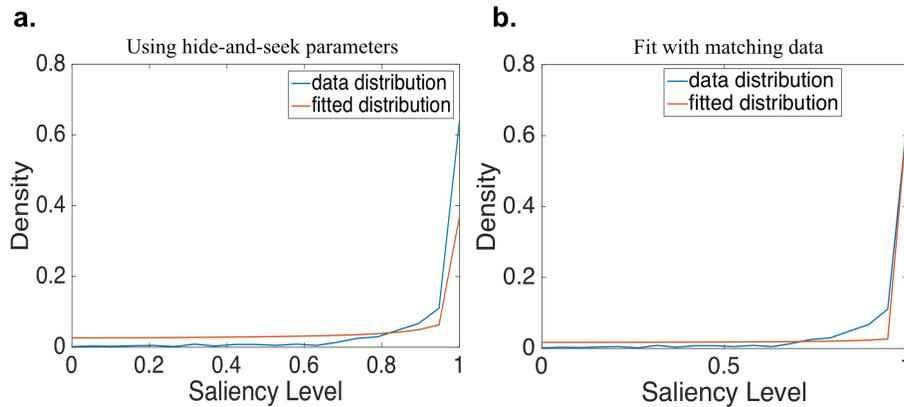


FIGURE 12: THE SCH MODEL CALIBRATED ON HIDER-SEEKER GAME DATA CAN PREDICT MATCHING GAME CHOICES.

The comparison between the matching data distribution and the fitted matching game distribution. (a) Uses all parameter estimates from the hider-seeker game to predict matching game results. BIC: 24.83 (b) allows  $\tau$  to be a free parameter in the matching game (fixing  $\mu$  and  $\lambda$  to the estimates from hider-seeker games) in order to test where there is any change in  $\tau$  between games. BIC: 24.51

## 6. Further results

### *a. How predictable is the matching rate across images?*

An additional hypothesis based on SCH is that the differences in the matching rate across images should be affected by their saliency distributions. When saliency is highly concentrated then the rate of choosing the same pixels (and hence matching) should also be highly concentrated in matching games. Specifically, we hypothesized that the number of local maxima<sup>20</sup> for the saliency distribution on an image should be negatively correlated with the matching rate in matching games. For example, most people should prefer to play the matching game on a white paper with only one star in the center (a single peak of the 2D

---

<sup>20</sup>This number is calculated using function `imregionalmax()` and a Gaussian filter, more details see Appendix Method. For all images, the number of peaks ranges from one to five. We also tested whether entropy of an image is correlated with its matching rate. There is a correlation but it is weaker and sensitive to outliers ( $p = 0.18$  for no feedback and  $p = 0.36$  for feedback session)

saliency distribution) instead of a white paper with four identical stars (four peaks of saliency) in four corners.

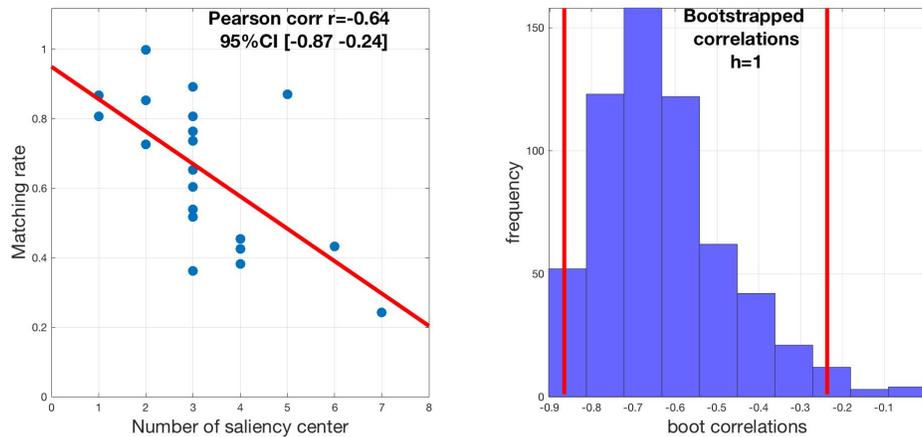


FIGURE 13 : CORRELATION ACROSS IMAGES BETWEEN MATCHING RATE AND NUMBER OF SALIENCY CENTERS FOR AN IMAGE.

Note: Right figure shows the bootstrapped values of correlations used to construct the confidence interval of bootstrapped correlations reported in the left figure.

Figure 13 shows that in matching games (both sub-sessions), the matching rate is negatively correlated with the number of saliency centers (Pearson  $r = -0.64$ ,  $p = 0.0001$ ). With larger image set and datasets, later work might be able to accurately predict the matching rate based on input context and corresponding attention maps.

### *b. Time pressure*

If the model we proposed is on the right track, manipulating the average strategic levels of the population should change the percentage of those choosing the most salient locations, which could change the associated seeker's advantage. One way to potentially lower average strategic levels is to force subjects to make

decisions more quickly, under time pressure. For example, Spiliopoulos, Ortmann, and Zhang (2018) found that people tend to choose more heuristic strategies when facing a time constraint, in several normal form games. If higher levels of strategic thinking take more time (e.g., Brocas et al. 2014), time pressure will correspond to reducing the value of  $\tau$ . Time pressure should therefore *increase* the seeker's advantage.

To test this hypothesis, we compared two new groups from mTurk. The control group (N=38) faced exactly the same experimental setting as the lab subjects (with a constraint of 6s for response). The treatment group (N=31) had a shorter time constraint of 2s. All other features of the design were unchanged.<sup>21</sup>

We first measured the seeker advantage between control and treatment groups. In the control condition, the online group still shows a significant seeker's advantage at precisely the same rate, 0.09, as in the local lab population. This seeker advantage increased significantly to 0.23 in the time pressure group no-feedback condition (p-value  $<10^{-4}$ , t-test) and to 0.15 (p-value  $<10^{-4}$ , t-test) in the time pressure group feedback condition<sup>22</sup>. Thus, our understanding of how saliency interacts with the depth of strategic thinking enabled us predict a large change in seeker advantage due to time pressure. Keep in mind that nothing in standard equilibrium analysis predicts these types of effects, because neither response time nor salience are part of the conventional game specification or behavioral analysis.

---

<sup>21</sup> For both conditions, trials with no click point recorded during the time constraint occurred in 13.3% of trials for the control group and 14.7% for the treatment group. The numbers of observations for feedback and no feedback group differed less than 5%. The smaller one is reported

<sup>22</sup> The same bootstrapping method that was used in previous seeker's advantage estimation was used here.

TABLE 3: BEST-FIT SET OF PARAMETERS

		Number of Observations	No feedback	Feedback
Baseline			0.071	0.071
Matching	Local control group	1147	0.35 (4e-4)	0.64 (4e-4)
	Online control group	458	0.41 (0.001)	0.63 (0.001)
	Online time-pressure group	388	0.31 (0.001)	0.63 (0.001)
Hider-seeker	Local control group	1090	0.09 (3e-4)	0.09 (3e-4)
	Online control group	441	0.09 (6e-4)	0.09 (6e-4)
	Online time-pressure group	372	0.23 (0.001)	0.15 (9e-4)

To investigate whether this increase in seeker's advantage was due to subjects choosing more salient locations under time pressure, we compared the mean saliency level of location choices in the control and treatment. For the hider role, the mean level of choice saliency significantly increased when people under time pressure, from 0.52 to 0.63 and 0.52 to 0.59 for the no-feedback condition and the feedback condition, respectively. The corresponding increases in saliency of seeker choices were from 0.65 to 0.81 and from 0.61 to 0.81.

### *c. Hidden Markov model*

The most direct way to measure visual attention is through eye movements. The first thing we looked at was whether the attention levels were distributed in the same way in our game environment versus free viewing (i.e., predicted by the SAM algorithm). We found that subjects often looked at the most salient locations first no matter what game conditions they were facing. The average saliency level of the first fixation is 0.92 for matching, 0.88 for seeking and 0.81 for hiding<sup>23</sup>. All those levels are significantly higher than 0.50 (t-test,  $P < 10^{-4}$ ).

To further understand the mechanism behind such strategic decision process, we modeled the gaze trajectory using a hidden Markov process (HMM). HMM is

<sup>23</sup> These percentages were calculated for both sub-sessions pooled in one game.

a common method for understanding sequential data and solving cognition problems (Rabiner 1989). A hidden Markov model hypothesizes a set of states, and rates of transition from state to state in discrete time periods. The forward and backward estimation algorithm (Murphy 2005) is then used to estimate Gaussian distributions of state-variable distributions, and transition rates. The HMM we applied has four states—a starting node and a decide node, and both unsalient and salient nodes. In the latter nodes, people are looking at either low- or high-salient locations, where the distributions of saliency within each of the two nodes are Gaussian distributed, and the mean and variance of the two Gaussians are estimated empirically. We briefly mention the main results here. More details are in an Appendix.

Although most subjects started with looking at the most salient object, the fixations afterwards varied by games and player roles. Subjects transition to the unsalient node in the hider-seeker game at rates of 32% and 20% respectively--the hidiers transition to looking at low-saliency more often. In the matching game, they transited from salient areas to unsalient areas only 8% of the time.

The results above support the critical assumption in the model that agents favor more salient areas in all games. Rather than guessing and inferring according to the existing patterns of the choice data, eye movement data directly points out that the biases came from the way people distribute their attention. In addition, it confirms that modeling attention using SAM algorithm tuned on free view data works well here, since the first fixation can be predicted with over 80% accuracy by this algorithm, even in game environments.

## 7. Conclusion

We measured visual salience in two-player games, in which player chooses locations in images, and desire to match or mismatch. Salience predicted from a finely calibrated neural network algorithm (SAM) is a strong predictor of observed choices.

Our approach specifies salience precisely. It is a use case of how the neuroscience of vision, filtered through psychology and helped by computer science, can help answer questions of long-standing interest in economics.

The foundation for the SAM algorithm is earlier algorithms (e.g., Itti, Koch, and Niebur 1998) specifying computationally how layers of visual cortex process orientation, contrast, and color at different spatial scales and integrate them. The particular SAM algorithm we apply has been calibrated empirically by eyetracking data from people freely looking at images. Following Crawford and Iriberrri (2007), salience is then embedded in a cognitive hierarchy model. Parameter estimates generate an observed seeker's advantage. Estimates can be used to predict from one type of game (HS) to another (matching), showing evidence of portability that Hargrave Heaps et al (2014) did not see in their data. Our novel evidence of cross-game portability answers the first question that motivated our paper.

A new prediction that time pressure should increase the seeker's advantage is, based on the hypothesis that lower-level SCH thinking is rapid and more salience-biased, this prediction is confirmed by new data.

The salience-dependent sequence of fixations over time, recorded by eyetracking, confirms the hypothesis that people look at more salient locations first. A hidden Markov model captures these transitions and shows some interesting differences across games.

The second motivating question is whether salience is a general principle that can be used to specify level-0 choice in CH or level-k models. The goal is to replace different level-0 specifications used for games with matrix games, private information games, and games with large number sets, based on some common principle.<sup>24</sup>

In our data visual salience provides a promising account of level 0. “Primary salience” (what is preferred or privately chosen) has been shown in early experiments by Mehta et al to explain matching rather well. Other games have not used the language of primary salience but are consistent with the idea that private preference (e.g., lucky numbers) influence choices.<sup>25</sup>

In matrix games, another type of salience which is promising as a level-0 basis is abstract properties of strategies (e.g., maximax) or strategy vectors (e.g. payoff equality) (following the early ideas of Mehta, Starmer and Sugden 1994ab on spatial and abstract norms of property rights). Wright and Leyton-Brown (2016) develop a “meta-model” of level 0 from structural features of game payoffs which, when embedded in CH, outpredicts the assumption of random level 0 play (see also Lopes-Perez et al 2014 and Leland and Schneider 2018). Evidence of differential attention to payoffs that are large or maximin is also consistent with experiments on endogenous allocation of attention by Avoyan and Schotter (2018).

A related type of salience appears to be evident in games with private information. In many such games, there is a natural mapping between private information states, and strategies. Two examples are sender-receiver games and

---

<sup>24</sup> Note that in this conception of level-0 as generally salient, not random, the assumption of randomization is equivalent to no strategy being any more salient than any other.

<sup>25</sup> Ho, Camerer, and Weigelt (1998) found a large number of Singaporean subjects choosing 80 or 88 in the 2/3-of-the-average p-beauty contest game (because “8” is a lucky number for many Chinese). In a Swedish lottery game LUPI choosing numbers from one to 5 digits, Östling et al. (2011) found that disproportionately many people picked numbers corresponding to birth years (e.g., 1900-2000).

auctions. In sender-receiver experiments, senders tend to choose “honest” response, using the state they observe as the message, more often than predicted by theory (Cai and Wang 2006; Wang, Spezio, and Camerer 2010, and Crawford 2003).

In auctions, bidders observe a signal of value. CH models based on level-0 bidders who simply bid their signal (rather than randomize) can explain some deviations from equilibrium bidding (Crawford and Iriberry 2007b). In both of these cases, it is likely that the private state or value signal is highly salient (as shown using eyetracking by Wang et al 2010). The hypothesis that level 0s are influenced by such salience helps explain overall behavior.

A promising path to explore is whether one or more concepts of salience could explain level-0 behavior in all these different results from visual image games, games from semantic or choice sets, matrix games, and private information games. In matrix games, for example, level-0 strategies that choose high-payoff rows or avoid low-payoff rows might because extreme quantities are salient (cf. Bordalo et al 2012, 2013ab). In private information games, the observed state may be visually or attentionally salient— that is, a level-0 sender may be guided by saliency rather than “honesty” per se. One way to proceed is to use a combination of off-the-shelf algorithms (such as SAM) and eyetracking methods<sup>26</sup> to see how well salience can be predicted, and whether it corresponds at all to payoff properties in matrix games or to the tendency to choose a private state as a strategy.

Another simple direction for future research is to consider other behavioral models, specifically quantal response equilibrium enhanced by salience in the same perturbation way (sQRE). Assuming that all agents weakly preferred salient

---

<sup>26</sup> Frydman and Mormann (2018) record fixations during risky money choices to test behavioral and attentional predictions of Bordalo, Gennaioli, and Shleifer (2012). Varying visual contrast (bottom-up salience) also has an effect of modulating attentional effects for gamble losses.

locations captured by parameter  $\mu$ , QRE predicts that both parties will favor salient locations and have seeker advantages when the softmax parameter  $\lambda$  is small. When  $\lambda$  is large, the equilibrium restriction can generate a hider advantage rather than seeker advantage. (More mathematical details in Appendix).

In general, quite a lot is known from visual neuroscience about the many drivers of salience, both bottom-up and top-down, though less is known about the net influence of multiple types of salience on choice (cf. Towal, Mormann, and Koch 2013). In principle a general theory of salience is within our grasp and could be applied to a wide range of social and economic choices. Indeed, salience has been increasingly recognized as a central construct to explain behavior in economic applications outside of game theory. Examples include:

1. In shopping, prices and taxes that are difficult to compute and not visibly posted are given less weight (Russo 1977; Ott and Andrus 2000; Hossain and Morgan 2006; Min Kim and Kachersky 2006; Chetty, Looney, and Kroft 2009).
2. Store items placed at eye level are chosen more often (including when experimentally varied, Dreze, Hoch, and Purk 1994).
3. Salient “ethnic or religious markers” are proposed as the basis of intergroup conflict (Esteban and Ray 2008).
4. In ordered or displayed choices, such as restaurant menus or roulette tables, there is often an “edge [or top-bottom] disadvantage” (Christenfeld 1995; Sundali and Croson 2006; Dayan and Bar-Hillel 2011).
5. Oppositely of (4), top positions on ballots garner more votes in elections (D. E. Ho and Imai 2006), leading many jurisdictions to randomize ballot order.
6. More extreme relative numbers (e.g. higher prices) appear to be more salient (see Bordalo, Gennaioli, and Shleifer (2012, 2013a, 2013b)).

We speculate that these examples may all have something in common: It is likely that differences in visual salience are at least *part of the reason* for the differences in salience which affect behavior. In principle, one could use eyetracking, clickstreams, memory, and other data to test this hypothesis by correlating measured salience to choices, and then establish causality by experimentally randomizing variables that influence salience (which can also be done in many field settings).

An illustrative use case about the potential value of visual salience comes from the world of “nudges”. Nudges are small changes in the architecture and design of choices that can improve choices of people who might make mistakes, without imposing a burden on highly-rational choosers.

Many nudges use physical design details to make information more salient. For example, Karing and Naguib (2018) had Kenyan parents bring their children for deworming treatments. They tested whether allowing parents to visibly signal their participation to others would improve participation in deworming (which is a public good because of infection risk). Note that the signaling is a one-sided matching game: Parents who dewormed prefer a type of signal that others will notice, to gain social approval. Salient signals are better.

Karing and Naguib used two different signals after deworming their children: A rubber bracelet that parents could wear around their wrist, or ink on their thumb. They note (p. 9), “Bracelets have a high visibility as they are worn around the wrist. Ink’s visibility is lower as it is applied to the thumb and only lasts for about 3 days to 2 weeks (on the skin/on the nail).” They found that awarding bracelets increased deworming rates by 24% (greater than financial incentive) but thumb ink had little effect.

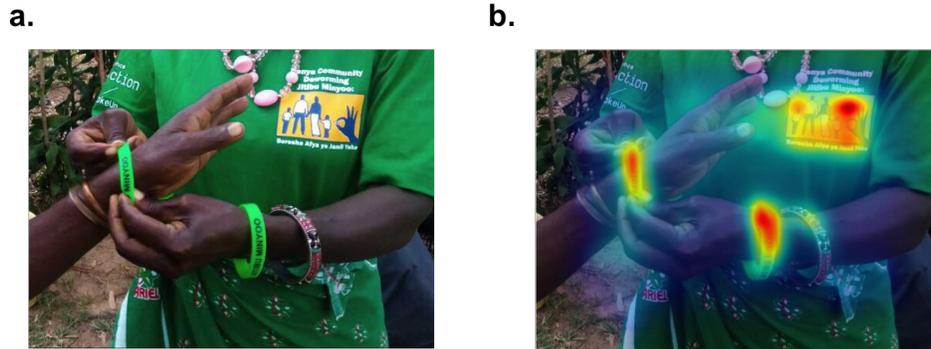


FIGURE 14: ORIGINAL IMAGE OF DEWORMING BRACELET (LEFT) AND GBVS SALIENCY MAP (RIGHT)

Source: Photograph courtesy of Anne Karing.

Why did the bracelet nudge work so well?

The authors' hunch about “high visibility” of bracelets is probably crucial to their success. A simple test of whether the bracelets are “visible” is to feed an image of a bracelet-wearing participant through the simple GBVS (Harel et al 2007) algorithm and see if the bracelet is visually salient, as predicted by the algorithm.<sup>27</sup> Figure 14 shows the result. The authors were right: The heatmap predicts intense focus on the two bracelets at the left and lower center (and also the t-shirt logo at upper right).

In this example, salience save lives. In general, ex ante predictions about what is visually salient might be useful for some other nudges where the details of displays and choice architecture are an important design feature.

---

<sup>27</sup> Thanks to Anne Karing for sharing this image. We were unable to obtain a similar picture of an inked thumb.

## REFERENCES

- Aumann, Robert J. 1974. "Subjectivity and Correlation in Randomized Strategies." *Journal of Mathematical Economics* 1 (1): 67–96.
- Awh, Edward, Artem V. Belopolsky, and Jan Theeuwes. 2012. "Top-down versus Bottom-up Attentional Control: A Failed Theoretical Dichotomy." *Trends in Cognitive Sciences* 16 (8): 437–43. <https://doi.org/10.1016/j.tics.2012.06.010>.
- Bacharach, Michael, and Michele Bernasconi. 1997. "The Variable Frame Theory of Focal Points: An Experimental Study." *Games and Economic Behavior* 19 (1): 1–45.
- Bar-Hillel, Maya. 2015. "Position Effects in Choice from Simultaneous Displays: A Conundrum Solved." *Perspectives on Psychological Science* 10 (4): 419–33.
- Baum, Leonard E., and Ted Petrie. 1966. "Statistical Inference for Probabilistic Functions of Finite State Markov Chains." *The Annals of Mathematical Statistics* 37 (6): 1554–63.
- Belot, Michèle, Vincent P. Crawford, and Cecilia Heyes. 2013. "Players of Matching Pennies Automatically Imitate Opponents' Gestures against Strong Incentives." *Proceedings of the National Academy of Sciences* 110 (8): 2763. <https://doi.org/10.1073/pnas.1209981110>.
- Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer. 2012. "Salience Theory of Choice under Risk." *The Quarterly Journal of Economics* 127 (3): 1243–85.
- . 2013a. "Salience and Asset Prices." *American Economic Review* 103 (3): 623–28.

- . 2013b. “Salience and Consumer Choice.” *Journal of Political Economy* 121 (5): 803–43.
- Brocas, Isabelle, Juan D. Carrillo, Stephanie W. Wang, and Colin F. Camerer. 2014. “Imperfect Choice or Imperfect Attention? Understanding Strategic Thinking in Private Information Games.” *Review of Economic Studies* 81 (3): 944–70.
- Cai, Hongbin, and Joseph Tao-Yi Wang. 2006. “Overcommunication in Strategic Information Transmission Games.” *Games and Economic Behavior* 56 (1): 7–36.
- Camerer, Colin F., Teck-Hua Ho, and Juin-Kuan Chong. 2004. “A Cognitive Hierarchy Model of Games.” *The Quarterly Journal of Economics* 119 (3): 861–98.
- Chetty, Raj, Adam Looney, and Kory Kroft. 2009. “Salience and Taxation: Theory and Evidence.” *American Economic Review* 99 (4): 1145–77.
- Christenfeld, Nicholas. 1995. “Choices from Identical Options.” *Psychological Science* 6 (1): 50–55.
- Cornia, Marcella, Lorenzo Baraldi, Giuseppe Serra, and Rita Cucchiara. 2016. “Predicting Human Eye Fixations via an LSTM-Based Saliency Attentive Model.” *ArXiv Preprint ArXiv:1611.09571*.
- Costa-Gomes, Miguel A., and Vincent P. Crawford. 2006. “Cognition and Behavior in Two-Person Guessing Games: An Experimental Study.” *American Economic Review* 96 (5): 1737–68.
- Crawford, Vincent P. 2003. “Lying for Strategic Advantage: Rational and Boundedly Rational Misrepresentation of Intentions.” *American Economic Review* 93 (1): 133–49.
- Crawford, Vincent P., Miguel A. Costa-Gomes, and Nagore Iriberri. 2013. “Structural Models of Nonequilibrium Strategic Thinking: Theory,

- Evidence, and Applications.” *Journal of Economic Literature* 51 (1): 5–62.
- Crawford, Vincent P., and Nagore Iriberry. 2007a. “Fatal Attraction: Saliency, Naivete, and Sophistication in Experimental “Hide-and-Seek” Games.” *American Economic Review* 97 (5): 1731–50.
- . 2007b. “Level-k Auctions: Can a Nonequilibrium Model of Strategic Thinking Explain the Winner’s Curse and Overbidding in Private-Value Auctions?” *Econometrica* 75 (6): 1721–70.
- Dayan, Eran, and Maya Bar-Hillel. 2011. “Nudge to Nobesity II: Menu Positions Influence Food Orders.” *Judgment and Decision Making* 6 (4): 333–42.
- Dreze, Xavier, Stephen J. Hoch, and Mary E. Purk. 1994. “Shelf Management and Space Elasticity.” *Journal of Retailing* 70 (4): 301–26.
- Esteban, Joan, and Debraj Ray. 2008. “On the Saliency of Ethnic Conflict.” *American Economic Review* 98 (5): 2185–2202.
- Falk, Ruma, Raphael Falk, and Peter Ayton. 2009. “Subjective Patterns of Randomness and Choice: Some Consequences of Collective Responses.” *Journal of Experimental Psychology: Human Perception and Performance* 35 (1): 203.
- Fragiadiakis, D., Ada Kovaliukaite, and D. Rojo Arjona. 2017. “Testing Cognitive Hierarchy Assumptions.” Working paper, 2017. 4, 13, 14, 15, 17.
- Frintrop, Simone, Erich Rome, and Henrik I. Christensen. 2010. “Computational Visual Attention Systems and Their Cognitive Foundations: A Survey.” *ACM Transactions on Applied Perception (TAP)* 7 (1): 6.
- Frydman, Cary, and Milica Milosavljevic Mormann. 2018. “The Role of Saliency and Attention in Choice under Risk: An Experimental Investigation.”
- Fudenberg, Drew, and Annie Liang. 2018. “Predicting and Understanding Initial Play.”

- Gales, Mark, and Steve Young. 2008. "The Application of Hidden Markov Models in Speech Recognition." *Foundations and Trends® in Signal Processing* 1 (3): 195–304.
- Hardy, Godfrey H., John E. Littlewood, and George Pólya. 1988. *Inequalities (Cambridge Mathematical Library)*. Cambridge University Press.
- Harel, Jonathan, Christof Koch, and Pietro Perona. 2007. "Graph-Based Visual Saliency." In *Advances in Neural Information Processing Systems*, 545–52.
- Hargreaves Heap, Shaun, David Rojo Arjona, and Robert Sugden. 2014. "How Portable Is Level-0 Behavior? A Test of Level-k Theory in Games With Non-Neutral Frames." *Econometrica* 82 (3): 1133–51.
- Hassan, Md Rafiul, and Baikunth Nath. 2005. "Stock Market Forecasting Using Hidden Markov Model: A New Approach." In *Intelligent Systems Design and Applications, 2005. ISDA'05. Proceedings. 5th International Conference On*, 192–96. IEEE.
- Henderson, John M., and Taylor R. Hayes. 2017. "Meaning-Based Guidance of Attention in Scenes as Revealed by Meaning Maps." *Nature Human Behaviour* 1 (10): 743.
- Ho, Daniel E., and Kosuke Imai. 2006. "Randomization Inference with Natural Experiments: An Analysis of Ballot Effects in the 2003 California Recall Election." *Journal of the American Statistical Association* 101 (475): 888–900.
- Ho, Teck-Hua, Colin Camerer, and Keith Weigelt. 1998. "Iterated Dominance and Iterated Best Response in Experimental "p-Beauty Contests"." *The American Economic Review* 88 (4): 947–69.
- Hossain, Tanjim, and John Morgan. 2006. ... "... Plus Shipping and Handling: Revenue (Non) Equivalence in Field Experiments on Ebay." *Advances in Economic Analysis & Policy* 5 (2).

- Itti, Laurent, and Pierre Baldi. 2009. "Bayesian Surprise Attracts Human Attention." *Vision Research* 49 (10): 1295–1306.
- Itti, Laurent, Christof Koch, and Ernst Niebur. 1998. "A Model of Saliency-Based Visual Attention for Rapid Scene Analysis." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 20 (11): 1254–59.
- Judd, Tilke, Frédo Durand, and Antonio Torralba. 2012. "A Benchmark of Computational Models of Saliency to Predict Human Fixations."
- Judd, Tilke, Krista Ehinger, Frédo Durand, and Antonio Torralba. 2009. "Learning to Predict Where Humans Look." In *Computer Vision, 2009 IEEE 12th International Conference On*, 2106–13. IEEE.
- Kahneman, Daniel, and Amos Tversky. 1972. "Subjective Probability: A Judgment of Representativeness." *Cognitive Psychology* 3 (3): 430–54.
- Karing, Anne, and Karim Naguib. 2017. "The Role of Social Signaling in Community Deworming Evidence from a Field Experiment in Kenya."
- Kneeland, Terri. 2015. "Identifying Higher-Order Rationality." *Econometrica* 83 (5): 2065–79.
- Krassanakis, Vassilios, Vassiliki Filippakopoulou, and Byron Nakos. 2014. "EyeMMV Toolbox: An Eye Movement Post-Analysis Tool Based on a Two-Step Spatial Dispersion Threshold for Fixation Identification." *Journal of Eye Movement Research* 7 (1).
- Leland, Jonathan W., and Mark Schneider. 2018. "A Theory of Focal Points in  $2 \times 2$  Games." *Journal of Economic Psychology* 65: 75–89.
- Manor, Barry R., and Evian Gordon. 2003. "Defining the Temporal Threshold for Ocular Fixation in Free-Viewing Visuocognitive Tasks." *Journal of Neuroscience Methods* 128 (1–2): 85–93.
- McKelvey, Richard D., and Thomas R. Palfrey. 1995. "Quantal Response Equilibria for Normal Form Games." *Games and Economic Behavior* 10 (1): 6–38.

- Mehta, Judith, Chris Starmer, and Robert Sugden. 1994. "The Nature of Saliency: An Experimental Investigation of Pure Coordination Games." *The American Economic Review* 84 (3): 658–73.
- Min Kim, Hyeong, and Luke Kachersky. 2006. "Dimensions of Price Saliency: A Conceptual Framework for Perceptions of Multi-Dimensional Prices." *Journal of Product & Brand Management* 15 (2): 139–47.
- Murphy, Kevin. 2005. *Hidden Markov Model (HMM) Toolbox for Matlab*. Retrieved May 25 2012.
- Nagel, Rosemarie. 1995. "Unraveling in Guessing Games: An Experimental Study." *The American Economic Review* 85 (5): 1313–26.
- Östling, Robert, Joseph Tao-yi Wang, Eileen Y. Chou, and Colin F. Camerer. 2011. "Testing Game Theory in the Field: Swedish LUPI Lottery Games." *American Economic Journal: Microeconomics* 3 (3): 1–33.
- Ott, Richard L., and David M. Andrus. 2000. "The Effect of Personal Property Taxes on Consumer Vehicle-Purchasing Decisions: A Partitioned Price/Mental Accounting Theory Analysis." *Public Finance Review* 28 (2): 134–52.
- Palfrey, Thomas R. 2016. *Quantal Response Equilibrium-a Stochastic Theory of Games*. Princeton University Press.
- Rabiner, Lawrence R. 1989. "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition." *Proceedings of the IEEE* 77 (2): 257–86.
- Riche, Nicolas, Matthieu Duvinage, Matei Mancas, Bernard Gosselin, and Thierry Dutoit. 2013. "Saliency and Human Fixations: State-of-the-Art and Study of Comparison Metrics." In *Computer Vision (ICCV), 2013 IEEE International Conference On*, 1153–60. IEEE.

- Rogers, Brian W., Thomas R. Palfrey, and Colin F. Camerer. 2009. "Heterogeneous Quantal Response Equilibrium and Cognitive Hierarchies." *Journal of Economic Theory* 144 (4): 1440–67.
- Rubinstein, Ariel, Amos Tversky, and Dana Heller. 1997. "Naive Strategies in Competitive Games." In *Understanding Strategic Interaction*, 394–402. Springer.
- Russo, J. Edward. 1977. "The Value of Unit Price Information." *Journal of Marketing Research*, 193–201.
- Spiliopoulos, Leonidas, Andreas Ortmann, and Le Zhang. 2018. "Complexity, Attention, and Choice in Games under Time Constraints: A Process Analysis." *Journal of Experimental Psychology: Learning, Memory, and Cognition* 44 (10): 1609.
- Stahl II, Dale O., and Paul W. Wilson. 1994. "Experimental Evidence on Players' Models of Other Players." *Journal of Economic Behavior & Organization* 25 (3): 309–27.
- Sundali, James, and Rachel Croson. 2006. "Biases in Casino Betting: The Hot Hand and the Gambler's Fallacy." *Judgement and Decision Making* 1 (1): 1.
- Towal, R. Blythe, Milica Mormann, and Christof Koch. 2013. "Simultaneous Modeling of Visual Saliency and Value Computation Improves Predictions of Economic Choice." *Proceedings of the National Academy of Sciences* 110 (40): E3858–67.
- Urrea, Luis Alberto. 2008. *The Devil's Highway: A True Story*. Back Bay Books.
- Wright, James R., and Kevin Leyton-Brown. 2016. "Models of Level-0 Behavior for Predicting Human Behavior in Games." *ArXiv Preprint ArXiv:1609.08923*.
- Zviran, Moshe, and William J. Haga. 1999. "Password Security: An Empirical Study." *Journal of Management Information Systems* 15 (4): 161–85.



## For Online Publication

### Appendix: Eye movement and fixation analysis using a Hidden Markov Model (HMM)

To preprocess the raw gaze data, we first extracted the 2D gaze positions for both eyes on each image during a particular subject's experiment time. All data used were those with validity code (0,0<sup>28</sup>). We took the midpoint of the two eyes' positions as the gaze position in this analysis. Then, we adopted a fixation detection algorithm, Eye Movements Metrics & Visualizations (EyeMMV) to get fixation centers and transitions using both spacial and durational constraint (Krassanakis, Filippakopoulou, and Nakos 2014). The minimum duration threshold we used for fixation detection was 100ms, which has been shown to be efficient according to Manor and Gordon (2003).

We first plot several fixation transition samples during the decision periods for all games (Figure 15).

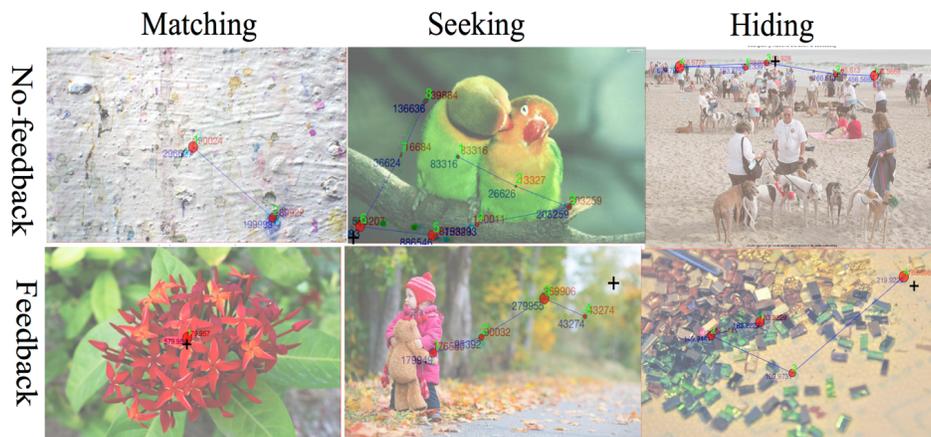


FIGURE 15: SAMPLE OF FIXATION MAPS

<sup>28</sup> In Tobii validity code scale ranges from 0-4 from most valid to least.

Each of the image above is a randomly selected fixation observation from a specific sub-session. The red circles on the images stand for fixation clusters of the gaze data. Both the number next to the circle and the radius of it indicate its duration time. The green numbers next to the circle show the order of visits and the black cross points to the final click point.

Figure 15 indicates that subjects often looked at the most salient location first no matter what game condition they were facing. The average saliency level of the first fixation is 0.92 for matching, 0.88 for seeking and 0.81 for hiding<sup>29</sup>. All those levels are significantly higher than 0.8 (t-test,  $P < 10^{-4}$ ).

If people always started by looking at the most salient area as fixation data showed, how did they end up choosing different locations later? How do they transition between salient and unsalient locations? To answer these questions, we applied a Gaussian Hidden Markov Model to all fixation data since these gaze traces revealed most information of what subjects were considering. The fundamental assumption of this model is whether such problem will satisfy Markov property, which simply means the probability distribution of any future event happening only depends on current states, and independent of the historical paths. Towal, Mormann, and Koch (2013) had tested the Markov property was valid in general gaze transition problem.

HMM is a probabilistic model of both visible emissions and hidden states (Baum and Petrie 1966). It has been largely used in solving speech detection problem (Gales and Young 2008), financial prices predictions (Hassan and Nath 2005) and other machine learning questions. Each state has a probability distribution over some possible outcomes (“tokens”). Here we assume such probability distribution is Gaussian, a version called Gaussian HMM.

---

<sup>29</sup> It was calculated for both sub-sessions in one game.

In this paper, the observations generated by a HMM corresponded to fixation transition vectors between hidden states. These vectors will be used to estimate the hidden states and transitioning probabilities. The underlying hypothesis is that people were deciding between very salient locations and less salient locations, which can be seen in HMM as transiting between a salient hidden state and another unsalient hidden state (Figure 16). In each hidden state, the HMM creates one observed point (one fixation) based on the Gaussian distribution of saliency in that state. For instance, if a person is currently as being in the on a salient state S, he is statistically likely to be looking at high-saliency locations (where the Gaussian distribution of saliency values in each of the states U and S is estimated from the data). After each state, one could either choose to stay on the same state, transit to the other state or decide.

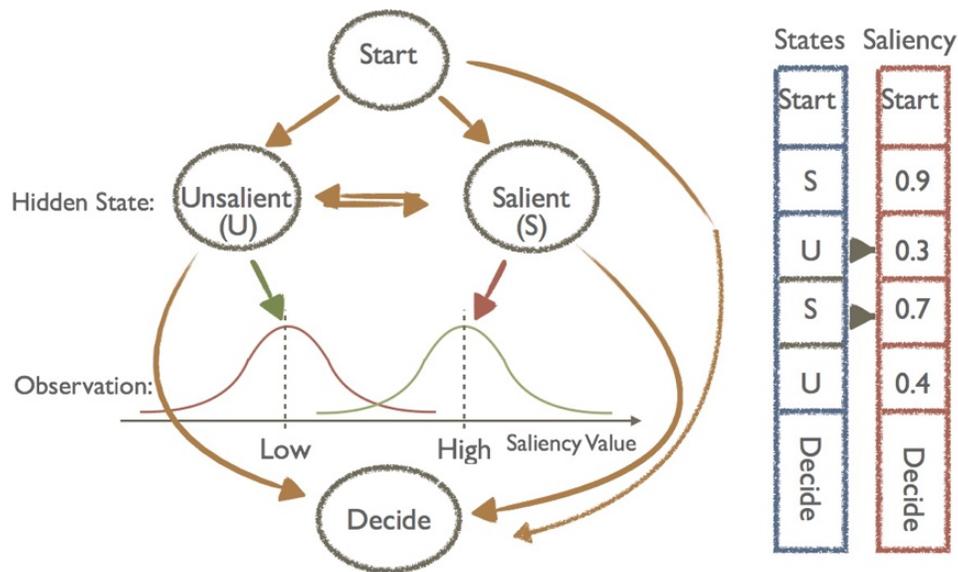


FIGURE 16: HMM MODEL

Left panel shows the relationship between hidden units and observations. On each hidden unit, the HMM emits an observation point (one fixation) on saliency space following a Gaussian distribution. The means of the Gaussian differ from high salient space to low salient space. The right panel presents an example emission out of a sample realized state transitioning. The red vector is what the fixation transition vector on saliency space looks like.

We used the HMM toolbox in Matlab (Murphy 2005) to get the best fitted transition matrix and Gaussian parameters in each game (Figure 17). This figure demonstrates the different gaze transition patterns among different games. In all three game roles, there is a strong tendency to transition from the starting state to a salient state (S); those transition probabilities range from 87 to 98%. In both states U and S, the predominant transition is stay within the same salience class (at rates from 48-61%). In matching games, people are much more likely to transition from U to S, searching for a high-salience point, than in the reverse direction from S to U (31% > 8%). The opposite is true in hiding and seeking; in those game roles, the transition from S to U is more common. After the first fixation, different from matching game, people transitioned more often to unsalient locations in hider-seeker game. This finding provides an evidence that people shared a common prior, the salient stimulus, during their strategic reasoning process.

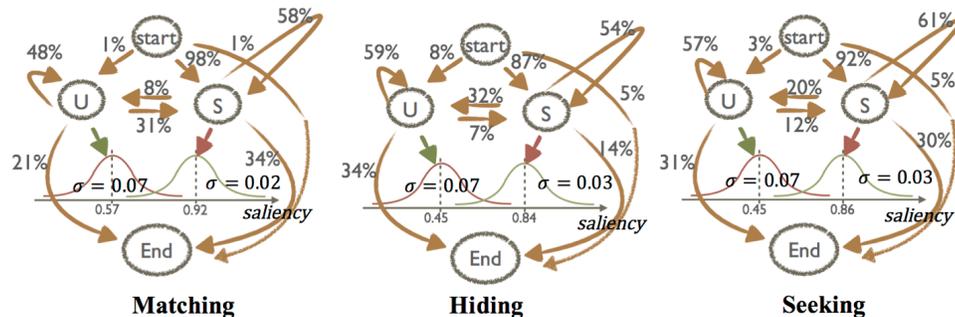


FIGURE 17: BEST-FITTED TRANSITION PROBABILITY

These three graphs showed the fitted transition probabilities for matching, hiding and seeking games. The probabilities on each arrow showed the estimated transition probability between any two states. The numbers near the red curve and green curve in the middle showed the estimated emission probability from each state to observations. We bootstrapped the data to get the confidence intervals for these estimates (Appendix, table 3).

TABLE 4: THE 95% CONFIDENCE INTERVALS OF ALL ESTIMATES IN FIGURE 10 ARE PRESENTED.

	Matching	Hiding	Seeking
Mean seeker	(0.54,0.60)	(0.42,0.48)	(0.42,0.48)
Mean hider	(0.90,0.91)	(0.83,0.86)	(0.85,0.87)
P(SI Begin)	(0.97,0.99)	(0.84,0.89)	(0.90,0.93)
P(UI Begin)	(0.01,0.02)	(0.06,0.11)	(0.02,0.04)
P(EndlBegin)	(0.004,0.01)	(0.04,0.06)	(0.04,0.07)
P(SIS)	(0.56,0.61)	(0.50,0.57)	(0.58,0.63)
P(EndlS)	(0.32, 0.35)	(0.16,0.12)	(0.17,0.19)
P(UIS)	(0.07,0.10)	(0.28,0.36)	(0.19,0.24)
P(SIU)	(0.24,0.39)	(0.05,0.10)	(0.09, 0.14)
P(UIU)	(0.42,0.55)	(0.56,0.61)	(0.54,0.60)
P(EndlU)	(0.17,0.25)	(0.32,0.36)	(0.29,0.34)

S means the salient state and U means the unsalient state. Each probability stands for a transition probability from one to another. The first two rows show the mean of the Gaussian distributions for each state.

## For Online Publication

### Appendix: A QRE alternative model

An alternative model is a saliency perturbed version of the quantal response model (QRE) (McKelvey and Palfrey 1995; Palfrey 2016) In this Appendix. we will introduce a QRE model including saliency similarly to how it was adding in

the SCH analysis. We will also show some results of how the predicted seeking advantage depends on the response parameter.

Assuming that agent  $i$  has the following utility function for choosing option  $n$  in a finite strategy space:

$$1. \quad u_{in} = M + \mu S_n$$

$M$  is the monetary reward he gets for choosing that option. He gets extra utility from choosing a more salient spot (this could be interpreted as a reduced-form way of including attentional costs, as salient locations are less costly to attend to).

Players play the game and maximize their expected utility given the salience-enhanced specification. The logit version of QRE strategy  $p$  satisfies:

$$2. \quad P_{in} = \frac{e^{\lambda E U_{in}(p-i)}}{\sum_m e^{\lambda E U_{im}(p-i)}}$$

For the particular game used in our experiment, the strategy space is finite with each choice corresponding to a pixel on an image, denoted by  $\{n | n = 1, 2, \dots, N\}$ . Therefore, in our game, such  $P_{in}$  described above with  $i = s$  or  $h$  plus the condition that  $\sum P_{in} = 1$  constitutes a quantal response equilibrium.

Without the loss of generality, hider's winning amount will be set to one and seeker's winning amount will be set to be  $\sigma$ . This way, the equilibrium probability can be further written as:

$$3. \quad P_{sn} = \frac{e^{\lambda(\sigma P_{hn} + \mu S_n)}}{\sum_m e^{\lambda(\sigma P_{hm} + \mu S_m)}}$$

$$4. \quad P_{hn} = \frac{e^{\lambda(1 - P_{sn} + \mu S_n)}}{\sum_m e^{\lambda(1 - P_{sm} + \mu S_m)}}$$

Notice that the denominators of equation 3 and 4 are actually independent of  $n$ . Therefore, let's denote each of the denominator as  $C_s$  and  $C_h$ , respectively. In the previous part, we have shown that all the choices need to have positive probability to be chosen in order to sustain an equilibrium. Therefore, we can transform 3 and 4 as below since the right hand sides of the equations are never zero:

$$5. \quad \frac{P_{sn}}{P_{hn}} = \frac{C_s e^{\lambda \sigma P_{hn}}}{C_h e^{\lambda(1-P_{sn})}}$$

$$6. \Leftrightarrow P_{sn} e^{\lambda(1-P_{sn})} = P_{hn} \frac{C_s}{C_h} e^{\lambda \sigma P_{hn}}$$

Now, we want to analyze the matching rate conditioning on the two parameters,  $\lambda$  and  $\sigma$ . The matching rate is  $\sum_{i=1}^n P_{si} P_{hi}$ . To analyze it, we need a theorem called *rearrangement inequality* (Hardy, Littlewood, and Pólya 1988).

*Rearrangement inequality:*

For two real sequences  $x_n$  and  $y_n$  in order:

$$7. \quad x_1 \leq x_2 \dots \leq x_n, \quad y_1 \leq y_2 \dots \leq y_n$$

and an arbitrary permutation  $x_{\sigma(n)}$

$$8. \quad \Rightarrow \sum_{i=1}^n x_i y_i \geq \sum_{i=1}^n x_{\sigma(i)} y_i \geq \sum_{i=1}^n x_{(n+1-i)} y_i$$

Now, let's sum up all the possible permutations together in the middle so that it becomes one.

$$9. \quad \Rightarrow n \sum_{i=1}^n x_i y_i \geq \sum_{\sigma} \sum_{i=1}^n x_{\sigma(i)} y_i \geq n \sum_{i=1}^n x_{(n+1-i)} y_i$$

$$10. \quad \sum_{i=1}^n x_i y_i \geq \frac{1}{n} \sum_i x_i \sum_{i=1}^n y_i \geq \sum_{i=1}^n x_{(n+1-i)} y_i$$

If,  $x_i$  and  $y_i$  are probabilities and add up to one, which is our case, equation 10 becomes:

$$11. \quad \sum_{i=1}^n x_i y_i \geq \frac{1}{n} \geq \sum_{i=1}^n x_{(n+1-i)} y_i$$

Let's discuss the matching rate into three cases. Since the RHS of equation 6 is an increasing function of  $P_h$ , we will focus on the first order derivative of the LHS. The conditions are as follows:

**Case 1:**  $P_{si} < \frac{1}{\lambda}$  for all i

From equation 6, we can see that when  $P_{si} < \frac{1}{\lambda}$ , for any  $P_{si} \geq P_{sj}$ ,  $P_{hi} \geq P_{hj}$ . It is easy to see by taking the derivative of the right hand side with respect to  $P_{hn}$ .

In addition, since  $P_{si} \leq 1$ , this case always happen when  $\lambda < 1$ .

Therefore, the left side of equation 11 applies here:  $\sum_{i=1}^n P_{si} P_{hi} \geq \frac{1}{n}$ . It indicates that there is a seeker's advantage for  $\lambda < 1$  and can be a seeker's advantage for larger values of  $\lambda$ .

**Case 2:**  $P_{si} > \frac{1}{\lambda}$  for all i

From equation 6, it can be seen that for any  $P_{si} \geq P_{sj}$ ,  $LHS_i = RHS_i \leq RHS_j = LHS_j$  so that  $P_{hj} \geq P_{hi}$  and  $\sum_{i=1}^n P_{si} P_{hi} \leq \frac{1}{n}$ .

Since n is finite, therefore, this condition becomes  $\lambda > \frac{1}{P_{s \min}}$ . As  $\lambda$  becomes large enough (people are closer to best-response) a hider's advantage can occur for sufficiently large  $\lambda$ .

