bioRxiv preprint doi: https://doi.org/10.1101/061473; this version posted June 30, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Individual differences in learning behaviours in humans: Asocial exploration tendency does not predict reliance on social learning

Author names and affiliations:

Wataru Toyokawa[a, b, 1, *], Yoshimatsu Saito[b, c], Tatsuya Kameda[c]

a: Department of Behavioural Science, Hokkaido University, N10W7, Kita-ku, Sapporo, Japan 060-0810

b: Japan Society for the Promotion of Science, 5-3-1 Kojimachi, Chiyoda-ku Tokyo, Japan 102-0083

c: Department of Social Psychology, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo, Japan 113-0033

1: Present address: School of Biology, University of St Andrews, St Andrews, Fife, UK, KY16 9TH

E-mail addresses:

WT (wt25@st-andrews.ac.uk), YS (middle158725071@gmail.com), and TK (tkameda@l.u-tokyo.ac.jp)

***Corresponding author:** Wataru Toyokawa (wt25@st-andrews.ac.uk)

Word count: 5834 (excluding references), 7720 (including references)

bioRxiv preprint doi: https://doi.org/10.1101/061473; this version posted June 30, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Abstract

A number of empirical studies have suggested that individual differences in asocial exploration tendencies in animals may be related to those in social information use. However, because the 'exploration tendency' in most previous studies has been measured without considering the exploration-exploitation trade-off, it is yet hard to conclude that the animal asocial 'exploration-exploitation' tendency may be tied to social information use. Here, we studied human learning behaviour in both asocial and social multi-armed bandit tasks. By fitting reinforcement learning models including asocial and/or social decision processes, we measured each individual's (1) asocial exploration tendency and (2) social information use. We found consistent individual differences in the exploration tendency in the asocial tasks. We also found substantive heterogeneity in the adopted learning strategies in the social task: One-third of participants were most likely to have used the copy-when-uncertain strategy, while the remaining two-thirds were most likely to have relied only on asocial learning. However, we found no significant individual association between the exploration frequency in the asocial task and the use of the social learning strategy in the social task. Our results suggest that the social learning strategies may be independent from the asocial search strategies in humans.

Keywords: Social learning; individual differences; reinforcement learning; multi-armed bandit; exploration-exploitation; copy-when-uncertain

bioRxiv preprint doi: https://doi.org/10.1101/061473; this version posted June 30, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

1. Introduction

To find better behavioural options in foraging, mate choice, nest search, etc., group living animals can benefit from asocial information-gathering strategy (e.g., reinforcement learning rules (Sutton and Barto, 1998; Trimmer et al., 2012)) and from strategic use of social information (Boyd and Richerson, 1985; Laland, 2004). Although there has been much recent interest in the inter-individual variation in both asocial and social learning behaviour (Mesoudi et al., 2016; Reader, 2015), little is known about whether and (if so) how they associate with each other.

The adaptive reason of the individual differences in asocial exploration tendency might be the trade-off between exploration and exploitation. Given the limited time/energy budget, a single animal must strike the right balance between trying unfamiliar behaviours to sample information (i.e., 'exploration') versus choosing known best behaviour (i.e., 'exploitation') so as to improve the long-term net decision performance (Cohen et al., 2007; Hills et al., 2014). The optimal balance of exploration-exploitation depends on the costs and benefits of information gathering, which may differ between individuals. For example, an individual with poor information processing performance may have large costs of exploration, an individual with shorter expected life-span may benefit less from sampling more information, while an individual experiencing a temporary volatile environment may be forced to explore so as to update their knowledge (Reader, 2015).

On the other hand, the individual variation in reliance on social information might come from the balance of cost and benefit of copying others (Mesoudi et al., 2016). For instance, an individual possessing inaccurate private information will potentially incur a large cost if relying solely on the private knowledge and hence may tend to copy others more (e.g., 'copy-when-uncertain' (Laland, 2004; Rendell et al., 2011)), an individual living in a large group may benefit more from following the majority

bioRxiv preprint doi: https://doi.org/10.1101/061473; this version posted June 30, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

(King and Cowlishaw, 2007), while an individual faced with a highly volatile environment may rely more on private information due to the potentially large cost from copying an out-of-date behaviour (Aoki and Feldman, 2014).

The increasing body of empirical studies has suggested that the individual differences in the asocial exploration tendency might associate with those in the social information use (but see Webster and Laland (2015)). For instance, the individual exploration propensity negatively correlates with individual tendency of copying conspecifics in barnacle geese *Branta leucopsis* (Kurvers et al., 2010a, 2010b) and zebra finches *Taeniopygia guttata* (Rosa et al., 2012), while the opposite is true in three-spined sticklebacks *Gasterosteus aculeatus* (Nomakuchi et al., 2009) and great tits *Parus major* (Marchetti and Drent, 2000). The interesting question here would be why each of these species shows such individual correlations between asocial and social search behaviour. Several possible mechanisms may be conceivable to them. For example, environmental volatility may increase asocial exploration tendency while also decreasing copying tendency. On the other hand, a common cognitive ability underlying both asocial and social learning may generate a positive correlation between them (Mesoudi et al., 2016).

However, the term ‘exploration’ has been used rather loosely in the previous literature, and has been confounded with other personality traits (reviewed in Réale et al. (2007)), which might have contributed to somewhat incoherent previous findings on the relation between asocial exploration and social information use. Broadly speaking, more active, neophilic, or bolder individuals tend to be labelled as ‘explorative’ while more inactive, neophobic, or shyer individuals tend to be labelled as ‘unexplorative’ (Réale et al., 2007). However, it was untested whether more active individuals actually gather information more (i.e., explore more) during the learning process compared to inactive individuals. Also, from the viewpoint of the exploration-

bioRxiv preprint doi: https://doi.org/10.1101/061473; this version posted June 30, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

93 exploitation trade-off, more explorative (exploitative) individuals are not necessarily
94 bolder (shyer): Especially in a changing environment, individuals who seldom explore
95 (i.e., who exploit the same option for a long time) may also be seen as very 'bold'
96 (e.g., Carere and Locurto, 2011; Groothuis and Carere, 2005; Koolhaas et al., 1999).
97 Therefore, it is yet hard to conclude that animal asocial 'exploration-exploitation'
98 tendency may be tied to the social information use. A more clear-cut measurement of
99 exploration behaviour is needed.

100 In this study, we focused on human learning behaviour in a multi-armed bandit (MAB)
101 problem, and saw whether the individual differences in asocial exploration tendency
102 might predict the reliance on social learning. In the MAB task, individuals have
103 multiple choice options, but at the outset they do not have exact knowledge of which
104 option is the most profitable (Figure 1a). In every round, each individual has to make
105 a decision whether to exploit (i.e., choosing the option that has higher estimated
106 reward value as of that round; see Methods section) or to explore (i.e., choosing the
107 other option with lower estimated reward value). Because the MAB problem embeds
108 the exploration-exploitation trade-off in its heart (Sutton and Barto, 1998), it is a
109 suitable test bed for unambiguously measuring exploration behaviour (Daw et al.,
110 2006; Keasar et al., 2002; Racey et al., 2011; Toyokawa et al., 2014). Fitting a
111 reinforcement learning model to each participant's decision data (O'Doherty et al.,
112 2003), we quantified each participant's asocial exploration tendency.

113 In addition to the asocial situation where participants engaged in the MAB task alone
114 (hereafter, 'solitary task'), participants also played the MAB task in a pairwise
115 situation ('paired task') in which they were able to observe the other participant's
116 choice (but not the peer's earned payoff) displayed on the monitor. To examine
117 whether the participants adopted social learning strategy in the paired task, we fitted
118 several asocial- and social-learning models to each participant's decision data, and

bioRxiv preprint doi: https://doi.org/10.1101/061473; this version posted June 30, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

then selected the most likely learning model individually. Also, we analysed each participant's gaze movement measured by an eye-tracker in order to confirm the participant's information use during the task. Finally, we examined whether the exploration tendency in the solitary task (i.e., asocial exploration) might predict the use of social learning strategy in the paired task.

2. Material and methods

2-1 Participants

Fifty-six right-handed undergraduate students were randomly selected from a subject pool at Hokkaido University in Japan to participate in the experiment. Of these 56 participants, 8 participants failed at eye tracker calibration, leaving us with 48 participants (24 females; Mean Age ± S.D. = 19.0 ± 0.90) to be included in data analysis. After the experimental session, participants received monetary rewards based on their performance in the experimental tasks as compensation for their participation (mean ± S.D. = 1253 ± 18.5 JPY).

2-2 Task overview: the restless 2-armed bandit

Participants performed a restless 2-armed bandit task on a computer screen (Figure 1a). We used a '2-armed' task as the simplest case of a MAB. Each participant had to repeatedly choose between two slot machines. The participant's goal was to maximise the total reward earned over a sequence of plays by deciding which machine to play for how many times.

Each round started with a 1-second interval during which a crossbar was shown at the centre of the grey background. After this interval, two boxes (i.e., slot machines) appeared on the left and right side of the screen. Participants chose the left box or

bioRxiv preprint doi: https://doi.org/10.1101/061473; this version posted June 30, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

143 right box by pressing the 'left' or 'right' key (respectively) on a keyboard with their
144 right hand. Participants had a maximum of 2 seconds to make their choices (decision
145 interval): If no choice was made during the decision interval, a 'TIME OUT' message
146 appeared in the centre of the screen for 2.5 seconds to signal a missed round
147 (average number of missed rounds per participant was 0.44 out of 80 rounds in the
148 'first solitary task', 0.52 out of 180 rounds in the 'paired task', and 0.15 out of 80
149 rounds in the 'second solitary task'). If participants responded within 2 seconds, the
150 frame of the chosen option turned to be bold for 1.5 + (2 – *response time*) seconds
151 so as to confirm their choice, followed by a 1-second display of earned points in the
152 chosen box. After showing the rewards (or showing 'TIME OUT'), i.e., after 4.5
153 seconds from the outset of the decision interval, the next round started with a
154 crossbar.

155 Each option yielded random points (50 points = 1 JPY) from a normal probability
156 distribution unique to each box, rounded up to the next integer, or truncated to zero if
157 it would have been a negative value (although this never happened). The mean
158 values of the probabilistic payoff were different between the two options. Additionally,
159 the mean values of payoff were changing during the task (Figure 1b). The standard
160 deviations of the probabilistic payoff distributions were identical for both boxes and
161 did not change during the task (S.D. = 10). Before each task started, participants
162 were informed about the total number of rounds (80 rounds in the first and second
163 solitary tasks, respectively; 180 rounds in the paired task) as well as about the
164 possibility that the mean payoff from the options might change at some points during
165 the task. However, they were not informed of the actual value of mean payoff from
166 each slot, when or how it would actually change, or the exact rate at which the payoff
167 points would be transformed into JPY after the experiment. To confirm that our
168 results would not change in different bandits' settings (e.g., mean payoff, pattern or
169 schedule of payoff changing), we used four different settings (hereafter, conditions)

bioRxiv preprint doi: https://doi.org/10.1101/061473; this version posted June 30, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

for the paired task as a between-pair design, randomly assigned for each pair (Online Supporting Material Figure S2).

The computer-based task was constructed using Python with PsychoPy package (Peirce, 2007, 2009) with Tobii SDK 3.0 for Python. Python code used for the task is available from the corresponding author. Further details of the settings of each task are available in Online Supporting Material S1-1.

bioRxiv preprint doi: https://doi.org/10.1101/061473; this version posted June 30, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

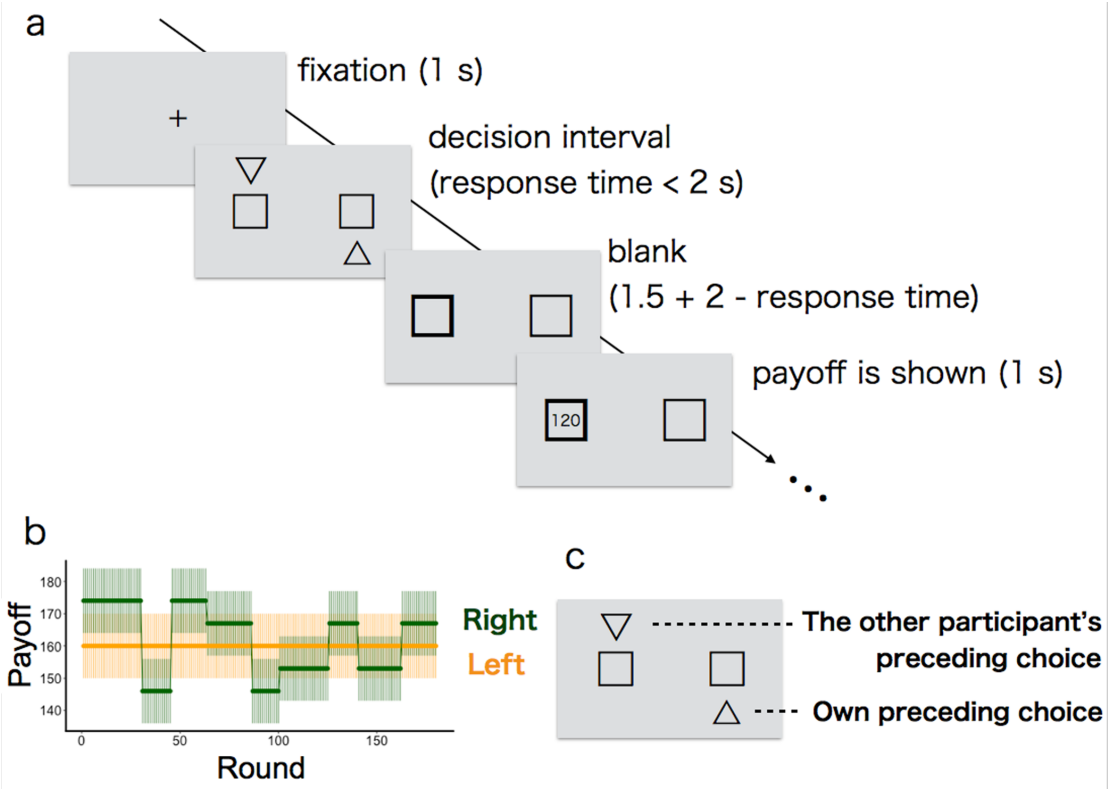


Figure 1. The restless 2-armed bandit task. (a) Illustration of the time line within a round. After fixation with a crossbar for one second, two slots (boxes) were presented. The participant had to choose one within 2 seconds (decision interval). When choosing one, the frame of the chosen option turned to be bold. Up to 3.5 seconds later, the number of payoff points earned was revealed to the participant (120 points in this example). After a further second, the next round started with a crossbar. (b) Example of mean payoffs for each option in paired task (bold lines). The payoff received for a particular choice is drawn from a Gaussian noise around each mean (shaded areas show 1 S.D. of this noise). Note that the most profitable slot ('optimal option') was switched several times due to the volatility; only one box was volatile and the other box's mean was fixed. The left-right location of the volatile box was counterbalanced across pairs in the paired task (the right box was volatile in this example). (c) Example of social information in the paired task. The other participant's choice in the preceding round was marked by a downward triangle (left in this example) while their own choice in the preceding round was marked by an upward triangle (right in this example).

bioRxiv preprint doi: https://doi.org/10.1101/061473; this version posted June 30, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

2-3 Experimental procedure

For each experimental session, two participants of the same sex, randomly selected from the participant pool, were called to the laboratory. Upon arrival, each participant was seated in a separate soundproof chamber equipped with a computer terminal and received general instructions about the experiment. After filling in the consent form (see Ethics section), they read further paper-based instructions individually. Participants remained strictly anonymous to each other throughout the experiment.

A single experimental session was composed of three tasks: (1) first solitary task, (2) paired task, and (3) second solitary task. Although participants were informed that there would be three tasks in total, they did not know any details of each task until reading the instruction about the next task during the 5-minute break between the tasks. The three tasks differed in both the mean payoffs of slots and their changing pattern over time (see Online Supporting Material S1-1 for details). Furthermore, only in the paired task were participants able to see social information—the other participant's choice made in the preceding round—in addition to their own preceding choice (Figure 1c). They were explicitly informed that both participants played the identical task so that they were able to understand that the social information could be informative. The session lasted for about 70 minutes in total.

2-4 Acquisition and processing of gaze data

We recorded each participant's gaze movement during the task at 300 Hz using Tobii TX300 eye trackers connected to 23-inch monitors (Tobii Technology, Stockholm, Sweden). Participants were seated with their head positioned on a chinrest, 70 cm in front of the screen. Before each task, the calibration of the eye-tracker was validated with five fixation points and re-calibrated if needed.

bioRxiv preprint doi: https://doi.org/10.1101/061473; this version posted June 30, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

We used Tobii's default setting of noise reduction and fixation classification to process the gaze data. Because we were interested in participants' information use before making each decision, we focused upon the gaze positions in the 1-second time window before each choice was made. Note that, because there was a 1-second interval between rounds, the 1-second time window did not include gaze positions in the previous round. The fixation positions of gazes were classified into the following five categories: *gazing-at-social-information*, *gazing-at-own-preceding-choice*, *gazing-at-the-left-box*, *gazing-at-the-right-box*, and *others* (see Online Supporting Material S1-2 for the gaze data processing). All eye-tracking data and Python code used in the gaze data processing are available from the corresponding author upon request.

2-5 Testing exploration tendency

In order to classify each choice in the solitary tasks as either exploration or exploitation, we fitted a standard reinforcement learning model to each participant's choice data using maximal likelihood estimation method (Daw et al., 2006; O'Doherty et al., 2003). Following the previous empirical studies on social learning strategies in humans (McElreath et al., 2005, 2008), our learning model consists of two parts. First, the values for choosing the options (i.e., Q-values (Sutton and Barto, 1998)) are recursively updated by experienced reward according to the Rescorla-Wagner rule (Trimmer et al., 2012) that is commonly used in the animal learning literature. Second, the Q-values are transformed to the choice probability for each option by the 'softmax' choice policy (Daw et al., 2006). The mathematical expression of this asocial learning model is shown in the next sub-section (Eq.1 and Eq.2; see Online Supporting Material S1-3 for full details of the model fitting procedure).

After obtaining the most likely parameters of the asocial learning model separately for each participant, we classified each choice of each participant as either exploitative

bioRxiv preprint doi: https://doi.org/10.1101/061473; this version posted June 30, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

or explorative according to whether the chosen box had the larger Q-value (exploitation) or the smaller Q-value (exploration) between the two options. Then, we summed up the number of explorations for each participant for each of the solitary tasks. We calculated Pearson's correlation coefficient between the numbers of explorations in the first and second solitary tasks so as to examine the individual consistency in exploration tendency.

2-6 Determining learning strategies

We examined four different learning models as candidates of participants' strategies in the paired task: *asocial learning model* (AL), *unconditional-copying model* (UNC), *copy-when-uncertain model* (CWU), and *random choice model* (Random). AL and Random are asocial, while UNC and CWU are social learning models. Since each participant interacted with only one anonymous peer without seeing the peer's earned payoff, we could not consider any 'frequency based' strategies (e.g., copy the majority) and 'model based' strategies (e.g., copy the prestigious individual or payoff-based copying) (Rendell et al., 2011). We fitted each of the models to each participant's choice data individually.

All models have the same updating rule for Q-values, called the Rescorla-Wagner rule. Q-value for the left option is updated as follows:

$$Q_{t+1,l} \leftarrow (1-\alpha)Q_{t,l} + \alpha\pi_{t,l}, \qquad \text{(Eq.1)}$$

where α $(0 \leq \alpha \leq 1)$ is a parameter determining the weight given to new experience (learning rate) and $\pi_{t,l}$ is the amount of payoff obtained from choosing the left option (l) in round t. When the left option was not chosen in round t, the left option's Q-value was not updated ($Q_{t+1,l} = Q_{t,l}$). The same updating rule applied to the right option. We set $Q_{0,l} = Q_{0,r} = 0$ because there was no reason to expect participants to have any prior preference for either option at the outset (McElreath et al., 2008).

bioRxiv preprint doi: https://doi.org/10.1101/061473; this version posted June 30, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Of course, we could have considered any other updating rules (e.g., Bayesian updating model (Payzan-LeNestour and Bossaerts, 2011)). However, because our scope here is examining whether the participants might conduct social learning or not, rather than quantifying detailed computational learning algorithms, we decided to focus only on the simple Rescorla-Wagner rule which has been shown to be evolutionary adaptive across a broad range of environmental conditions (Trimmer et al. 2012).

The Q-values were transformed into the choice probability in different ways between the four models, shown as follows.

2-6-1 Asocial learning model (AL)

In the asocial learning model, the probability of choosing the left option in round $t+1$ is given by the following softmax rule:

$$P_{t+1,l} = \frac{\exp\left(Q_{t,l}/\beta\right)}{\exp\left(Q_{t,l}/\beta\right)+\exp\left(Q_{t,r}/\beta\right)}, \qquad \text{(Eq.2)}$$

where β ($\beta > 0$) is a parameter that measures the influence of the difference between Q-values on choice ('temperature' parameter). As $\beta \to \infty$, choice is completely random (i.e., $P_{t+1,l} = 1/2$; highly explorative). As $\beta \to 0$, choice becomes deterministic, in favour of the option with the higher Q-value (i.e., highly exploitative). Therefore, β regulates the individual's asocial exploration tendency.

This asocial learning model was also used in the last subsection (section 2-5) to estimate the exploration tendencies in the solitary tasks.

2-6-2 Unconditional-copying model (UNC)

Next we considered unconditional/unselective copying, the simplest case of social information use (Laland, 2004). The individuals copied the other participant's choice

bioRxiv preprint doi: https://doi.org/10.1101/061473; this version posted June 30, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

296 at the fixed rate λ ($0 \leq \lambda \leq 1$), otherwise their decisions were determined by the
297 softmax rule (Eq. 2). In the case that the peer chose the left option in round t, the
298 focal participant's probability of choosing the left option at round $t+1$ is given by:

$$P_{t+1,l} = (1-\lambda)\frac{\exp(Q_{t,l}/\beta)}{\exp(Q_{t,l}/\beta)+\exp(Q_{t,r}/\beta)} + \lambda, \quad \text{(Eq.3-1)}$$

299 whereas when the peer chose the right option in round t, the participant's probability
300 of choosing the left option at round $t+1$ is given by:

$$P_{t+1,l} = (1-\lambda)\frac{\exp(Q_{t,l}/\beta)}{\exp(Q_{t,l}/\beta)+\exp(Q_{t,r}/\beta)}. \quad \text{(Eq.3-2)}$$

301 When $\lambda = 1$, the individuals always copy the peer's choice. When $\lambda = 0$, the
302 individuals rely only on asocial learning. We assumed that the individuals also relied
303 only on the asocial learning rule when the peer missed the preceding round (i.e.,
304 when no choice was made by the peer).

305 ***2-6-3 Copy-when-uncertain model (CWU)***

306 A number of studies have suggested that animals are selective in timing to use social
307 information, depending on the degree of uncertainty that they are experiencing (e.g.,
308 Coolen et al., 2003; Galef, 2009; van Bergen et al., 2004). To quantify the uncertainty
309 level concerning which of the two slots is more rewarding at a given round, we used
310 absolute difference in Q-values: The closer the Q-values between the two options,
311 the higher the uncertainty. In the copy-when-uncertain model, when the peer chose
312 the left option in round t, the focal participant's choice probability for the left option at
313 $t+1$ is given by:

$$P_{t+1,l} = (1-\gamma H_t)\frac{\exp(Q_{t,l}/\beta)}{\exp(Q_{t,l}/\beta)+\exp(Q_{t,r}/\beta)} + \gamma H_t, \quad \text{(Eq.4-1)}$$

314 whereas when the peer chose the right option in round t,

bioRxiv preprint doi: https://doi.org/10.1101/061473; this version posted June 30, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

$$P_{t+1,l} = (1 - \gamma H_t)\frac{\exp(Q_{t,l}/\beta)}{\exp(Q_{t,l}/\beta)+\exp(Q_{t,r}/\beta)}, \quad \text{(Eq.4-2)}$$

where

$$H_t = -\sum_k \left(\frac{Q_{t,k}}{Q_{t,l}+Q_{t,r}}\right) log_2 \left(\frac{Q_{t,k}}{Q_{t,l}+Q_{t,r}}\right). \quad \text{(Eq.4-3)}$$

γ ($0 \leq \gamma \leq 1$) is a parameter that determines the upper limit of copying probability, and H_t (i.e., information entropy, $0 \leq H_t \leq 1$) determines the actual copying rate at round t, where $k \in \{l, r\}$. When both Q-values are equal ($Q_l = Q_r$), the uncertainty becomes max ($H = 1$), leading the copying probability to be maximal (i.e., the individual copies the peer at the probability γ when uncertainty is the highest). As the difference between Q-values becomes larger, H approaches to 0, with the result that the choice is mostly determined by the asocial learning.

2-6-4 Random choice model

We also considered the case that the choices are made randomly at fixed rate regardless of the Q-values, so as to verify that the participants did not behave just randomly in the experiment. The choice probability for the left option is always a fixed rate ε ($0 \leq \varepsilon \leq 1$); hence, the probability of choosing the right option is $1 - \varepsilon$.

2-7 Model selection

We used Bayesian Model Selection (BMS) (Rigoux et al., 2014; Stephan et al., 2009) that estimates how likely it is that each learning model generated the data of a *randomly* chosen participant (see Online Supporting Material S1-4). Using BMS, we were also able to calculate the probability that each model had generated a *given* participant's data (Stephan et al., 2009). Because here we were interested in inter-individual variation of social information use, we mainly focused on the latter 'individual-level' probabilities for each model, rather than the former 'group-level'

bioRxiv preprint doi: https://doi.org/10.1101/061473; this version posted June 30, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

probability that BMS was originally aimed at. To verify this BMS estimation for each participant's learning model, we also conducted the standard AIC comparison (see Online Supporting Material S1-5).

2-8 The relation between gaze movement and learning strategy

We examined the gaze movement so as to confirm the participants' information use during the paired task. We focused on '*gazing-at-social-information before choice'* (yes = 1/no = 0) for each round as a binary response variable in a binomial generalised linear mixed model (GLMM), including the following three random effects: individuals, pairs, and bandit's conditions. We considered estimated learning strategy (social = 1/asocial = 0), information uncertainty (i.e., similarity between Q-values [Eq. 4-3]), rounds, and possible 2-way interactions as fixed effects. Model selection was done based on each AIC value (Online Supporting Material S1-6).

2-9 The relations between asocial exploration and social learning

To examine whether the asocial exploration tendency might predict the use of social learning strategy, we analysed a binomial GLMM with social = 1/asocial = 0 learning as a binary response variable, with random effects of pairs and bandit's conditions. We considered the asocial exploration tendency (frequency of explorative choices in the first solitary task), asocial learning performance (average payoff earned per choice in the first solitary task), and possible 2-way interactions as fixed effects. Model selection was done based on each AIC value (Online Supporting Material S1-7).

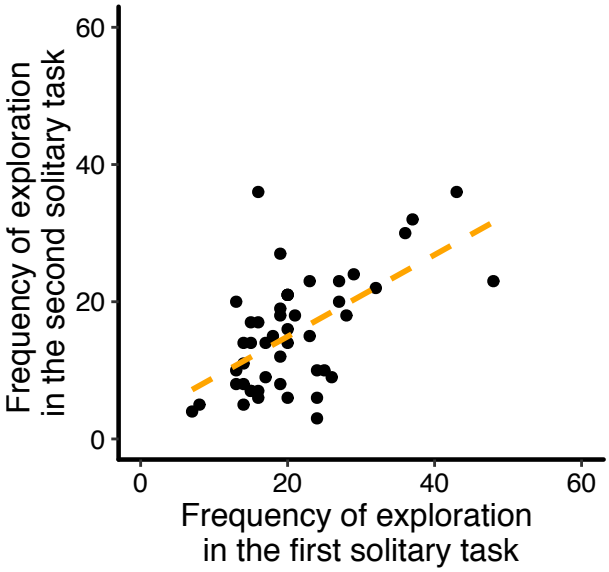
3. Results

bioRxiv preprint doi: https://doi.org/10.1101/061473; this version posted June 30, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

359 ***3-1 Individual consistency in the asocial exploration tendency***

360 The frequency of explorative choices in the first solitary task was positively correlated
361 with that in the second solitary task (r = 0.58, p < 0.001; Figure 2). This result
362 indicates that participants exhibited a stable asocial exploration tendency across the
363 two solitary tasks. Results of the fitting of the asocial learning model are shown in
364 Online Supporting Result S2-1.

365



366

367 **Figure 2. Individual consistency in the asocial exploration frequency.** The x-axis
368 shows the frequency of explorative choices in the first solitary task. The y-axis shows
369 the frequency of explorative choice in the second solitary task. The dotted line is a
370 linear regression.

371

372 ***3-2 Detecting social learning strategies***

373 The Bayesian Model Selection (BMS) resulted in the heterogeneous distribution of
374 learning strategies among the participants (Figure 3a). The estimated likelihood of

bioRxiv preprint doi: https://doi.org/10.1101/061473; this version posted June 30, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

each model for the whole data was 0.47 for the Asocial Learning model (AL), 0.10 for the Unconditional Copying model (UNC), 0.41 for the Copy-when-uncertain model (CWU), and 0.02 for the Random model.

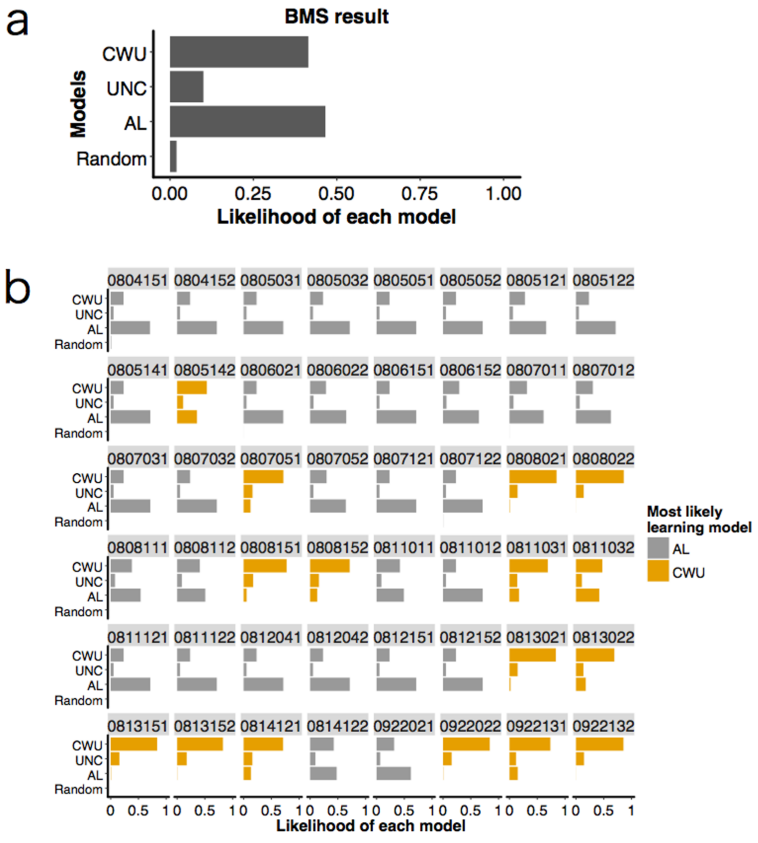
Focusing on each individual's probabilities for each model, the BMS revealed that 32 participants were most likely to have adopted AL, while the remaining 16 participants were most likely to have adopted CWU (Figure 3b). No individuals were most likely to have used UNC or Random. We hereafter call the former 32 individuals 'asocial learners' and the latter 16 individuals 'social learners' for simplicity. We also checked each model's AIC values for each participant and confirmed that the result was not qualitatively changed (although three participants were most likely to have adopted the UNC model instead; see Online Supporting Material S2-2 and Figure S6).

Table 1 shows means (± 1 SDs) of best-fitted parameters for asocial and social learners. The mean value of γ (i.e., maximum copying probability) for social learners (CWU) was 0.162 (± 0.072), which means that, when uncertainty level was max (i.e., setting $Q_l = Q_r$ so that $H = 1$ and $\exp(Q_t/\beta)/\sum \exp(Q_t/\beta) = 0.5$, see Eq. 4), there was about 58% chance of choosing the same option as chosen by the other participant in the preceding round. Note that asocial learners should choose whichever option by 50% chance (i.e., randomly) when they have the maximum uncertainty. Therefore, there was at most 8% increase in the probability of "copying others' choice" by social learners as compared to asocial learners. Together with the smaller frequency of social learners (N = 16) than asocial learners (N = 32), we will revisit this rarity of social learning in Discussion.

bioRxiv preprint doi: https://doi.org/10.1101/061473; this version posted June 30, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

	Asocial Learning (AL)	*Copy-when-uncertain (CWU)*
α (learning rate)	0.842 ± 0.187	0.787 ± 0.192
β (temperature)	11.31 ± 10.39	6.03 ± 2.03
γ (maximum copying probability)	N.A.	0.162 ± 0.0720
N	32	16

Table 1. Fitted parameter values (mean ± 1 S.D.) that provide the maximum likelihood for asocial learners (N = 32) and for social (copy-when-uncertain) learners (N = 16).



bioRxiv preprint doi: https://doi.org/10.1101/061473; this version posted June 30, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Figure 3. Results of Bayesian Model Selection. We considered the following four models: random choice model, asocial learning model (AL), unconditional copying model (UNC), and copy-when-uncertain model (CWU). (a) Distributions of probabilities concerning how likely it is that each model was adopted by a randomly chosen participant (i.e., group-level model likelihood). (b) Likelihood of each model for each participant (i.e., individual-level model likelihood). The numbers shown above each cell indicate participants' IDs. Thirty-two participants were most likely to have used the asocial learning strategy (shown in grey); the remaining 16 participants were most likely to have used the copy-when-uncertain strategy (shown in orange).

3-3 Social learning did not improve the performance

Given the result that some participants used social learning, an interesting question is whether social learners performed better than asocial learners in the paired task. We compared each individual's average payoff earned per choice using a Gaussian-GLMM with a fixed effect of learning type (asocial = 0/social = 1), and with two random effects of pairs and bandit's conditions. The fixed effect was not significant, implying that social learners performed no better than asocial learners (*coefficient* [95% confidence interval] = 0.38 [-0.33, 1.12], $p = 0.28$; see Online Supporting Figure S8).

3-4 The relation between gaze movement and learning strategy

The result of BMS could be argued as an artefact of the limited number of models considered. To confirm that estimated 'social learners' actually saw the social information, we also examined the participants' gaze patterns. We used the estimated learning strategy as a dummy fixed effect (i.e., asocial learner = 0; social learner = 1). The binomial-GLMM for the probability of looking at social information revealed the significant effects of uncertainty (coefficient [odds ratio] = -0.63 [0.53], $p = 0.036$), round (coefficient [odds ratio] = 0.24 [1.25], $p < 0.001$), the interaction

bioRxiv preprint doi: https://doi.org/10.1101/061473; this version posted June 30, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

between uncertainty and learning type (coefficient [odds ratio] = 1.45 [4.25], p = 0.0028), and the intercept (coefficient [odds ratio] = -3.25 [0.039], p < 0.001). However, the fixed effect of learning strategy itself was not significant (coefficient [odds ratio] = -0.65 [0.52], p = 0.37). The confidence intervals are shown in Online Supporting Table S2.

The significant negative effect of uncertainty suggests that the participants tended not to look at the social information when uncertainty (here defined as the closeness of the two options in terms of Q-values) was high. However, the significant positive interaction between uncertainty and learning strategies suggests that, when uncertainty was high, the social learners tended to look at the social information more frequently than asocial learners. This interaction effect is thus consistent with the behavioural-choice pattern predicted by the CWU model as compared to the AL model.

3-5 The relation between asocial exploration tendency and social learning strategy

Our results showed that (1) there were consistent individual differences in asocial exploration tendency across the two solitary tasks and that (2) use of social learning was heterogeneous between individuals. Given this, we finally examined whether the asocial exploration tendency in the solitary task might predict the use of social learning strategy in the paired task. We analysed a binomial GLMM that predicts the use of social learning strategy in the paired task (Online Supporting Material S1-7).

The selected GLMM contains both fixed effects of exploration frequency at the first solitary task and performance at the first solitary task. However, none of them were statistically significant (exploration tendency: coefficient [odds ratio] = -1.21 [0.30], p = 0.16, performance: coefficient [odds ratio] = -1.05 [0.35], p = 0.07). Hence, we cannot say either of them predicts a participant's reliance on social learning. Looking

bioRxiv preprint doi: https://doi.org/10.1101/061473; this version posted June 30, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

456 at the scatter plots (Online Supporting Material Figure S8) rather than just
457 considering p-values of the GLMM parameters, however, the average performance of
458 the social learners in the first solitary task was not higher than the expected
459 performance from completely random choices (Student's t-test: t_{15} = 0.64, p = 0.53),
460 while that of asocial learners was better than the chance-level (Students' t-test: t_{31} =
461 6.15, p < 0.001; Figure S8a). It might suggest that the asocial learning performance
462 in the first solitary task has a negative effect on the use of social learning in the
463 paired task. On the other hand, the frequencies of explorations in the first solitary
464 task were not different between asocial and social learners (Figure S9).

465

466 **4. Discussion**

467 In this study, we investigated human search strategies in the asocial/social 2-armed
468 bandit tasks, respectively, and tested whether the individual differences in asocial
469 exploration tendency in isolated settings might predict the use of social learning in
470 group settings.

471 Across the first and second solitary tasks, our results showed the consistent
472 individual differences in asocial exploration tendency (Figure 2). Since participants
473 were not informed how the environmental change would occur in advance, it was
474 virtually impossible for them to calculate the optimal exploration schedule (Gittins et
475 al., 2011). Therefore, individual differences in calculation abilities are unlikely to
476 explain this individual variation. Instead, it is known that human exploration tendency
477 in a learning/decision task may be dependent on dopaminergic functions in the
478 prefrontal cortical region of the brain (Daw et al., 2006; Frank et al., 2009), which is
479 associated with tracking informational uncertainty (Yoshida and Ishii, 2006). Sensing
480 decision uncertainty accurately is important in deciding when to explore in multi-

bioRxiv preprint doi: https://doi.org/10.1101/061473; this version posted June 30, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

armed bandit problems (Cohen et al., 2007). The individual differences in asocial exploration tendency shown in our result may be related to such individual differences in neural activity for sensing uncertainty, although we did not investigate direct neural mechanisms here.

As for the social task, we found a heterogeneous distribution of learning strategies: Two-thirds of the participants seems to have used the *asocial learning* strategy while the rest seems to have used the *copy-when-uncertain* strategy (Figure 3b), even though a few of them might have adopted the *unconditional-copying* strategy instead (Online Supporting Figure S6). The copy-when-uncertain type of social information use is predicted by adaptive evolutionary models (e.g., Boyd and Richerson, 1988), and has been repeatedly reported in empirical studies about non-human animals (e.g., Galef et al., 2008; Kendal et al., 2009) as well as humans (e.g., Kameda and Nakanishi, 2002; Morgan et al., 2011; Muthukrishuna et al., 2015). Our result replicated those findings. However, in the previous empirical studies, uncertainties were externally manipulated by, for example, varying the task difficulty (Morgan et al., 2011), changing the option numbers (Muthukrishuna et al., 2015), or changing costs required for getting asocial information (Kameda and Nakanishi, 2002). Instead of fixing such parameters via experimental manipulations, our copy-when-uncertain model tracks each individual's internal information uncertainty (i.e., entropy) that can dynamically change during the course of learning. Our result thus provides a finer picture about possible plasticity in human learning strategies, shedding light on the individual flexibility of social information use during the course of problem solving.

It is noteworthy that, however, there were other forms of uncertainty that we did not consider here. For example, there were uncertainties concerning how noisily the payoff would be generated from a slot machine (i.e., variance of the payoff distribution), how often the environment would change (i.e., 'unexpected uncertainty'),

bioRxiv preprint doi: https://doi.org/10.1101/061473; this version posted June 30, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

and what kind of distribution payoff was generated from (i.e., 'structural uncertainty' or ambiguity) (Cohen et al., 2007; Payzan-LeNestour and Bossaerts, 2011; Payzan-LeNestour et al., 2013). Further study is needed to investigate how those different types of uncertainty may affect the timing of social information use.

Technically, we used AIC as an approximation to the log-evidence for the Bayesian Model Selection (see Online Supporting Method). Compared to other forms of calculation (e.g., Free Energy approximation), AIC tends to prefer complex models because of the weak penalty for having more parameters (Penny, 2012; Stephan et al., 2009). It might thus have caused over-estimation of the frequency of social learners because our social learning models have more parameters than do asocial models. Nevertheless, the gaze movement patterns were consistent with the behavioural-choice pattern from the BMS result—individuals categorised as copy-when-uncertain learners saw the social information more often than asocial learning individuals when they faced high uncertainty (Result 3-4). Importantly, the BMS only considered the behavioural (i.e., choice) data, which was measured independently from the eye-tracking. Therefore, the result confirmed that we successfully captured the significant pattern of individual differences in learning strategies.

Although evolutionary theory generally tends to suggest heavy reliance on social learning in a broader range of situations (Rendell et al., 2011), only one-third of participants seemed to have used social information in our experiment. In addition to the low prevalence of social learners among participants, the maximum copying probability of social learners was also low (at most 8% higher than that of asocial learners). This rarity of social learning might be because social learning was not so useful in the paired task (Result 3-3). Social information is useful if and only if others' behaviours can filter better options (Rendell et al., 2010). In our paired task, however, there was only one other individual playing the same task. Thus, the filtering effect

bioRxiv preprint doi: https://doi.org/10.1101/061473; this version posted June 30, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

533 from social learning was minimal. Also, there might be no 'worth-copying' peer (e.g.,
534 expert or veteran) because both participants started the task at the same time: When
535 a focal participant was naïve to the task, so was the other participant. Therefore,
536 information about the peer's choice might not be so accurate compared to the own
537 (asocial) learning experiences. Additionally, a recent empirical study suggests that
538 the reliance on social information becomes stronger as the number of choice options
539 increases (Muthukrishuna et al., 2015). Having only two options, therefore, our
540 current study might underestimate the potential use of social learning. Further studies
541 are needed to investigate whether social information use may change with group size
542 and/or number of options.

543 We also explored the possible associations between behaviour in the solitary task
544 and social learning in the paired task. Different from the previous empirical findings
545 about non-human animals, we found no relation between asocial exploration
546 tendency and the reliance on social learning. One possible reason for this difference
547 might come from the inconsistent definition of 'exploration' in the previous literature.
548 As described in Introduction, asocial exploration tendency has often been
549 confounded with other personality traits that could relate to more general asocial
550 learning ability (Réale et al., 2007). Better asocial learners may show less social
551 information use because they possess more accurate private information (e.g.,
552 Kurvers et al., 2010a, 2010b), while the opposite (better asocial learners rely more on
553 social learning) could also be plausible if both asocial and social learning reflect a
554 common basic cognitive ability (Mesoudi et al., 2016). Indeed, a number of studies
555 have shown that asocial learning ability correlates with the use of social information
556 (Bouchard et al., 2007; Katsnelson et al., 2011; Mesoudi, 2011). Although it was not
557 statistically significant, our result might also suggest that better asocial learners tend
558 to ignore social information (Result 3-5; Figure S8a). Importantly, in our task, the
559 asocial exploration tendencies were not correlated with the asocial learning

bioRxiv preprint doi: https://doi.org/10.1101/061473; this version posted June 30, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

performance (first solitary task: r = -0.20, p = 0.17; second solitary task: r = -0.14, p = 0.35; Online Supporting Figure S7). Our results that the asocial learning performance might have a negative effect on the use of social learning (but not exploration tendency) may suggest the importance of drawing a distinction between information-gathering behaviour (e.g., exploration tendency) and learning performance.

Overall, our study supports that humans are very selective about when to use social information. On the other hand, we should also acknowledge that our experimental set-up did not allow considering any 'copy-from-whom' (Heyes, 2015; Laland, 2004; Rendell et al., 2011) types of strategies, which might cause overlooking the potential social information use. Comparing 'who' strategies with 'when' strategies in the same framework will provide more comprehensive understanding on social learning for both human and non-human animals. We also believe that the computational learning model can be a strong tool for quantitative empirical investigations on animal social learning strategies (Daw et al., 2006; McElreath et al., 2005, 2008; O'Doherty et al., 2003; Sutton and Barto, 1998).

5. Ethics

This study was approved by the Institutional Review Board of the Centre for Experimental Research in Social Science at Hokkaido University (No. H26-01). Written informed consent was obtained from all participants before beginning the task.

6. Data accessibility

Behavioural data are available in Online Supporting Materials

bioRxiv preprint doi: https://doi.org/10.1101/061473; this version posted June 30, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

583

584 7. Competing interest

585 We have no competing interest.

586

587 8. Authors' contributions

588 WT designed the study, performed the experiment, carried out the data analysis and
589 drafted the manuscript; YS carried out the acquisition of gaze data and participated in
590 the data analysis; TK participated in the design of the study, supervised the study
591 and participated in writing the manuscript. All authors gave final approval for
592 publication.

593

594 9. Acknowledgments

595 We are grateful to Shinsuke Suzuki for valuable discussions related to the
596 computational learning models, and Mike Webster for helpful comments on earlier
597 draft of this manuscript.

598

599 10. Funding

600 JSPS KAKENHI Grant Number 25245063 and 25118004 to Tatsuya Kameda, and
601 JSPS KAKENHI grant-in-aid for JSPS fellows 24004583 to Wataru Toyokawa. The
602 funders had no role in design study, data collection and analysis, decision to publish,
603 or preparation of the manuscript.

bioRxiv preprint doi: https://doi.org/10.1101/061473; this version posted June 30, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

11. References

Aoki K, Feldman MW. 2014 Evolution of learning strategies in temporally and spatially variable environments: A review of theory. *Theor. Popul. Biol.* 91, 3-19. (doi:10.1016/j.tpb.2013.10.004)

Bouchard J, Goodyer W, Lefebvre L. 2007 Social learning and innovation are positively correlated in pigeons (*Columba livia*). *Anim. Cogn*. 10, 259–266. (doi: 10.1007/s10071-006-0064-1)

Boyd R, Richerson PJ. 1985 *Culture and the evolutionary process*. Chicago, IL: University of Chicago Press.

Boyd R, Richerson PJ. 1988 An evolutionary model of social learning: The effects of spatial and temporal variation. In T. Zentall & B. G. Galef (Eds.), *Social learning: A psychological and biological approach* (pp. 29-48). Hillsdale, NJ: Erlbaum.

Carere C, Locurto C. 2011. Interaction between animal personality and animal cognition. *Curr. Zool*, 57(4), 491-498.

Cohen JD, McClure SM, Yu AJ. 2007 Should I stay or should I go? How the human brain manages the trade-off between exploitation and exploration. *Phil. Trans. Roy. Soc. B.* 362, 933–942. (doi: 10.1098/rstb.2007.2098)

Coolen I, van Bergen Y, Day RL, Laland KN. 2003 Species difference in adaptive use of public information in sticklebacks. *Proc. Roy. Soc. B*. 270, 2413-2419. (doi: 10.1098/rspb.2003.2525)

bioRxiv preprint doi: https://doi.org/10.1101/061473; this version posted June 30, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

626 Daw ND, O'Doherty JP, Dayan P, Seymour B, Dolan RJ. 2006. Cortical substrates
627 for exploratory decisions in humans. *Nature* 441, 876–879.
628 (doi:10.1038/nature04766)

629 Frank MJ, Doll BB, Oas-Terpstra J, Moreno F. 2009 Prefrontal and striatal
630 dopaminergic genes predict individual differences in exploration and
631 exploitation. *Nat. Neurosci*. 12, 1062–1068. (doi:10.1038/nn.2342)

632 Galef, B. G. 2009 Strategies for Social Learning: Testing Predictions from Formal
633 Theory. *Adv. Study Behav*. 39, 117-151. (doi: 10.1016/S0065-
634 3454(09)39004-X)

635 Galef BG, Dudley KE, Whiskin EE. 2008 Social learning of food preferences in
636 'dissatisfied' and 'uncertain' Norway rats. *Anim. Behav*., 75(2), 631-637.
637 (doi:10.1016/j.anbehav.2007.06.024)

638 Gittins J, Glazebrook K, Weber R. 2011 *Multi-armed Bandit Allocation Indices 2nd*
639 *Edition*. Sussex: John Wiley & Sons, Inc.

640 Groothuis TG, Carere C. 2005 Avian personalities: characterization and epigenesis.
641 *Neuroscience & Biobehavioral Reviews*, 29(1), 137-150.

642 Heyes C. 2015 Who Knows? Metacognitive Social Learning Strategies. *Trends. Cog*.
643 *Sci.* 20, 204–213. (doi: 10.1016/j.tics.2015.12.007)

644 Hills TT, Todd, PM, Lazer D, Redish AD, Couzin ID, CSR Group. 2014 Exploration
645 versus exploitation in space, mind, and society. *Trends Cog. Sci.* 19, 46-54.
646 (doi: http://dx.doi.org/10.1016/j.tics.2014.10.004)

647 Kameda T, Nakanishi D. 2002 Cost-benefit analysis of social/cultural learning in a
648 nonstationary uncertain environment: an evolutionary simulation and an

bioRxiv preprint doi: https://doi.org/10.1101/061473; this version posted June 30, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

649 experiment with human subjects. *Evol. Hum. Behav*. 23, 373–393.
650 (doi:10.PiiS1090-5138(02)00101-0)

651 Katsnelson E, Motro U, Feldman MW, Lotem A. 2011. Individual-learning ability
652 predicts social- foraging strategy in house sparrows. *Proc. Roy. Soc. B* 278,
653 582–589. (doi: 10.1098/rspb.2010.1151)

654 Keasar T, Rashkovich E, Cohen D, Shmida A. 2002 Bees in two-armed bandit
655 situations: foraging choices and possible decision mechanisms. *Behav. Ecol.*
656 13, 757–765. (doi: 10.1093/beheco/13.6.757)

657 Kendal RL, Kendal JR, Hoppitt W, Laland KN. 2009 Identifying Social Learning in
658 Animal Populations: A New 'Option-Bias' Method. *PLoS ONE* 4, e6541.
659 (doi:10.1371/journal.pone.0006541)

660 King AJ, Cowlishaw G. 2007 When to use social information: the advantage of large
661 group size in individual decision making. *Bio. Lett.* 3, 137-139. (doi:
662 10.1098/rsbl.2007.0017)

663 Koolhaas JM, Korte SM, De Boer SF, Van Der Vegt BJ, Van Reenen CG, Hopster H,
664 De Jong IC, Ruis MAW, Blokhuis HJ. 1999 Coping styles in animals: current
665 status in behavior and stress-physiology. *Neuroscience & Biobehavioral*
666 *Reviews*, 23(7), 925-935.

667 Kurvers RHJM, Prins HHT, van Wieren SE, van Oers K, Nolet BA, Ydenberg RC.
668 2010a The effect of personality on social foraging: shy barnacle geese
669 scrounge more. *Proc. R. Soc. B.* 277: 601–608 (doi: 10.1098/rspb.2009.1474)

670 Kurvers RHJM, van Oers K, Nolet BA, Jonker RM, van Wieren SE, Prins HHT,
671 Ydenberg RC. 2010b Personality predicts the use of social information. *Ecol.*
672 *Lett.* 13: 829–837 (doi: 10.1111/j.1461-0248.2010.01473.x.)

bioRxiv preprint doi: https://doi.org/10.1101/061473; this version posted June 30, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

673 Laland KN. 2004 Social learning strategies. *Learn. Behav*. 32, 4–14.
674 (doi:10.3758/BF03196002)

675 Marchetti C, Drent PJ. 2000 Individual differences in the use of social information in
676 foraging by captive great tits. *Anim. Behav*. 60, 131–140.
677 (doi:10.1006/anbe.2000.1443)

678 McElreath R, Lubell M, Richerson PJ, Waring TM, Baum W, Edsten E, Efferson C,
679 Paciotti B. 2005 Applying evolutionary models to the laboratory study of social
680 learning. *Evol. Hum. Behav*. 26, 483–508.
681 (doi:10.1016/j.evolhumbehav.2005.04.003)

682 McElreath R, Bell AV, Efferson C, Lubell M, Richerson PJ, Waring T. 2008 Beyond
683 existence and aiming outside the laboratory: Estimating frequency-dependent
684 and pay-off-biased social learning strategies. *Phil. Trans. Roy. Soc. B* 363,
685 3515–3528. (doi: 10.1098/rstb.2008.0131)

686 Mesoudi A. 2011 An experimental comparison of human social learning strategies:
687 payoff-biased social learning is adaptive but underused. *Evol. Hum. Behav*.
688 32, 334-342. (doi: 10.1016/j.evolhumbehav.2010.12.001)

689 Mesoudi A, Chang L, Dall SRX, Thornton A. 2016 The Evolution of Individual and
690 Cultural Variation in Social Learning. *Trends. Ecol. Evol*. (doi:
691 http://dx.doi.org/10.1016/j.tree.2015.12.012)

692 Muthukrishuna M, Morgan TJH, Henrich J. 2015 The when and who of social
693 learning and conformist transmission. *Evol. Hum. Behav*. 37, 10-20.
694 (doi:10.1016/j.evolhumbehav.2015.05.004)

bioRxiv preprint doi: https://doi.org/10.1101/061473; this version posted June 30, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

695 Morgan TJH, Rendell LE, Ehn M, Hoppitt W, Laland KN. 2011 The evolutionary basis
696 of human social learning. *Proc. R. Soc. B* 279, 653–662. (doi:10.
697 1098/rspb.2011.1172)

698 Nomakuchi S, Park PJ, Bell MA. 2009 Correlation between exploration activity and
699 use of social information in three-spined sticklebacks. *Behav. Ecol.* 20, 340–
700 345. (doi: 10.1093/beheco/arp001)

701 O'Doherty JP, Dayan P, Friston K, Critchley H, Dolan, RJ. 2003 Temporal difference
702 models and reward-related learning in the human brain. *Neuron* 38, 329–
703 337. (doi:10.1016/S0896-6273(03)00169-7)

704 Payzan-LeNestour E, Bossaerts P. 2011 Risk, Unexpected Uncertainty, and
705 Estimation Uncertainty: Bayesian Learning in Unstable Settings. *PLoS*
706 *Comput. Biol.* 7, e1001048. (doi:10.1371/journal.pcbi.1001048)

707 Payzan-LeNestour E, Dunne S, Bossaerts P, O'Doherty, JP. 2013 The Neural
708 Representation of Unexpected Uncertainty during Value-Based Decision
709 Making. *Neuron* 79, 191-201 (doi:
710 http://dx.doi.org/10.1016/j.neuron.2013.04.037)

711 Peirce, JW. 2007 PsychoPy - Psychophysics software in Python. *J. Neurosci.*
712 *Methods* 162, 8-13. (doi: 10.1016/j.jneumeth.2006.11.017)

713 Peirce JW. 2009 Generating stimuli for neuroscience using PsychoPy. *Front.*
714 *Neuroinform.* 2:10. (doi:10.3389/neuro.11.010.2008)

715 Penny WD. 2012 Comparing dynamic causal models using AIC, BIC and free energy.
716 *NeuroImage,* 59, 319–330. (doi: 10.1016/j.neuroimage.2011.07.039)

bioRxiv preprint doi: https://doi.org/10.1101/061473; this version posted June 30, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

717 Racey D, Young ME, Garlick D, Pham JNM, Blaisdell AP. 2011 Pigeon and human
718 performance in a multi-armed bandit task in response to changes in variable
719 interval schedules. *Learn. Behav.* 39, 245–258. (doi: 10.3758/s13420-011-
720 0025-7)

721 Reader SM. 2015 Causes of Individual Differences in Animal Exploration and
722 Search, *Topics in Cognitive Science* 7, 451–468.

723 Réale D, Reader SM, Sol D, McDougall PT, Dingemanse, NJ. 2007 Integrating
724 animal temperament within ecology and evolution. *Biological Reviews*, 82,
725 291–318.

726 Rendell L, Boyd R, Cownden D, Enquist M, Eriksson K, Feldman MW, Forgaty L,
727 Ghirlanda S, Lillicrap T, Laland KN. 2010 Why copy others? Insights from the
728 social learning strategies tournament. *Science* 328, 208–213. (doi:
729 10.1126/science.1184719)

730 Rendell L, Fogarty L, Hoppitt WJE, Morgan TJH, Webster MM, Laland KN. 2011
731 Cognitive culture: theoretical and empirical insights into social learning
732 strategies. *Trends. Cog. Sci.* 15, 68-76. (doi:10.1016/j.tics.2010.12.002)

733 Rigoux L, Stephan KE, Friston KJ, Daunizeau J. 2014 Bayesian model selection for
734 group studies — Revisited. *NeuroImage* 84, 971-985. (doi:
735 10.1016/j.neuroimage.2013.08.065)

736 Rosa P, Nguyen V, Dubois F. 2012 Individual differences in sampling behaviour
737 predict social information use in zebra finches. *Behav. Ecol. Socio. Biol*. 66,
738 1259-1265. (doi: 10.1007/s00265-012-1379-3)

bioRxiv preprint doi: https://doi.org/10.1101/061473; this version posted June 30, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

739 Stephan KE, Penny WD, Daunizeau J, Moran RJ, Friston KJ. 2009 Bayesian model
740 selection for group studies. *NeuroImage* 46, 1004–1017. (doi:
741 10.1016/j.neuroimage.2013.08.065)

742 Sutton RS, Barto AG. 1998 *Reinforcement Learning: An Introduction*. Cambridge,
743 MA: MIT press.

744 Toyokawa W, Kim H, Kameda T. 2014 Human Collective Intelligence under Dual
745 Exploration-Exploitation Dilemmas. *PLoS ONE* 9, e95789. (doi:10.
746 1371/journal.pone.0095789)

747 Trimmer PC, McNamara JM, Houston AI, Marshall JAR. 2012 Does natural selection
748 favour the Rescorla-Wagner rule? *Journal of Theoretical Biology*, 302,39-52.
749 (http://dx.doi.org/10.1016/j.jtbi.2012.02.014)

750 van Bergen Y, Coolen I, Laland KN. 2004 Nine-spined sticklebacks exploit the most
751 reliable source when public and private information conflict. *Proc. Roy. Soc. B*.
752 271, 957-962. (doi: 10.1098/rspb.2004.2684)

753 Webster M, Laland K. 2015 Space-use and sociability are not related to public-
754 information use in ninespine sticklebacks. *Behav. Ecol. Scio. Biol.* 69, 895-
755 907. (doi: 10.1007/s00265-015-1901-5)

756 Yoshida W, Ishii S. 2006 Resolution of uncertainty in prefrontal cortex. *Neuron* 50,
757 781–789.