RESEARCH ARTICLE

Undesirable Choice Biases with Small Differences in the Spatial Structure of Chance Stimulus Sequences

David Herrera, Mario Treviño*

Instituto de Neurociencias, Universidad de Guadalajara, Guadalajara, Jalisco, México

* mariomtv@hotmail.com

Abstract

In two-alternative discrimination tasks, experimenters usually randomize the location of the rewarded stimulus so that systematic behavior with respect to irrelevant stimuli can only produce chance performance on the learning curves. One way to achieve this is to use random numbers derived from a discrete binomial distribution to create a 'full random training schedule' (FRS). When using FRS, however, sporadic but long laterally-biased training sequences occur by chance and such 'input biases' are thought to promote the generation of laterally-biased choices (*i.e.*, 'output biases'). As an alternative, a 'Gellerman-like training schedule' (GLS) can be used. It removes most input biases by prohibiting the reward from appearing on the same location for more than three consecutive trials. The sequence of past rewards obtained from choosing a particular discriminative stimulus influences the probability of choosing that same stimulus on subsequent trials. Assuming that the long-term average ratio of choices matches the long-term average ratio of reinforcers, we hypothesized that a reduced amount of input biases in GLS compared to FRS should lead to a reduced production of output biases. We compared the choice patterns produced by a 'Rational Decision Maker' (RDM) in response to computer-generated FRS and GLS training sequences. To create a virtual RDM, we implemented an algorithm that generated choices based on past rewards. Our simulations revealed that, although the GLS presented fewer input biases than the FRS, the virtual RDM produced more output biases with GLS than with FRS under a variety of test conditions. Our results reveal that the statistical and temporal properties of training sequences interacted with the RDM to influence the production of output biases. Thus, discrete changes in the training paradigms did not translate linearly into modifications in the pattern of choices generated by a RDM. Virtual RDMs could be further employed to guide the selection of proper training schedules for perceptual decision-making studies.

Introduction

In typical two-alternative discrimination tasks, subjects are required to make choices between two options offered [1–4]. They receive rewards after correct responses on those trials in which they respond to the side where the 'positive' discriminative stimulus (S^D) appears, whereas they are credited with errors when they choose the incorrect stimulus on the other side (*i.e.* the stimulus delta, S^Δ). A major problem for the data analysis from these kind of experiments involves deciding whether fluctuations in the discrimination learning curve away from chance represent genuine changes in the subject's behavior with respect to the S^D, or whether they derive from some behavior that responds some other features of the situation (*v.gr.* the reward history) [4–8]. Thus, aiming to improve their measurements, experimenters commonly randomize the locations occupied by the rewarded stimuli, so that systematic behavior with respect to irrelevant stimuli/factors can only produce chance performance on the learning curve [9]. By definition, these presentation series contain an equal number of opposing sequences, balancing the frequency of reward for both options.

A practical way to create a 'full random training schedule' (FRS) for discrimination tasks is to use random numbers derived from a discrete binomial distribution. When using FRS, however, long but sporadic training sequences, in which reward is placed in the same location, occur by chance. We will refer to such repeating chunks of laterally-biased training sequences as 'input bias'. To justify such an unconstrained training sequence generation, one must either assume that the subjects are not influenced by input biases or that the effects of input biases are washed out over time. Yet, empirical evidence indicates that during the early training phase subjects do not immediately link their responses to the S^D and they also tend to establish behavioral strategies based on recent reward history [2–4,10]. For these reasons, choosing the right order and combination of discriminative stimuli during training is extremely important, particularly when the stimulus trials are larger than the number of display locations (two for dichotomic tasks). For example, the number of trials in which the same comparison stimulus is scheduled as correct can induce temporary stimulus preferences that interfere with the reinforcement contingencies intended by the experimenter, whereas disruptive location preferences can result from repeated presentations of correct stimuli at the same location [10–13].

One effort to restrain the source of such input biases for training sequences is the 'Gellerman-like training schedule' (GLS; [2–4,14]), a template designed to render the S^D on each trial unpredictable from the outcome of previous trials, such that its location is not correct for more than three consecutive times. In this way, obvious artifacts such as alternating sequences are not introduced.

Here, we compared the choice patterns produced by a virtual 'Rational Decision Maker' (RDM) in response to computer-generated FRS and GLS training sequences. Through computer simulations, we found that, although the GLS presented fewer input biases than the FRS, the virtual RDM produced more steady-state output biases with GLS than with FRS under a variety of test conditions. This confirms the fact that the statistical and temporal properties of training sequences interacted with the RDM to influence the production of output biases.

Results

Although it is clear that FRS contains an implicit source of input biases, we do not understand exactly how switching from FRS to GLS affects the production of output biases produced by a simple 'Rational Decision Maker' (RDM; see equations below). This is a very relevant question because choice biases are deleterious for experiments in psychophysics: they affect the estimation of perceptual thresholds [3,4,6], are associated with higher error rates, less reinforcement

and also usually reflect a general disengagement from animals performing the behavioral task [1,4,6,15].

Reinforcement learning theories provide a powerful theoretical framework for the description of choice behavior in dynamic environments: they state that future actions are chosen to maximize a long-term sum of positive outcomes, which can be accomplished through a set of value functions that represent the amount of expected reward associated with particular states or actions [13]. To match behavior to income, animals must integrate the rewards earned earlier from particular behaviors, and maintain an appropriate representation of the value of competing alternatives (*i.e.* reward frequency). Quantitatively, this idea is captured by the matching law [16,17], which states that the long-term average ratio of choices matches the long-term average ratio of reinforcers [16,18,19]:

$$\frac{C_R}{C_L} = c \left(\frac{R_R}{R_L} \right)^a \qquad (1)$$

where C_R and C_L denote the number of steady-state responses, and R_R and R_L the number of steady-state reinforcers for the right (R) and left (L) options, respectively. The coefficient a denotes the sensitivity to the reinforcement ratio, and c is a bias term unrelated to reinforcer frequency or magnitude [18,19]. This equation describes an empirical regularity of behavior in which the average tendency of subjects to make their choices is in correspondence with the frequency of reward they receive from choosing it. It predicts that the amount of laterally biased choices (output biases) should be proportional to the amount of laterally biased rewards (input biases).

The reward history from a S^D influences the probability of choosing it again on subsequent trials [4,10–13,20]. Because location habits are a common feature of animal behavior [4,21], it is crucial to carefully choose the training sequences of locations for the S^D. Such sequences should prevent better, or worse, than chance performance arising from systematic behavior based upon location, and should also prevent the accidental reinforcement of any such behavior.

Typically, FRS have been used to reduce the predictability of reward, aiming to make discrimination tasks independent of spatial cues. However, FRS contain a source of input biases which may influence the production of output biases. Thus, it would be desirable to test GLS instead of FRS because the former has fewer input biases.

By assuming that choices comply with matching behavior, we hypothesized that the reduced number of input biases in GLS would decrease the production of output biases compared to FRS. Such a reduction could increase the task efficiency because biased choices are linked to errors (in balanced tasks). If true, this would suggest a positive interaction between the GLS and the decision-making process. If not, it could either reflect a non-linearity in the decision-making process or that biased choices do not comply with matching behavior.

Our specific aim was to compare the choice patterns produced by a RDM in response to FRS and GLS training schedules (Fig 1A). First, we made linear arrays of pseudo-random numbers derived from a discrete binomial distribution to create FRS and GLS training schedules (100 trials x 100 repetitions). The GLS algorithm was identical to the FRS one, but it additionally prohibited sequence repetitions of more than three consecutive trials [2–4,14]. Reinforcement was provided on each trial for one of the two mutually exclusive options (*i.e.* probability of reward per trial = 100%; probability of reward per side = 50%). To create a virtual RDM, we implemented an algorithm that made choices based on past rewards and choices according to an exponentially-weighted moving average filter (EWMA), but was insensitive to differences in discriminative input signals [1]. With this model, we assumed that the integration of past
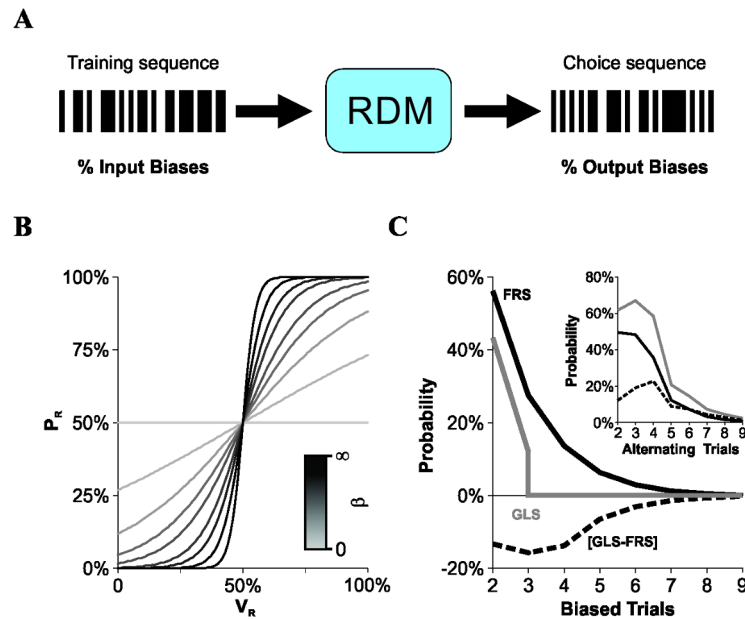
Fig 1. Virtual decision maker and training schedules with different input bias probabilities. (A) A training sequence with a % input biases is processed by a virtual 'Rational Decision Maker' (RDM) to produce choice sequences with a % output biases. (B) The RDM algorithm varied according to a sigmoidal function, in which the probability of choosing to the right (P_R, y-axis) depended on the weight function for the right option (V_R, x-axis) and on the slope of the curve (β; darker lines represent higher β values; Eq 2). (C) The probability of finding laterally biased sequences (y-axis) of different lengths (x-axis) is smaller using GLS (gray line) than FRS (black line). The opposite occurred for alternating sequences (inset). The differences between group probabilities are displayed by dotted lines and are always negative. We calculated the probabilities for biased and alternating sequences by creating independent sets of randomized sequences. Average probabilities for biased/alternating 2-trial-sequences with the FRS schedule converge towards 50% when the number of trials is increased (*v.gr.* from 100 to 1000; not illustrated).

rewards is imperfect (*i.e.* leaky), which translates into a finite effective memory on estimates of income, making them local rather than global (in time) [13]. Indeed, the EWMA provides a description for short-term memory in which a reinforcer produces smaller effects into current choices as one considers responses that extend further into the past [12,22]. We chose to use the EWMA based on a series of quantitative observations made on the choice records from diverse animal models [4,11,13,19,23], and because it has fewer free parameters than other alternative models *v.gr.* [24]. 'Memory gradients' have been also described with hyperbolic functions [19,25], but this won't be addressed here.

The probability that the RDM chose a particular option was computed by comparing a uniformly distributed pseudo-random number against a logistic curve with the following form:

$$P_R = \frac{1}{1 + e^{\beta(V_R - V_L + B)}} \qquad (2)$$

where β corresponds to the slope of the sigmoid (Fig 1B, darker lines represent higher β values; see also [26,27]), and V_R and V_L are weight functions for the right and left options, respectively, and B is a positive or negative constant (with randomized sign) that can be used to represent individual biases towards one or other option (*i.e.* a lateral bias, hereafter referred to as 'intrinsic laterality' [28]). In this system, V_R and V_L served to horizontally shift P_R towards one or the other option, influencing future choices. Also, the weight functions for the chosen (*x*) and unchosen (*y*) options were updated on a trial-by-trial basis using the Euler approximation of

derivatives:

$$v_{x,i} = \begin{cases} 1 & \text{if rewarded} \quad (a) \\ k_1 \cdot v_{x,i-1} & \text{if not rewarded} \quad (b) \end{cases} \tag{3}$$

$$v_-(y,i) = k_-2 \quad [\![\cdot v]\!] \; _-(y, i-1) \qquad \text{if un} - \text{chosen} \tag{4}$$

According to these equations, every time a choice was rewarded, its corresponding weight value was increased to one (Eq 3(A)), whereas the weight values for the un-rewarded (Eq 3(B)) and the un-chosen (Eq 4) options decreased $(1-k_1)$ and $(1-k_2)$, respectively (on every trial). These k values are directly linked to a mono-exponentially decaying process with a $\tau = -1/\ln(k)$. An initial fast reduction of the weight value of the un-rewarded option (small k_1), followed by a slower reduction in the weight value for the un-chosen option ($k_2 > k_1$) are consistent with the law of effect [28,29] and could be a plausible representation of short-term memory [11].

To quantify the amount of biased and alternating sequences in the training and choice records, we implemented a pair-wise alignment method that consisted in sliding each query sequence along the record of interest. A sequence was considered to be present in the record when the alignment matched perfectly for the entire length of the query sequence (*i.e.* sequence similarity = 1). The average probability of occurrence of a sequence was then calculated by dividing the number of times that it was found in the record by the maximum number of times that it could fit within the total length of the record without interfering with any identical sequence that generated a count in previous trials. Note that comparing the average probabilities instead of the number of cases is fundamental because the alignment method implicitly increases the counts for shorter sequences contained in longer ones [4]. With this tool, we next analyzed the input bias probability for complementary sequences using training trials produced by both sequence-generation algorithms and found that, as expected, it was smaller for GLS (gray line) than for FRS (black line; Fig 1C). Notably, biased sequences with a length of two and three trials were also more likely to occur in FRS than in GLS, making the overall difference in input bias probabilities always negative ($\Delta P = P_{input\_bias[GLS]} - P_{input\_bias[FRS]}$; dotted line; Fig 1C). Furthermore, because the probability of finding biased sequences of length > 3 in GLS is always zero, this implies that the difference between these distributions must always be negative. Such a difference could mean that other un-biased sequences must have changed their probability of occurrence. We counted the number of complementary alternating sequences from the same training schedules and found that they were indeed more frequent in GLS (gray line) than in FRS (black line; Fig 1C, inset). Thus, switching from FRS to GLS fully re-arranged the probabilities of observing biased and alternating sequences during training. However, the reward probability in steady state was well balanced between the two options for both groups (FRS: 50.66% ± 0.50% $n = 10,000$; GLS: 50.00% ± 0.50% $n = 10,000$). Note that we calculated the probabilities for biased and alternating sequences by creating independent sets of randomized sequences (100 repetitions).

We next asked how GLS and FRS influenced the production of output biases when processed by the same virtual RDM. To avoid misestimating output probabilities, we tested the RDM with 100 trial sequences that were re-randomized according to GLS and FRS for every condition. We mapped for a variety of internal mechanisms of the RDM: multiple β values, intrinsic laterality and different k values. For this analysis, we measured the probability of occurrence of output biases of variable length dividing the number of times that each biased sequence was found in the choice record by the maximum number of times that it could fit in the choice record if the choices had been completely biased. To facilitate the visualization of
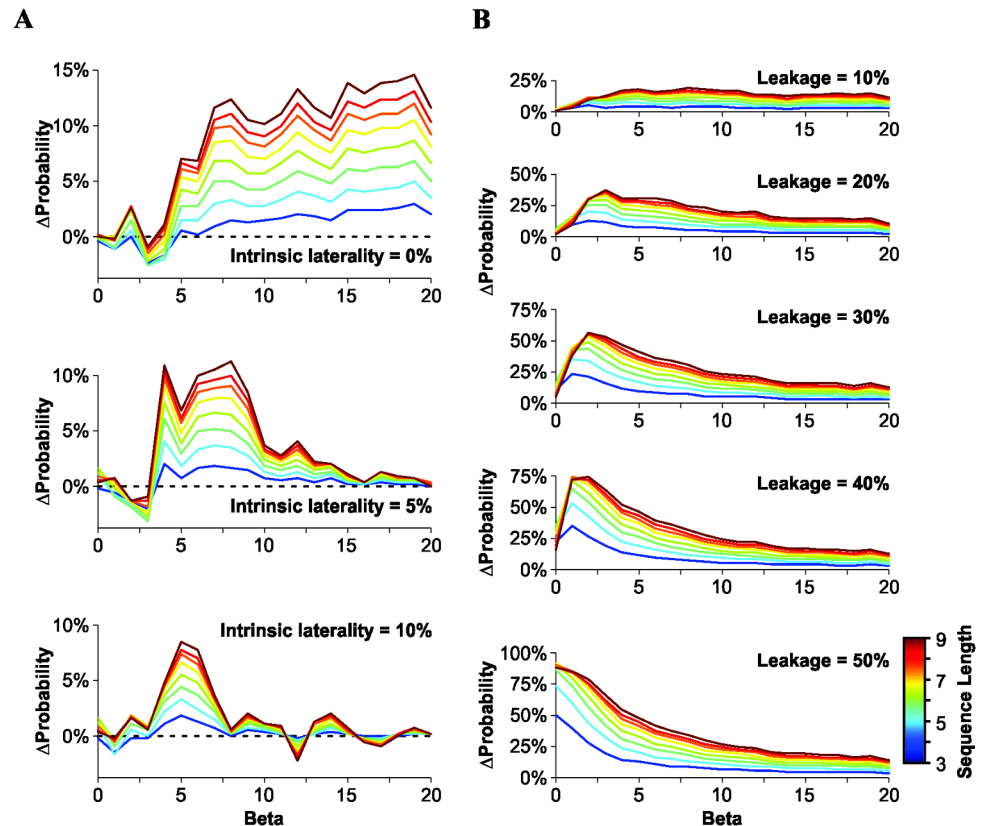
Fig 2. Training the RDM with FRS and GLS. Results are displayed as the difference in the probability of finding output biased choices produced with a 'Gellerman-like' (GLS) minus a 'Full Random' (FRS) training schedule, after being processed by the 'Rational Decision Maker' (RDM) algorithm (*i.e.* $\Delta P = P_{output\_bias[GLS]} - P_{output\_bias[FRS]}$). ΔP values were computed as a function of β (x-axis; Eq 2) and of sequence length (from 3 to 9; darker lines represent longer sequences). ΔP values were always positive (for $\beta \geq 3$), yet they decreased with intrinsic laterality (A; $k_1 = 0.5$, $k_2 = 0.5$, leakage = 0%), but increased with leakage (B; $k_1 = 0.5$, $k_2 = 0.5$, intrinsic laterality = 0%).

doi:10.1371/journal.pone.0136084.g002

results, we subtracted the output bias probability obtained when using GLS from that one produced when using FRS ($\Delta P = P_{output\_bias[GLS]} - P_{output\_bias[FRS]}$). Surprisingly, the output bias probability was higher with GLS than with FRS (for $\beta \geq 3$), and group differences increased with the length of the biased sequence (darker lines for output biased sequences of longer length; Fig 2A upper panel). We approximated the slopes of the fitted logistic curves (*i.e.* β) taken from normalized data from published psychometric responses from the auditory system in rats and humans ($\beta \approx 12.6$, $\beta \approx 13.9$, respectively; [29]), and from the visual ($\beta \approx 8.9$; [30]) and somatosensory ($\beta \approx 8.4$; [31]) systems in monkeys. All these β values were bigger than 3. These numbers indicate that our results might have a relevant impact for the estimation of psychometric curves.

To investigate whether the sign of ΔP depended on the intrinsic components of the RDM, we introduced a source of individual laterality with fixed magnitude but variable polarity (Eq 2). This number represents an implicit preference of the RDM for a particular option, at the individual but not at the population level (see also [4]). Notably, although ΔP decreased with higher intrinsic laterality values, the production of output biases was still larger when using GLS compared to FRS (Fig 2A; middle panel, intrinsic laterality: 5% and lower panel, intrinsic laterality: 10%).
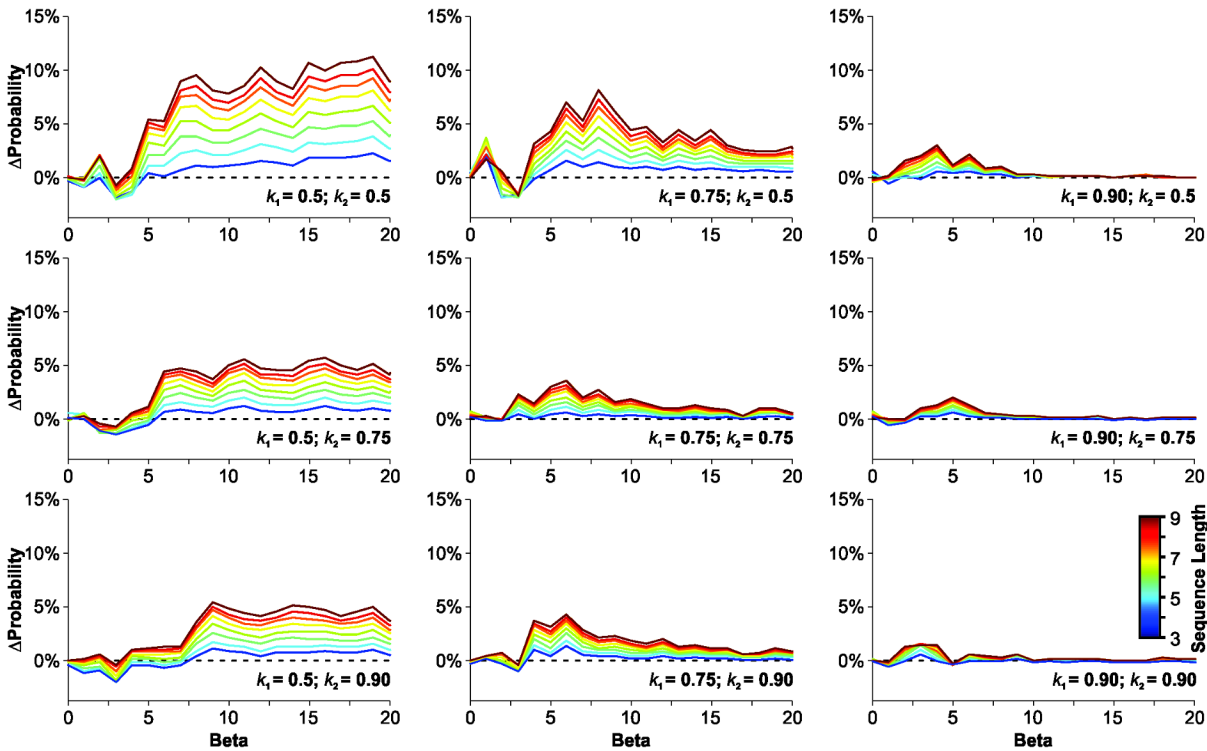
Fig 3. Testing the RDM with different 'short-term memory' decaying rates. Results are displayed as the difference in the probability of finding output biased choices produced with a 'Gellerman-like' (GLS) minus a 'Full Random' (FRS) training schedule, after being processed by the 'Rational Decision Maker' (RDM) algorithm (*i.e.* $\Delta P = P_{output\_bias[GLS]} - P_{output\_bias[FRS]}$). Panels are arranged in columns ($k_1 = 0.5$, 0.75 and 0.90) and rows ($k_2 = 0.5$, 0.75 and 0.90). Increasing k values implies longer short-term memory representatioτns. k values of 0.5, 0.75 and 0.90 correspond to a τ of 1.44, 3.47 and 9.49, respectively.

doi:10.1371/journal.pone.0136084.g003

Next, we asked whether an imperfection in the decision-making process could reverse the differences observed in the production of output biases. We introduced a graded leakage in the effective probability of choosing a particular option ($EP_R = [\varepsilon \cdot P_R]$; [32]) but found that ΔP remained positive for all values tested, and even displayed an increased output bias production with bigger information leakage (Fig 2B).

According to the matching law, each reward contributes equally to increasing the probability of choosing an option [16]. This contribution is conceptualized as being stable over time because these mathematical descriptions are valid only for choice behavior in a steady-state scenario [12]. In a dynamic system, however, the reward changes its location over time and choice behavior is strongly influenced by recently obtained reinforcers [12]. In such scenario, k_1 and k_2 represent constants that determine the decay rate for the contribution of past rewards to current choice (Eq 3(B) and Eq 4; see also [13]). Higher k values imply a higher impact of reward history on current choice, whereas lower values represent a more prolonged effect of reinforcement history. Based on this idea, we investigated whether ΔP could invert its sign after changing the relative contribution of past rewards. We tested all possible combinations for three different discrete values for k_1 and k_2, respectively. In Fig 3 we show that ΔP values were positive for all combinations tested (covering both $k_1 > k_2$ and $k_1 < k_2$), yet group differences became smaller as k values increased. Also, k_1 had a stronger effect (than k_2) in dissipating ΔP for high β values (Fig 3).

Because output biases were generally higher with GLS than with FRS, we pondered whether these group differences also led to differential rewards obtained by the RDM. Reward was
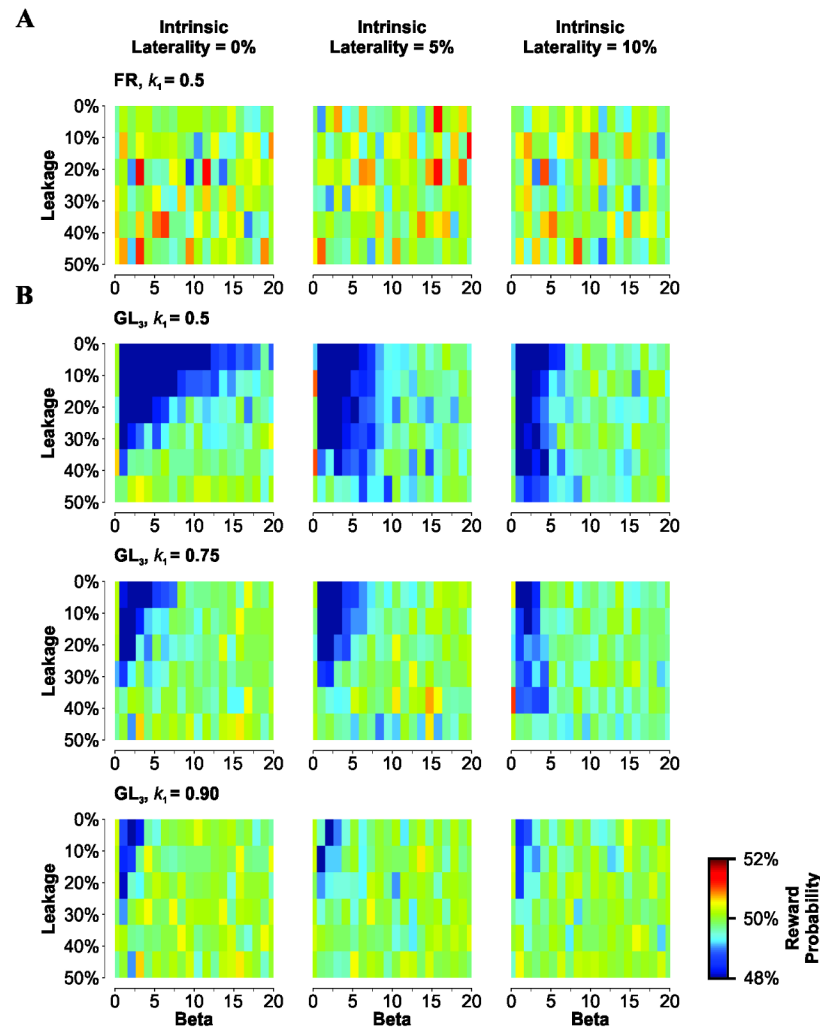
Fig 4. Reduced reward probability when training the RDM with GLS. Reward probability was estimated by computing the group average correlation between reward location and the side to which the choices were made. (A) Irrespective of leakage (y-axis) and β (x-axis) values, the reward obtained with a 'Full Random' training schedule (FRS) processed by the 'Rational Decision Maker' (RDM) with increasing intrinsic laterality values (displayed in columns) was always around chance level (50%). (B) There is an apparent drop in reward (dark regions) obtained with a 'Gellerman-like' training schedule (GLS) when processed by the RDM for small leakage and β values. However, the drop tends to disappear upon increasing information leakage (y-axis), intrinsic laterality (columns) and increasing k_1 (rows; $k_2 = 0.5$ for all panels).

calculated by using the correlation between reward location and the corresponding choice record from the RDM. As expected for a balanced reward distribution, the reward probability was near chance level when the RDM was trained with the FRS (Fig 4A). When using GLS, however, the reward probability dropped systematically from this level by approximately 2–3% for $\beta \leq 15$ and with an information leakage ≤ 30 (dark regions in panels from Fig 4B). This apparent reduction in reward probability vanished upon increasing β (x-axis), the information leakage (y-axis), the intrinsic laterality (columns in Fig 4), and also upon increasing k_1 (rows in Fig 4).

Finally, we probed how the drop in reward probability obtained with GLS training related to changes in the output bias probability. In this case, we selected blocks of 30 training trials and counted the number of times that each sequence occurred within the choice record for
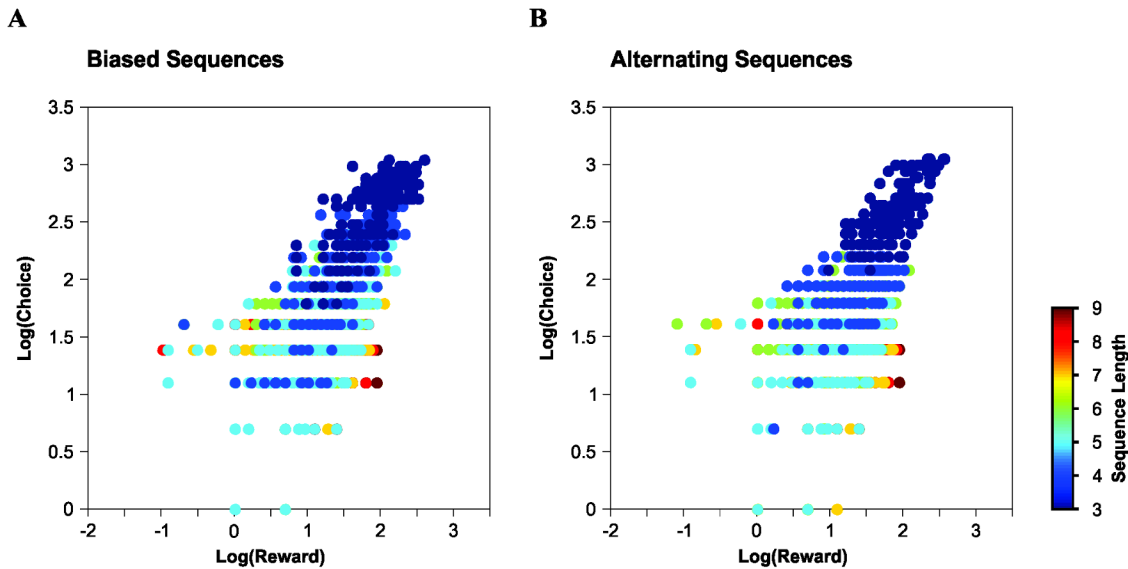
Fig 5. Output choice biases comply with the generalized matching law. The log frequency of choices is linearly related to the log frequency of rewards for biased sequences (A) and alternating sequences (B) obtained with both a 'Full Random' (FRS) and a 'Gellerman-like' (GLS) training schedules processed by the 'Rational Decision Maker' (RDM). Each point represents the frequency of occurrence of a particular choice sequence *versus* the sum of reward values obtained from choosing those specific sequences.

each training repetition of the RDM (*i.e.* sequence frequency), and calculated the corresponding reward value as the sum of the correct choices for those found sequences (*i.e.* reward frequency). Fig 5 shows the log frequency (y-axis) for biased (Fig 5A) and alternating (Fig 5B) choice sequences as a function of the log net amount of reinforcer received (x-axis). The slopes for both models were positive indicating that all biased choices produced by the RDM complied with the matching law.

Discussion

In two-alternative discrimination tasks, researchers usually present the positive stimulus (S^D) on either the right or left side of the apparatus according to a fully random training schedule (FRS). It is assumed that such orders of training stimuli provide the subjects an opportunity of making only half of their responses correct through 'chance' alone. In actual experience, however, it has been found that additional factors such as habits of alternation may result in an accuracy higher than 50%. The uncertainty in the interpretation of such results is in part due to the selection of the orders of training stimuli. Gellerman (1933) argued that the FRS allows various incorrect habits, such as location alternation, to produce an accuracy of performance as high as 70%. He attributed the general uncertainty on where to put the criterion for learning to the faulty selection of orders of alternating stimuli [14]. To remedy this situation, he proposed some criteria necessary to ensure balanced random sequences, as in the 'Gellerman-like training schedule' (GLS) that we used here. Indeed, the GLS removes input biases by prohibiting the S^D from appearing at the same location for more than three consecutive trials [2–4,21].

Although it was clear to us that the FRS contained an implicit source of input biases, we could not predict how switching from FRS to GLS training scheme would affect the production of output biases. We thus created a virtual 'Rational Decision Maker' (RDM) that was sensitive to past rewards, trained it with FRS and GLS sequences, and measured its response properties under a variety of test conditions by changing β, intrinsic laterality, information leakage and k values. Unexpectedly, although GLS contained fewer input biases than FRS, the RDM

produced more output biases when trained with GLS rather than with FRS (for $\beta \geq 3$). Interestingly, we introduced an internal source of intrinsic laterality [4] and found that it contributed in dissipating the differences in output biases between groups. One possibility is that such differences were still existent but became much smaller, because intrinsic laterality increased the portion of biased choices that are independent of the training schedule. Moreover, since alternating choices increase with smaller β values, this should reduce the relative impact of the intrinsic laterality in the choices made by the RDM. Also, regardless of the relative differences in the production of output biases between FRS and GLS (*i.e.* a between-group comparison), we confirmed that all biased sequences produced by the RDM complied with the matching law (*i.e.* a within-subject relationship).

The output choices of the RDM depended on a dynamic competition between weight values for both options (V_R and V_L) that were raised when rewarded but decayed for un-rewarded and un-chosen options. When the weight values were similar (*i.e.* $V_R/V_L \approx 1$), the decision process tended to be at random, favoring alternation between options. In contrast, when the weight value for one option grew relative to the opposite one (*v.gr.* $V_R >> V_L$; $V_R/V_L \rightarrow \infty$ and $V_R/V_L \rightarrow 0$), the likelihood of choosing the former option increased ($P_R >> P_L$), creating a tendency in the RDM to produce more output biases. Accordingly, to promote alternation in the RDM, it was necessary to decrease the weight value of the favored option (by not rewarding it) while increasing the weight value of the opposite option (by rewarding it) until a $V_R/V_L \approx 1$ balance was reached.

Weight values for un-rewarded and un-chosen options decayed mono-exponentially at different rates, which we mapped for different values. However, irrespective of which of these functions decayed faster, the RDM always produced more output biases with GLS than with FRS. Indeed, for both weight value representations, smaller k values led to a faster decay rate that promoted alternation, whereas a bigger values lead to a longer integration time window, making the choices of the RDM (*i.e.* output) less sensitive to the local (temporal) variations in training conditions (*i.e.* input). Thus, the decay time constants of the un-rewarded and un-chosen options strongly determined the period during which the RDM was sensitive to the input stream of training sequences. Clearly, the fastest way to remove an output bias in the RDM would have been by using an input bias towards the opposite option, as has been suggested previously [2–4]. The lack of input biases in GLS implied that long sequences of biased choices could not be rewarded (and un-rewarded; $P_R = 1-P_L$), thus reducing the efficiency of GLS, compared to FRS, to counterbalance the emergence of output biases. In contrast, the naturally occurring input biases in FRS provided the right opportunity to produce sequences of un-rewarded choices, which contributed to adjust the weight values, reduce output biases and promote alternation. Thus, the GLS training sequence contains information about which stimulus will be the rewarded one, at least in cases where three repetitions have occurred already. Because the RDM is sensitive to past rewards, it might seem then less surprising that the GLS led to more output biases than the FRS.

One important limitation of our study is that all the results presented here are strictly dependent on: i) the equations that we used to describe the RDM and ii) the fact that these equations successfully describe, by using very few free parameters, some basic principles of animal choice behavior [12,13,19,24]. However, depending on how the RDM diverges from real animal choice behavior, the results could lose their potential relevance. It is thus crucial to test these predictions with animal models. We have started doing some attempts in this direction [4]. Using GLS for training, we have shown that, depending on the degree of discriminative information, mice display different amounts of biased choices [1]. In conditions of high discriminability, the mice learned to discriminate and produce a reduced number of side-biased choices. However, as discriminability decreased, they increased the usage of biased

choices to solve the task [4]. Thus, when the discriminative information was low, the reward history became the most important cue for guiding their behavior [4,10,11]. An interesting possibility is that such biased choices could constitute a valid strategy for solving tasks when discriminative cues are scarce. These observations suggest that output biases could indeed be modulated by quantitative differences in the discriminability of sensory stimuli. How exactly the discriminative information regulates the strength of the dependency on past rewards (and choices [12]) to produce biased choices remains an open question.

The GLS has been a standard for decades in research on discrimination learning [14]. Using numerical simulations, we here showed that a simple virtual RDM unexpectedly produced more (and not fewer) biased choices when trained with GLS instead of FRS. Thus, it might be desirable to train with GLS but switch to FRS once the subjects reach a steady-state choice behavior. Certainly, as the complexities of the training procedures increase, the challenges in making appropriate combinatorial decisions concerning stimulus order and location sequences become progressively greater [9,33].

Our results demonstrate that the statistical and temporal properties of training streams interact with the internal components of the RDM to influence the production of output biases. They highlight the fact that discrete changes in the training paradigms do not translate linearly into changes in the output choices made by this simple choice maker. Therefore, manipulations of the properties of input streams must be made with extreme precaution because they may have a relevant impact on the estimation of psychophysical measures that are assessed through choice behavior. We propose the usage of virtual RDMs to select appropriate training streams for perceptual decision-making studies.

Acknowledgments

Author Contributions

Conceived and designed the experiments: MT. Performed the experiments: MT. Analyzed the data: MT DH. Contributed reagents/materials/analysis tools: MT. Wrote the paper: MT DH.

References

1. Trevino M, Oviedo T, Jendritza P, Li SB, Kohr G, De Marco RJ (2013) Controlled variations in stimulus similarity during learning determine visual discrimination capacity in freely moving mice. Sci Rep 3: 1048. doi: 10.1038/srep01048 PMID: 23308341

2. Prusky GT, West PW, Douglas RM (2000) Behavioral assessment of visual acuity in mice and rats. Vision Res 40: 2201–2209. S0042-6989(00)00081-X [pii]. PMID: 10878281

3. Trevino M, Frey S, Kohr G (2012) Alpha-1 adrenergic receptors gate rapid orientation-specific reduction in visual discrimination. Cereb Cortex 22: 2529–2541. bhr333 [pii]; doi: 10.1093/cercor/bhr333 PMID: 22120418

4. Trevino M (2014) Stimulus similarity determines the prevalence of behavioral laterality in a visual discrimination task for mice. Sci Rep 4. Article.

5. Grunow A, Neuringer A (2002) Learning to vary and varying to learn. Psychon Bull Rev 9: 250–258. PMID: 12120786

6. Busse L, Ayaz A, Dhruv NT, Katzner S, Saleem AB, Scholvinck ML, et al. (2011) The detection of visual contrast in the behaving mouse. J Neurosci 31: 11351–11361. 31/31/11351 [pii]; doi: 10.1523/JNEUROSCI.6689-10.2011 PMID: 21813694

7. Schwartz B (1980) Development of complex, stereotyped behavior in pigeons. J Exp Anal Behav 33: 153–166. PMID: 16812164

8. Bhagavatula PS, Claudianos C, Ibbotson MR, Srinivasan MV (2014) Behavioral lateralization and optimal route choice in flying budgerigars. PLoS Comput Biol 10: e1003473. doi: 10.1371/journal.pcbi.1003473 PCOMPBIOL-D-13-01914 [pii]. PMID: 24603285

9. Fellows BJ (1967) Chance stimulus sequences for discrimination tasks. Psychol Bull 67: 87–92. PMID: 6045339

10. Catania AC (1971) Reinforcement schedules: the role of responses preceding the one that produces the reinforcer. J Exp Anal Behav 15: 271–287. PMID: 16811513

11. Killeen PR (1994) Mathematical principles of reinforcement. In: Behavioral and Brain Sciences. pp. 105–172.

12. Lau B, Glimcher PW (2005) Dynamic response-by-response models of matching behavior in rhesus monkeys. J Exp Anal Behav 84: 555–579. PMID: 16596980

13. Sugrue LP, Corrado GS, Newsome WT (2004) Matching behavior and the representation of value in the parietal cortex. Science 304: 1782–1787. doi: 10.1126/science.1094765 304/5678/1782 [pii]. PMID: 15205529

14. Gellermann LW (1933) Chance Orders of Alternating Stimuli in Visual Discrimination Experiments. The Pedagogical Seminary and Journal of Genetic Psychology 42: 206–208. doi: 10.1080/08856559.1933.10534237

15. Huh N, Jo S, Kim H, Sul JH, Jung MW (2009) Model-based reinforcement learning under concurrent schedules of reinforcement in rodents. Learn Mem 16: 315–323. 16/5/315 [pii]; doi: 10.1101/lm.1295509 PMID: 19403794

16. Herrnstein RJ (1961) Relative and Absolute strength of Response as a Function of Frequency of Reinforcement1,2. Journal of the Experimental Analysis of Behavior 4: 267–272. PMID: 13713775

17. Baum WM (1974) On two types of deviation from the matching law: bias and undermatching. J Exp Anal Behav 22: 231–242. PMID: 16811782

18. Stuttgen MC, Yildiz A, Gunturkun O (2011) Adaptive criterion setting in perceptual decision making. J Exp Anal Behav 96: 155–176. doi: 10.1901/jeab.2011.96–155 2011–0006 [pii]. PMID: 21909162

19. Corrado GS, Sugrue LP, Seung HS, Newsome WT (2005) Linear-Nonlinear-Poisson models of primate choice dynamics. J Exp Anal Behav 84: 581–617. PMID: 16596981

20. Trevino M, Aguilar-Garnica E, Jendritza P, Li SB, Oviedo T, Kohr G, et al. (2011) Discrimination learning with variable stimulus 'salience'. Int Arch Med 4: 26. 1755-7682-4-26 [pii]; doi: 10.1186/1755-7682-4-26 PMID: 21812982

21. Sutherland NS (1964) VISUAL DISCRIMINATION IN ANIMALS. Br Med Bull 20: 54–59. PMID: 14104098

22. Davison M, Baum WM (2000) Choice in a variable environment: every reinforcer counts. J Exp Anal Behav 74: 1–24. doi: 10.1901/jeab.2000.74–1 PMID: 10966094

23. Brunton BW, Botvinick MM, Brody CD (2013) Rats and humans can optimally accumulate evidence for decision-making. Science 340: 95–98. 340/6128/95 [pii]; doi: 10.1126/science.1233912 PMID: 23559254

24. Sanabria F, Thrailkill E (2009) Pigeons (Columba livia) approach Nash equilibrium in experimental Matching Pennies competitions. J Exp Anal Behav 91: 169–183. doi: 10.1901/jeab.2009.91–169 PMID: 19794832

25. Rachlin H, Raineri A, Cross D (1991) Subjective probability and delay. J Exp Anal Behav 55: 233–244. doi: 10.1901/jeab.1991.55–233 PMID: 2037827

26. Carvalho TP, Buonomano DV (2009) Differential effects of excitatory and inhibitory plasticity on synaptically driven neuronal input-output functions. Neuron 61: 774–785. S0896-6273(09)00080-4 [pii]; doi: 10.1016/j.neuron.2009.01.013 PMID: 19285473

27. Wolfart J, Debay D, Le MG, Destexhe A, Bal T (2005) Synaptic background activity controls spike transfer from thalamus to cortex. Nat Neurosci 8: 1760–1767. nn1591 [pii]; doi: 10.1038/nn1591 PMID: 16261132

28. Priddy KL, Keller PE (2005) Artificial Neural Networks: An Introduction (SPIE Tutorial Texts in Optical Engineering, Vol. TT68). Washington: SPIE-The International Society for Optical Engineering.

29. Brunton BW, Botvinick MM, Brody CD (2013) Rats and humans can optimally accumulate evidence for decision-making. Science 340: 95–98. 340/6128/95 [pii]; doi: 10.1126/science.1233912 PMID: 23559254

30. Shadlen MN, Kiani R (2013) Decision making as a window on cognition. Neuron 80: 791–806. S0896-6273(13)00999-9 [pii]; doi: 10.1016/j.neuron.2013.10.047 PMID: 24183028

31. Romo R, Hernandez A, Zainos A, Salinas E (2003) Correlated neuronal discharges that increase coding efficiency during perceptual discrimination. Neuron 38: 649–657. S0896627303002873 [pii]. PMID: 12765615

32. Usher M, McClelland JL (2001) The time course of perceptual choice: the leaky, competing accumulator model. Psychol Rev 108: 550–592. PMID: 11488378

33. Iversen I (1997) Matching-to-sample Performance In Rats: A Case Of Mistaken Identity? J Exp Anal Behav 68: 27–45. PMID: 16812849