Inferring Intentional Agents From Violation of Randomness

Yuan Meng (yuan_meng@berkeley.edu)
Thomas L. Griffiths (tom_griffiths@berkeley.edu)
Fei Xu (fei_xu@berkeley.edu)
Department of Psychology, University of California, Berkeley
Berkeley, CA 94720 USA

Abstract

Humans have a strong "cognitive compulsion" to infer intentional agents from violation of randomness and such an agency–nonrandomness link emerges early in development. In two studies, we directly quantified, formalized, and compared both ends of this link for the first time. In Experiment 1, two groups of participants viewed the same 256 binary sequences (e.g., AABAAABA) and classified each as generated by agents/non-agents or by nonrandom/random processes. We found a strong correlation between two judgments: sequences viewed as more agentive also tended to be judged as less random. In Experiment 2, another two groups were asked to produce sequences that others might appreciate as agentive or nonrandom. Participant-generated sequences in the two conditions had a substantial overlap, indicating common guiding principles of agency and nonrandomness generation. Taken together, the present studies provide evidence for a shared cognitive basis of agency detection and subjective randomness.

Keywords: agency; subjective randomness; agency–nonrandomness link; animate-inanimate distinction

Introduction

We can accept a certain amount of luck in our explanations, but not too much. The question is, how much?

– Richard Dawkins, The Blind Watchmaker (1996)

When we look at something as delicate and *orderly* as the eye, it is only natural to believe that such a work of art must be designed by *someone*, an intentional agent with a purpose in mind. In contrast, it takes "a very large leap of the imagination to think the other way around" (Dawkins, 1996, p.7).

This cognitive compulsion to infer agents from order may have given birth to thousands of religions shared by the vast majority of people on Earth (Barrett, 2000; Keil & Newman, 2015) and it dates back to early childhood (Friedman, 2001) and infancy (Ma & Xu, 2013; Ma, Berthiaume, Hoch, & Xu, under revision; Newman, Keil, Kuhlmeier, & Wynn, 2010). By 10 and 12.5 months of age respectively, infants appreciate that only agents can create regular visual (e.g., YYRYYRYYR, Y and R stand for yellow and red balls; Ma & Xu, 2013) and auditory (e.g., TTTCTTTCTTTCTTTC, T and C stand for tambourine and cowbell sounds; Ma et al., under revision) sequences. Similar appreciation has been found in different tasks. For instance, 12-month-olds expect agents (e.g., a ball) but not inanimate objects (e.g., a perceptually similar ball with eyes) to bring order to a disorderly pile of blocks.

Keil and Newman (2015) used these findings to argue that during the first few months of life, infants observe a bulk of ordering and disordering events together with their causes and come to "associate only agents but not non-agents with many kinds of ordered and nonrandom sequences" (p.132). Notwithstanding counterexamples such as molecular self-assembly or evolution, this association is often true in our universe where entropy tends to increase over time; in contrast, it takes *energy*, *information*, and *goal-direction* to go against the force of nature, all of which strongly indicate agentive causes even to the youngest humans (see Baillargeon, Scott, & Bian, 2016, for a review). Based on these experiences and intuitive theories, deviation from randomness often indicates nonrandom generation processes behind the scenes (Griffiths & Tenenbaum, 2001, 2003, 2004, 2007; Sim & Xu, 2013; Williams & Griffiths, 2013). When the specifics are unknown, people often default to agents as the "causal placeholders", thinking that they did it somehow (Saxe, Tenenbaum, & Carey, 2005; Wu, Muentener, & Schulz, 2015).

If the above line of reasoning holds true, people's judgment of agency should align with their judgment of nonrandomness—the less random something looks, the more agentive it strikes us as. This prediction might be hinted by Kushnir, Xu, and Wellman's (2010) finding that 20-month-olds use violation of random sampling to infer the preference of an agent ("She always picked this toy despite its rarity, so she must really *like* it."). It is possible that nonrandomness not only indicates the psychological states of an already known agent, but also cues its very existence. Imagine if you see a haystack in the middle of a desert (which apparently violates the surrounding vegetation distribution)—you may well conclude that *someone must have been there* apart from that she is very fond of hay. Alternatively, as Feldman and Tremoulet (2008)[1] suggested, the *intermediate* level of nonrandomness is the strongest cue of agency: "Too simple—a simple periodic noise burst, say—and it's an inanimate source, say, a rotating pulsar. Too complex—a totally patternless sequence—and it's just random electromagnetic interference. To seem intelligent it has to be somewhere in between: patterned, but neither perfectly periodic nor completely chaotic" (p.22).

Empirical evidence is needed to test the two possibilities. To the best of our knowledge, there is no research directly examining the relationship between human intuitions about agency and randomness. The closest work to date (Ma & Xu, 2013; Ma et al., under revision), for instance, lacks a measure of the regularity of the stimuli. Without quantifying

[1] Feldman and Tremoulet (2008) only briefly mentioned this hypothesis at the end of their paper, the bulk of which focused on detecting agency from observed motion patterns. The idea of using computational models (e.g., finite state automaton) to formalize agent detection is shared by our paper, but we dived into a very different domain—static visual sequences, where there is a substantial literature on the formalism of subjective randomness to draw on.

subjective agency and subjective nonrandomness, however, it is hard to tell whether or not they go hand in hand. In addition, without an independent measure of regularity, one may find herself trapped in circular reasoning—*"What is regularity?" "It's the evidence from which people infer intentional agents." "How do people infer intentional agents?" "From regularity."* Also, past studies only looked at a small subset of all possible binary sequences of equal length (9 digits in Ma & Xu, 2013; 12 digits in Ma et al., under revision). Chances are that regular sequences in those studies *happened to* look both regular to the researchers (which is most likely why they were chosen in the first place) and agentive to the infants. We cannot rule out the possibility that there exist a considerable amount of "irregular" sequences that look like the work of an agent, or reversely, "regular" sequences that only call for an inanimate cause. Therefore, to systematically investigate the agency–nonrandomness link, we need to include a wide range of sequences rather than a selected few.

Linking agency and randomness

To shed light upon the agency-nonrandomness link, we conducted two experiments. In Experiment 1, one group of participants viewed 256 binary sequences of length 8 (e.g., ΛΛΓΛΓΛΛΛ) and classified the source of each sequence into agentive or non-agentive entities while another group classified the source of each as random or nonrandom processes[2]. Should agency detection be tightly related to subjective randomness, we would expect a high correlation between ratings from the two groups. Sometimes, we not only need to detect other agents, but also wish to be detected by others. For instance, if we are abducted and locked in a truck, flashing the taillight in a "meaningful" way may attract the police and save our lives. Do we actually have such a good intuition about what kind of sequences *others* may appreciate as agentive? Is it related to our intuition of what others will view as nonrandom? We looked into these questions in Experiment 2 by asking another two groups of participants to generate an 8-digit binary sequence that they thought might receive the highest agency or nonrandomness score. We examined whether the mean scores of participant-generated sequences were higher than that of all 256 in Experiment 1, as well as the overlap between sequences generated by the two groups.

If people indeed make similar judgments about agency and nonrandomness, then the question is whether they solve the two problems in a similar way. In the last decade, Griffiths and Tenenbaum (2001, 2003, 2004, 2007) formalized the problem of randomness detection as a statistical inference of the *data generation process* given the data: that is, when judging if a given sequence X is random, people are comparing the probability that X was generated by a random process ($P(\text{random}|X)$) against the probability that X was generated by a regular process ($P(\text{regular}|X)$). The ratio of these two probabilities, or "posterior odds", is given by Bayes' theorem (below is the log-odds form):

$$\log \frac{P(\text{random}|X)}{P(\text{regular}|X)} = \log \frac{P(X|\text{random})}{P(X|\text{regular})} + \log \frac{P(\text{random})}{P(\text{regular})}, \tag{1}$$

The *subjective randomness* of sequence X is defined as the only part that depends on X—the log-likelihood ratio:

$$\text{randomness}(X) = \log \frac{P(X|\text{random})}{P(X|\text{regular})}. \tag{2}$$

If X results from flipping a fair coin or the equivalent, then $P(X|\text{random})$ is simply $\frac{1}{2}^{l(X)}$ ($l(X)$ is the length of X)—the heart of the problem thus becomes evaluating $P(X|\text{regular})$. Griffiths and Tenenbaum (2003, 2004) specified $P(X|\text{regular})$ using a hidden Markov model (HMM) that associates each symbol x_i (e.g., `H`) in X with a hidden state z_i (e.g., repeating `H`). The probability that X is generated by a certain HHM is obtained by summing over the probability that X is generated by each of all possible states Z under this model:

$$P(X) = \sum_{Z} P(X, Z), \tag{3}$$

Knowing that each x_i solely depends on z_i and each z_i is determined by z_{i-1}, we can rewrite Equation 3 as below:

$$P(X, Z) = P(z_0) \prod_{i=2}^{n} P(z_i|z_{i-1}) \prod_{i=1}^{n} P(x_i|z_i). \tag{4}$$

In this study, the regular generation process is defined by 22 repeating "motifs" of length 1 (repeating `H` or `T`) to 4 (e.g., repeating `HTHH`, `TTHH`, *etc.*). Each symbol in the 22 motifs corresponds to a hidden state, which amounts to 72. The prior of each motif is α^k (k is the length of a motif) and the probability of continuing with a motif is δ. Using this HMM, we can estimate the subjective randomness of all 256 sequences.

Experiment 1: Judging sequences

Method

Participants Seventy-four participants with a United States IP address took part in Experiment 1 on Amazon Mechanical Turk (http://www.mturk.com/, "MTurk") for a payment of \$3.5. A past acceptance rate equal to or greater than 93% was required for participation. 40 participants (20 women; mean age = 37.15, *SD* = 12.63, range: 19–70 years) were randomly assigned to the agency judgment task and 34 (17 women; mean age = 33.98, *SD* = 10.45, range: 23–67 years) to the nonrandomness judgment task. Another 36 were excluded for failing one or both instruction check questions.

Stimuli and procedure To begin, participants read a cover story corresponding to their task:

Agency condition. "Welcome to year 3017! Imagine you are a space rescuer whose job is to search for astronauts lost in deep space. These days spaceships are all equipped with a radio transmitter. To call for help, astronauts can use it

[2] Griffiths and Tenenbaum (2003) only showed each participant half (128) of the stimuli—that is, one either saw a sequence (e.g., `HTHHTHTT`) or its complement (e.g., `THTTHTHH`).

to send out sequences made of two types of radio waves—Lambda (Λ) and Gamma (Γ). However, both types of waves may also be produced by natural phenomena such as celestial body activities—in this case, Λ and Γ are equally likely to appear. Everyone in space knows that by current technical standard, radio receivers can only pick up 8 waves in a row—that is, you can only detect sequences that have 8 waves (as mentioned before, each wave is either Λ or Γ). If you receive a sequence and think it was produced by natural phenomena, you will ignore it and stay on course. If you think it came from humans, then you will go to them. In this study, you will see about 250 sequences. Your task is to decide whether each sequence was generated by a natural phenomenon or by a human astronaut. Please answer as quickly and accurately as possible!"

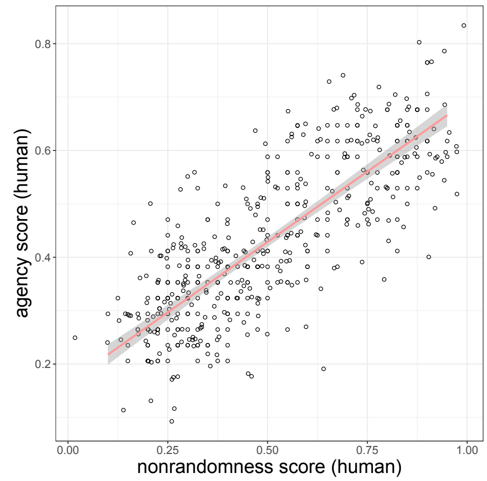
Nonrandomness condition. "You're about to see some sequences made of 8 symbols. Each symbol is either Lambda (Λ) or Gamma (Γ)—which may represent heads or tails, even or odd digits, successes or failures, or other event outcomes. For instance, the sequence $\Lambda\Gamma\Gamma\Lambda$ could stand for "tails, heads, heads, tails", "even, odd, odd, even", *etc*. Some of these sequences were created by tossing an actual fair coin, which means they are random series—in this case, Λ and Γ are equally likely to appear. However, other sequences may be generated by nonrandom processes, such as computer programs, successes and losses of a basketball team, and so on. In this study, you will see about 250 sequences. Your task is to decide whether each sequence was generated by a random process or by a nonrandom process. Please answer as quickly and accurately as possible!"

Two quizzes immediately followed to test participants on the sources of sequences as well as the chance of two symbols appearing in the natural phenomenon or the random processs scenario. Images (pixel resolution: 700×525) of 256 binary sequences of length 8 were then displayed on the screen. In the agency judgment task, participants were asked to classify the source of each sequence into "human astronaut" or "natural phenomenon", and in the nonrandomness judgment task, into "nonrandom process" or "random process". The display order was randomized and the relative location of choices counterbalanced between participants. The whole process was self-paced and took an average of 20 minutes.

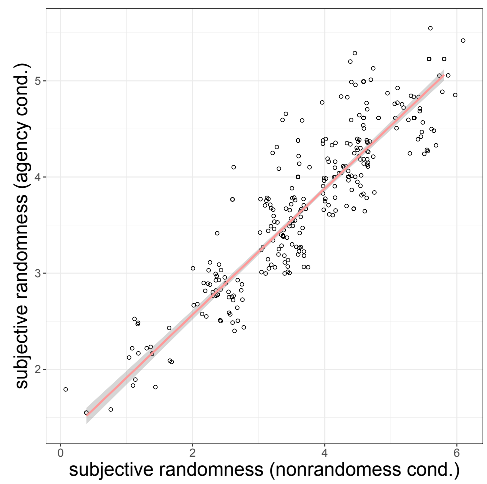
Results

An alpha level of .05 was used for all statistical analyses. Each sequence received an *agency score* (the proportion of participants classifying its source as "human astronaut") as well as a *nonrandomness score* (the proportion of participants classifying its source as "nonrandom process").

To examine the stability of participants' judgments, we looked at the scores of 128 sequences (e.g., $\Lambda\Lambda\Gamma\Gamma\Lambda\Lambda\Gamma\Lambda$) and their complements (e.g., $\Gamma\Gamma\Lambda\Lambda\Gamma\Gamma\Lambda\Gamma$)—in theory, they should be the same. Indeed, the sequence-complement correlation was high in the agency condition, $r(126) = .84$, 95% CI [.78, .89], $p < .001$, adjusted $R^2 = .71$, and even higher in the nonrandomness condition, $r(126) = .93$, 95% CI [.91, .95], $p < .001$, adjusted $R^2 = .87$, suggesting that participants were making reliable judgments during the task.



(a) Human judgments.



(b) Model predictions.

Figure 1: The agency-nonrandomness link in Experiment 1.

Of central interest to Experiment 1 was the correlation between agency and nonrandomness scores. We found a strong positive correlation between the two, $r(254) = .84$, 95% CI [.81, .88], $p < .001$, adjusted $R^2 = .72$ (see Figure 1a).

To see whether detecting agents was a similar problem to detecting deviation from randomness, we tested if the same model (as specified earlier) fit data from both tasks reasonably well. Since our models predicted subjective *randomness*, we recoded agency scores into non-agency scores (1 - agency score) and nonrandomness scores into randomness scores (1 - nonrandomness score). Fitting the non-agency data gave $\delta = .44$ and $\alpha = .11$, with correlation $r(254) = .72$, $p < .001$, and fitting the randomness data gave $\delta = .45$, $\alpha = .12$, with correlation $r(254) = .75$, $p < .001$. Model predictions in the two conditions were strongly correlated, $r(254) = .96$, 95% CI [.95, .97], $p < .001$, adjusted $R^2 = .92$ (see Figure 1b).

Discussion

In Experiment 1, we found a strong correlation between participants' agency and nonrandomness judgments—that is, sequences viewed as less random were rated as more agentive. A hidden Markov model fit human data in both tasks reasonably well and produced highly overlapping predictions for subjective non-agency and subjective randomness, which suggests that people are solving highly similar problems when detecting agents and violation of randomness.

At the beginning of this paper, we discussed two possible forms the agency–nonrandomness link. Feldman and Tremoulet (2008) suggested that mid-level randomness is most agentive, according to which we should find a reverse U-shaped/U-shaped relation between the degree of agency/non-agency and the degree of randomness. However, this was neither the case in human judgments (Figure 1a) nor model predictions (Figure 1b). Instead, the findings in Experiment 1 provided evidence for a *linear* agency–nonrandomness link.

Experiment 2: Generating sequences

Participants A total of 212 participants with a United States IP address who did not participate before took part in Experiment 2 on MTurk for a payment of \$0.5. A past acceptance rate equal to or greater than 93% was required for participation. 105 participants (38 women; mean age = 35.59, SD = 10.54, range: 18–67 years) were randomly assigned to the agency generation task and 107 (34 women, 1 other; mean age = 36.36, SD = 11.63, range: 19–65 years) to the non-randomness generation task. Another 78 were excluded for failing one or both instruction check questions or generating sequences that were not binary or of length 8.

Method

Stimuli and procedure To begin, participants read a cover story corresponding to their task:

Agency condition. "Welcome to year 3017! Imagine you are an astronaut who is lost in deep space after an accident. These days spaceships are all equipped with a radio transmitter. To call for help, you can use it to send out sequences made of two types of radio waves—Lambda (Λ) and Gamma (Γ). However, both types of waves may also be produced by natural phenomena such as celestial body activities—in this case, Λ and Γ are equally likely to appear. Everyone in space knows that by current technical standard, radio receivers can only pick up 8 waves in a row—that is, you should only send sequences that have 8 waves (as mentioned before, each wave is either Λ or Γ). If space rescuers receive your sequence and thinks it was produced by natural phenomena, they will ignore it and stay on course. If they think it came from humans, then they will come to you.I n order to save yourself, what sequence will you send to space?"

Nonrandomness condition. "You're about to write down a sequence made of 8 symbols. Each symbol is either Lambda (Λ) or Gamma (Γ)—which may represent heads or tails, even or odd digits, successes or failures, or other event outcomes. For instance, the sequence ΛΓΓΛ could stand for 'tails, heads, heads, tails', 'even, odd, odd, even', *etc*. Sequences like this could be created by tossing an actual fair coin, which means they are random series. However, they could also be generated by nonrandom processes, such as computer programs, successes and losses of a basketball team, and so on. In this study, you will come up with one sequence. Your task is to make it look least random—that is, this sequence should NOT look like the product of a random process; instead, it should look like it's generated by a nonrandom process. In order to fulfill your task, what sequence will you write down?"

Two quizzes immediately followed to test participants on what kind of sequence they should generate and the chance of two symbols appearing in the natural phenomenon or the random processs scenario. Then they entered a sequence in 8 text entry cells (the task requirement was visible). The whole process was self-paced and took about 2–3 minutes.

Results

Since symbols have no inherent meanings, sequences (e.g., ΛΓΛΛΓΓΛΓ) and their complements (e.g., ΓΛΓΓΛΛΓΛ) were coded as the same form (e.g., ABAABBAB). Participants in the agency condition generated 31 unique sequences while those in the nonrandomness condition generated a total of 35. Table 1 summarized unique sequences generated in both conditions (16 overlapping sequences are marked in yellow).

To begin, we looked at whether participants were able to generate *good* sequences with respect to the task requirement. To do so, we assigned agency and nonrandomness scores in Experiment 1 to participant-generated sequences in Experiment 2 (e.g., the score of ABAABBAB would be the mean score of ΛΓΛΛΓΓΛΓ and ΓΛΓΓΛΛΓΛ). In the agency condition, participant-generated sequences had higher scores (M = .56, SD = .19) compared to all 256 sequences (M = .43, SD = .14), mean difference = .13, 95% CI [.08, .16], $t(154.56)$ = 6.58, p < .001, d = .86. Participant-generated sequences (M = .62, SD = .22) also received higher nonrandomness scores than that of the whole set (M = .50, SD = .23), mean difference = .12, 95% CI [.09, .17], $t(286.11)$ = 5.87, p < .001, d = .58. In both the agency and the nonrandomness condition, the *frequency* of sequences being generated was positively correlated with their scores, r(103) = .75, 95% CI [.66, .83], p < .001, adjusted R^2 = .56, r(105) = .53, 95% CI [.38, .65], p < .001, adjusted R^2 = .27, respectively.

To see if the agency–nonrandomness link exists in sequence generation, we examined the overlap of participant-generated sequences in the two conditions. First, 87 out of 105 "agentive" sequences were also generated in the non-randomness condition and a similarly large proprotion of "nonrandom" sequences—83 out of 107—were found in the agency condition as well. Among the 31 unique agentive sequences and 35 unique nonrandom sequences, 16 were shared by both. The question is whether this overlap was due to chance, a problem that is often faced by bioinformatics scientists when deciding given a genome with N genes, whether one gene list with a genes overlaps with another with b genes

Table 1: Participant-generated sequences in Experiment 2.

nonrandom	freq.	score	agentive	freq.	score
ABABABAB	21	0.60	AAAAAAAA	40	0.76
AAAAAAAA	17	0.90	ABABABAB	20	0.40
AAAABBBB	15	0.85	AAAAAAAB	3	0.65
AABBAABB	14	0.56	AAABBBAA	3	0.50
AAAAAAAB	3	0.85	AABABBAB	3	0.29
AAABAAAB	2	0.73	ABBAABBA	3	0.37
AAABAABA	2	0.53	AAAABBBB	2	0.56
AAABBBAA	2	0.73	AAABAAAB	2	0.65
AABBABAB	2	0.25	AAABAABA	2	0.54
AABBBAAB	2	0.40	AAABBAAA	2	0.65
ABABAABB	2	0.26	AAABBBAB	2	0.35
ABBABBAB	2	0.49	AABBAABB	2	0.47
AAAABAAB	1	0.60	ABAABBAA	2	0.34
AAABABBA	1	0.21	ABABBABA	2	0.25
AAABBAAA	1	0.73	AAAAAABB	1	0.60
AAABBABA	1	0.33	AAAAABAA	1	0.63
AAABBBAB	1	0.48	AAAABBAA	1	0.57
AABAAABA	1	0.51	AAABABAB	1	0.32
AABABAAA	1	0.75	AABABAAA	1	0.56
AABAABBB	1	0.30	AABBBAAA	1	0.51
AABABBAB	1	0.23	AABBBABA	1	0.34
AABABBBA	1	0.23	AABBBBBB	1	0.57
AABBAAAB	1	0.39	ABAABAAA	1	0.41
AABBABBA	1	0.33	ABAABAAB	1	0.43
AABBBAAA	1	0.69	ABAABABB	1	0.28
ABAAABAB	1	0.33	ABAABBAB	1	0.24
ABAABBAA	1	0.39	ABABBBAA	1	0.28
ABAABBBA	1	0.24	ABBAAABA	1	0.29
ABBABAAB	1	0.19	ABBAAABB	1	0.34
ABBABBAA	1	0.30	ABBBAABB	1	0.31
ABBABBBA	1	0.40	ABBBBBBA	1	0.66
ABBBAABA	1	0.20			
ABBBAABB	1	0.35			
ABBBABAA	1	0.24			
ABBBBBBA	1	0.90			

if they have an intersection of *t* genes. In our study, the "genome" was all 128 unique sequences while participant-generated agentive and nonrandom sequences were the two "gene lists". Using Fisher's exact test implemented by the `GeneOverlap` R package (Version 1.12.0; Shen & Sinai, 2013), we found that the 16-sequence overlap between two conditions was unlikely to arise by mere chance, $p < .001$.

Discussion

Participants in Experiment 2 showed a good sense of what sequences may strike others as agentive or nonrandom: in both the agency and the nonrandomness condition, they generated sequences with higher scores than all 256 sequences. Crucially, we found a statistically meaningful overlap between agentive and nonrandom sequences, indicating that people not only make similar judgments about agency and nonrandomness, but may also be guide by similar intuitions when generating stimuli that are agentive or nonrandom.

General Discussion

The present studies provide evidence for a shared cognitive basis of agency detection and subjective randomness. In Experiment 1, participants made similar judgments about agency and nonrandomness: sequences viewed as more agentive also tended to be judged as less random. A hidden Markov model with 72 states and 22 motifs fitted human performance in both tasks equally well and produced identical predictions regarding the degree to which people *should* view each sequence as agentive or nonrandom. In Experiment 2, participants did a good job generating sequences that others might see as agentive or nonrandom. Sequences in these two conditions had a substantial overlap, indicating common guiding principles of agency and nonrandomness generation.

Our work contributes to a growing body of literature on the perceived link between order and agency (e.g., Barrett, 2000; Friedman, 2001; Ma & Xu, 2013; Ma et al., under revision; Newman et al., 2010) by *directly* quantifying, formalizing, and comparing both ends of this link for the first time.

As Williams and Griffiths (2013) pointed out, randomness judgments often boil down to *relative frequency* (Are two equally likely events equally frequent?) and *sequential dependence* (Do earlier events influence subsequent events?). However, past researchers studied the link between these two apsects of randomness (or lack thereof) and agency *separately*—for instance, Kushnir et al. (2010) focused on the former while Ma and Xu (2013) the latter. In our study, both biased frequency (e.g., AAAAAAAB) and high sequential dependence (e.g., AAABBAAA) lead to high agency ratings, which may help unify past findings. Future work will look at how differently or similarly frequency and dependence are weighted in our nonrandomness and agency judgments.

Although violation of randomness plays an important role in agency detection, we do not claim that it is the only cue to agency or always linked to the latter. Given certain background knowledge or context[3], "irregular" stimuli may look agentive—a seemingly chaotic drip painting still looks like the work of a human, and "regular" stimuli can appear non-agentive—we would not necessarily think a mechanical watch is alive because it tick-tocks every second. Without such information, however, intentional agents are perhaps the best guess for a nonrandom outcome. Follow-up studies should take a closer look at how rich knowledge may be integrated into the way we reason about agents and randomness.

Another question is whether the current findings generalize to other types of stimuli, such as matrices, numbers, geometric shapes, *etc*. Answering this question requires us to understand and formalize randomness in these domains, on which there are increasingly more studies in recent years (e.g., Griffiths & Tenenbaum, 2007; Hsu, Griffiths, & Schreiber, 2010). We plan to investigate the agency–nonrandomness link us-

[3]This was suggested by two anonymous reviewers.

ing new types of stimuli. Even for binary sequences, it is worth looking at if what we found applies to longer sequences where more interesting regularities may emerge, such as the repeating triads in Ma and Xu's (2013) study. Also, as the length extends, it may become difficult to find a global pattern and one may have to focus on local patterns. How will these factor into our agency and nonrandomness judgments?

On the computational level, our study is the first step towards formalizing how people perceive agency. A more precise account requires us to directly estimate the probability distribution of a certain stimulus being generated by an agent, $P(X|\text{agent})$, which can be achieved by using a much larger sample size (e.g., the Big Bell Test invited more than 100,000 people from all over the world to generate random responses; see http://thebigbelltest.org/ for details) as well as applying sampling methods such as Markov chain Monte Carlo with People ("MCMCP", Sanborn, Griffiths, & Shiffrin, 2010).

MCMCP is also able to capture each person's judgment. What looks agentive to some may look inanimate to others; understanding individual differences may allow us to appreciate the complexity of human agency perception and on top of that, explain far-reaching psychological and societal consequences, such as the endorsement of Intelligent Design, the denial of natural selection in favor of creationism, and so on. In regards to other social phenomena, past studies explored the relationship between randomness and perceived efficacy of rituals (Legare, & Souza, 2014), belief in conspiracy theories (Dieguez, Wagner-Egger, & Gauvrit, 2015), *etc.*. We wonder whether people's agency intuition plays a similar or a different role in these situations, especially using the more intricate characterization of agency that MCMCP provides.

In his book *Scienceblind*, Shtulman (2017) argued that many misconceptions of science stick not just because of ideology or the media, but also because they have deep roots in our intuitive theories. As mentioned before, our agency–nonrandomness link may be one such root. For science educators, the question at hand is, how *malleable* is it? Given the right kind and amount of evidence, will we update our belief? For instance, by understanding causal mechanisms by which nonrandomness could arise from non-agentive sources (e.g., a ball rolling down a xylophone produces orderly sounds, but it is not viewed as an agent by adults or even infants, Schachner, Carey, & Kelemen, 2013), can we weaken or break this link?

Acknowledgments

We thank Shaun O'Grady for suggesting a critical change in the stimuli and Fred Callaway for helpful discussions.

References

Baillargeon, R., Scott, R., & Bian, L. (2016). Psychological reasoning in infancy. *Annual Review of Psychology*, *67*, 159–186.

Barrett, J. L. (2000). Exploring the natural foundations of religion. *Trends in Cognitive Sciences*, *4*(1), 29–34.

Dawkins, R. (1996). *The blind watchmaker: Why the evidence of evolution reveals a universe without design*. New York, NY: W. W. Norton & Company.

Dieguez, S., Wagner-Egger, P., & Gauvrit, N. (2015). Nothing happens by accident, or does it? A low prior for randomness does not explain belief in conspiracy theories. *Psychological Science*, *26*(11), 1762–1770.

Feldman, J., & Tremoulet, P. D. (2008). *The attribution of mental architecture from motion: Towards a computational theory* (Tech. Rep. No. RuCCS TR-87). Department of Psychology, Rutgers University.

Friedman, W. J. (2001). The development of an intuitive understanding of entropy. *Child Development*, *72*(2), 460–473.

Griffiths, T. L., & Tenenbaum, J. B. (2001). Randomness and coincidences: Reconciling intuition and probability theory. In J. D. Moore & K. Stenning (Eds.), *Proceedings of the 23rd annual conference of the cognitive science society* (pp. 370–375). Austin, TX: Cognitive Science Society.

Griffiths, T. L., & Tenenbaum, J. B. (2003). Probability, algorithmic complexity and subjective randomness. In R. Alterman & D. Kirsh (Eds.), *Proceedings of the 25th annual conference of the cognitive science society* (pp. 480–485). Austin, TX: Cognitive Science Society.

Griffiths, T. L., & Tenenbaum, J. B. (2004). From algorithmic to subjective randomness. In S. Thrun, L. K. Saul, & B. Schölkopf (Eds.), *Advances in neural information processing systems* (Vol. 16, pp. 953–960). Cambridge, MA: MIT Press.

Griffiths, T. L., & Tenenbaum, J. B. (2007). From mere coincidences to meaningful discoveries. *Cognition*, *103*(2), 180–226.

Hsu, A., Griffiths, T. L., & Schreiber, E. (2010). Subjective randomness and natural scene statistics. *Psychonomic Bulletin & Review*, *17*, 624–629.

Keil, F. C., & Newman, G. E. (2015). Order, order everywhere, and only an agent to think: The cognitive compulsion to infer intentional agents. *Mind and Language*, *30*(2), 117–139.

Kushnir, T., Xu, F., & Wellman, H. M. (2010). Young children use statistical sampling to infer the preferences of other people. *Psychological Science*, *21*(8), 1134–1140.

Legare, C. H., & Souza, A. L. (2014). Searching for control: Priming randomness increases the evaluation of ritual efficacy. *Cognitive Science*, *38*(1), 152–161.

Ma, L., Berthiaume, V. G., Hoch, J., & Xu, F. (under revision). Do Infants Infer Intentional Agents from the Perception of Auditory Regularity?

Ma, L., & Xu, F. (2013). Preverbal infants infer intentional agents from the perception of regularity. *Developmental Psychology*, *49*(7), 1330–1337.

Newman, G. E., Keil, F. C., Kuhlmeier, V. A., & Wynn, K. (2010). Early understandings of the link between agents and order. *Proceedings of the National Academy of Sciences*, *107*(40), 17140–17145.

Sanborn, A., Griffiths, T. L., & Shiffrin, R. M. (2010). Uncovering mental representations with Markov chain Monte Carlo. *Cognitive Psychology*, *60*(2), 63–106.

Saxe, R. J., Tenenbaum, J. B., & Carey, S. (2005). Secret agents: Inferences about hidden causes by 10- and 12-month-old infants. *Psychological Science*, *16*(12), 995–1001.

Schachner, A., Carey, S., & Kelemen, D. (2013). Inferring the causes of patterned sounds: Were those notes caused by an agent, or an inanimate force? In *The 8th biennial meeting of the cognitive development society.* Memphis, TN.

Shen, L., & Sinai, M. (2013). Geneoverlap: Test and visualize gene overlaps [Computer software manual]. Retrieved from `http://shenlab-sinai.github.io/shenlab-sinai/` (R package version 1.12.0)

Shtulman, A. (2017). *Scienceblind: Why our intuitive theories about the world are so often wrong*. New York, NY: Basic Books.

Sim, Z., & Xu, F. (2013). Infants' early understanding of coincidences. In M. Pauen, N. Sebanz, & I. Wachsmuth (Eds.), *Proceedings of the 35th annual conference of the cognitive science society* (pp. 3402–3407). Austin, TX: Cognitive Science Society.

Williams, J. J., & Griffiths, T. L. (2013). Why are people bad at detecting randomness? A statistical argument. *Journal of Experimental Psychology: Learning, memory, and cognition*, *39*(5), 1473–1490.

Wu, Y., Muentener, P., & Schulz, L. E. (2015). The invisible hand: Toddlers connect probabilistic events with agentive causes. *Cognitive Science*, *1*(23), 1–23.