Behavioural Brain Research 372 (2019) 111983

Contents lists available at ScienceDirect

Behavioural Brain Research

journal homepage: www.elsevier.com/locate/bbr

Short communication

Effects of reward size and context on learning in macaque monkeys

Lorenzo Ferrucci[1], Simon Nougaret[1], Emiliano Brunamonti, Aldo Genovesio*

Department of Physiology and Pharmacology, SAPIENZA, University of Rome, Piazzale Aldo Moro 5, 00185, Rome, Italy

ARTICLE INFO

Keywords:
Monkeys
Learning
Reward
Context

ABSTRACT

The outcome of an action plays a crucial role in decision-making and reinforcement learning processes. Indeed, both human and animal behavioural studies have shown that different expected reward values, either quantitatively or qualitatively, modulate the motivation of subjects to perform an action and, as a consequence, affect their behavioural performance. Here, we investigated the effect of different amounts of reward on the learning of macaque monkeys using a modified version of the object-in-place task. This task offers the opportunity to shape rapid learning based on a set of external stimuli that enhance an animal's accuracy in terms of solving a problem. We compared the learning of three monkeys among three different reward conditions. Our results demonstrate that the larger the reward, the better the monkey's ability to learn the associations starting with the second presentation of the problem. Moreover, we compared the present results with those of our previous work using the same monkeys in the same task but with a unique reward condition, the intermediate one. Interestingly, the performance of our animals in our previous work matched with their performance in the largest and not intermediate reward condition of the present study These results suggest that learning is mostly influenced by the reward context and not by its absolute value.

1. Main text

Reinforcement learning is defined as learning which actions to perform in order to maximize reward [1]. Neuroscience research with non-human primates typically uses this paradigm to train monkeys to perform behavioural tasks designed to study the neural mechanisms underlying specific cognitive processes. Animals learn to match the execution of specific actions and behaviours with a positive reinforcement and to avoid behaviours that lead to an absence of reward. Modulation in the schedule of reward delivery is a powerful tool that can accelerate the training [2] and offer insight into the neural basis of reward processing [3]. There are several pieces of evidence confirming that reward size is correlated with an increase of motivational level that enhances accuracy during a task [4,5]. Here, we investigated whether and how the size of reward influenced the learning speed in macaque monkeys. We also compared these results with those of our previous study [6] using the same animals and the same task but with a unique reward size in order to explore the importance of reward context in the learning processes.

Three male macaque monkeys (*Macaca Mulatta*) were part of the study: monkey M (five years and ten months old — 7.5 Kg), monkey S (six years and two months old — 8.0 Kg) and monkey D (six years and eight months old – 7.5 Kg). Animal care, housing and experimental procedures conformed to the European (Directive 210/63/EU) and Italian (DD.LL. 116/92 and 26/14) laws on the use of non-human primates in scientific research. The experiments were carried out while the monkey's heads were fixed by means of a head-holder, surgically implanted for allowing forthcoming electrophysiological recordings. The animals were preanesthetized with ketamine (10 mg/kg, i.m.) and anesthetized with isoflurane through a constant flux of isoflorane/air mixture (1–3%, to effect). Antibiotics and analgesics were administered postoperatively. During the experiments, monkeys sat in a primate chair with the head fixed in front of a monitor touch screen (3M™ MicroTouch™ M1700SS 17" LCD touch monitor, 1280 × 1024 resolution). The control of the appearance of the stimuli on the touch screen, reward delivery and monkey's behavioural responses were monitored with the non-commercial software package, CORTEX (NIMH, Bethesda, USA). After each correct trial, the animals were rewarded with apple sauce.

In the Objects-In-Place-Reward task (OIPR), we wanted to assess whether and to what extent the first stage of a learning process, the one-trial learning [7,6] is affected by the amount of expected reward. During the OIPR discrimination learning, two objects, one rewarded and the other unrewarded, were always displayed at the same places

* Corresponding author.
E-mail address: aldo.genovesio@uniroma1.it (A. Genovesio).

[1] These authors contributed equally.

https://doi.org/10.1016/j.bbr.2019.111983
Received 6 March 2019; Received in revised form 3 May 2019; Accepted 25 May 2019
Available online 26 May 2019

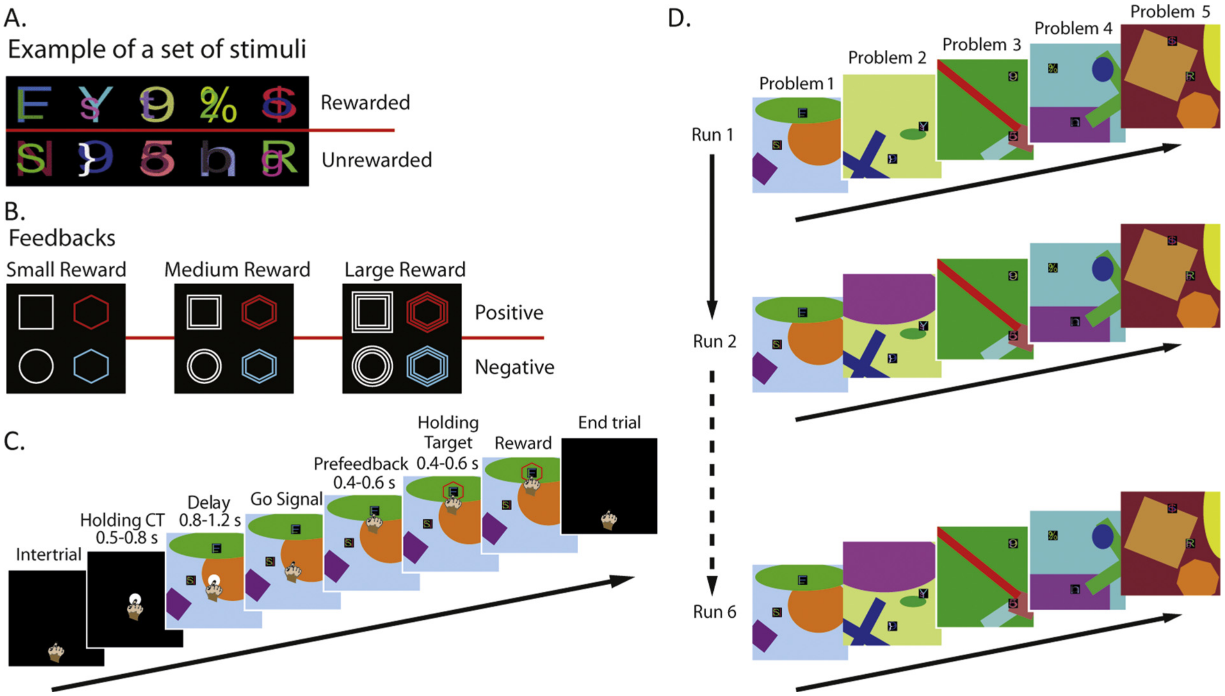

Fig. 1. Task. (A) Examples of stimuli displayed as objects in a problem. Each problem consisted of one rewarded stimulus leading to positive feedback and a reward and one unrewarded stimulus leading to negative feedback. (B) Positive and negative feedback for each amount of reward. Red hexagon(s) or white square(s) was/were displayed around the correct answer if chosen. Blue hexagon(s) or white circle(s) was/were displayed around the wrong object if chosen. One item was displayed in the small reward condition, two concentric items were displayed in the medium reward condition and three concentric items were displayed in the large reward condition. (C) Example of the temporal sequence of a trial. (D) Example of temporal sequence of the six runs. Five problems comprised the first run and were subsequently represented in the same order for six runs. Six runs comprised one complete session.

within a unique background, the scene, and formed the problem that monkeys had to solve. The scene was composed of a random colour background and three geometrical figures randomly selected. An object consisted of two superimposed coloured ASCII characters pseudorandomly generated (Fig. 1A). The randomizations allowed us to show unique patterns of objects-in-scene on the screen and to create unique problems to solve. A trial began with a central target (CT) being presented on the screen, represented by a white circle (Fig. 1C). The monkey had to touch the CT and hold the contact for 0.5 or 0.8 s to let the scene and the objects appear. After a delay period of 0.8 or 1.2 s, the CT disappeared, indicating the go signal for the monkey to touch one of the two objects displayed in the scene. Then, after the monkey maintained the touch for a pre-feedback period (0.4 or 0.6 s), visual feedback surrounding the chosen object was presented. The appearance of the feedback indicated the beginning of an additional holding period (holding target period, 0.4 or 0.6 s) during which time the monkey was required to keep holding the hand on the target until reward delivery in case of a correct response. After a correct response, a reward was delivered and then the screen turned black, while after an error no reward was delivered, and all the objects turned off. If the monkey aborted the trial prior to the feedback appearance, the same problem was presented again. Any completed trial, whether correct or incorrect, was followed by the next problem in the sequence, without the presentation of a correction trial after errors. Five different problems were presented consecutively to the animal, repeated for six runs, for a total of 30 trials (Fig. 1D). Inside each session, the five problems were arbitrarily divided into three different types, based on the amount of reward given: small reward (SR), medium reward (MR) and large reward (LR). Inside each set of five problems the reward sizes were pseudorandomly assigned, in order to have always the three rewards represented and never have three times the same reward. Consequently, in a given day in which the monkeys typically performed on average around 12 sessions, they experienced a similar number of problems associated with each reward size. The shapes of the feedback indicated a correct or incorrect response and the size of the received or missed reward (Fig. 1B). White square(s) or red hexagon(s) indicated a correct response and white circle(s) or a blue hexagon(s) were indicative of an incorrect response (Fig. 1B). The object was surrounded by one shape if a small reward was received (0.15 ml) or missed, by two shapes if a medium reward (twice the amount of small reward, 0.3 ml) was received or missed and by three shapes if a large reward (four times the amount of small reward, 0.6 ml) was received or missed (Fig. 1B).

During the first run, monkeys had no indication regarding which one of the two objects was the rewarded one and they had to guess using a trial and error strategy. Furthermore, monkeys had no indication about the amount of reward associated with the problem to solve before choosing for the first time one of the two targets. After five problems, the first run ended and the second run started with the same five problems presented in the same order. The amount of reward delivered was fixed for each problem along the six runs. The session ended after the monkey completed six runs. We included in the analysis only the complete sessions, that is the sessions with all the 30 trials performed. Each monkey performed 180 sessions, for a total amount of 5400 trials with approximatively 1800 trials performed in each condition of reward.

From this dataset, we calculated the learning curves for the three experimental conditions, small, medium and large reward, for each monkey (Fig. 2 left). As expected, performance did not differ from chance level in the first run in any of the reward level conditions (Monkey M: 50%, 49.3%, 52.3%; Monkey S: 50%, 51.3%, 51.7%; Monkey D: 49.3%, 47.5%, 46.5%; small, medium and large reward respectively; $p > 0.05$ exact binomial test). The learning effect appears in the second run, in which two of three monkeys showed a performance significantly different from chance for all the three reward conditions, while the third one only for the MR and LR conditions (Monkey M: 56.3%, 69.5%, 68.7%; Monkey S: 51.3%, 72%, 86.2%;

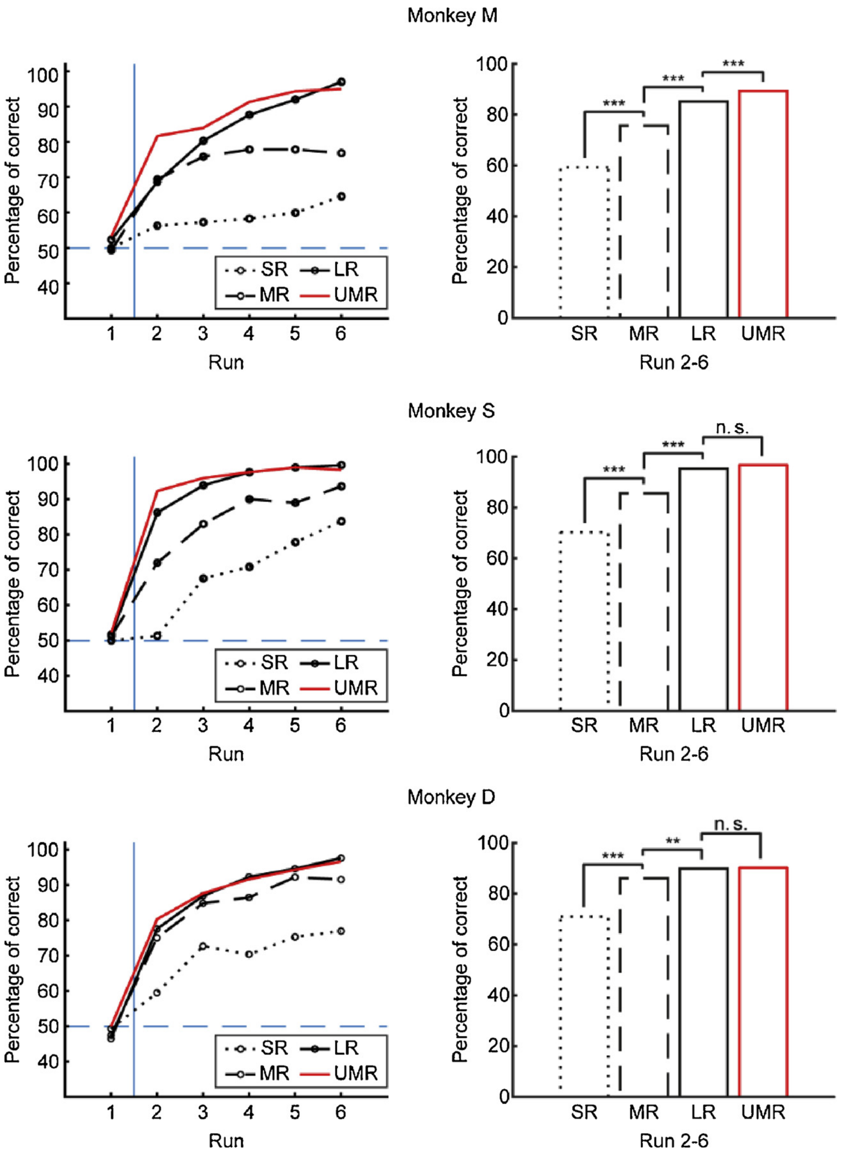

Fig. 2. Learning curves and average performance for the three monkeys. Left: the curves show the mean percentage of correct choices for the six runs in all trials performed in the three different reward conditions of the OIPR task (Small Reward: SR, Medium Reward: MR, Large Reward: LR) in black and the results obtained in our previous study (Ferrucci et al. 2019, Unique Medium Reward: UMR) in red. Vertical lines represent the separation between the first runs performed randomly by the monkeys and the five remaining runs. Horizontal dashed lines represent the chance level (50%). Right: average percentage of correct choices from the second to the sixth run in the same four conditions. The stars represent significant differences between the number of correct choices in a condition divided by the total number of choices made in this condition (** means $p < 0.01$, *** means $p < 0.001$, n. s. means non-significant).

Monkey D: 59.5%, 75.1%, 77.6%; small, medium and large reward respectively; $p < 0.05$ exact binomial test, except for Monkey S in SR, $p > 0.05$). The comparison between the proportions of correct trials in the three conditions during the second run showed a significant effect for the reward size in all of the monkeys (3-sample test for equality of proportions with continuity correction, $p < 0.01$). We also evaluated the learning, calculating the proportion of correct trials for each reward level from the second to the sixth run taken together (Fig. 2 right). The comparison among the three conditions showed a significant difference in all three monkeys (3-samples test for equality of proportions with continuity correction, $p < 0.01$). Pairwise comparisons showed a significant difference between each reward couple (SR-MR, SR-LR and MR-LR) in the three learning curves for all monkeys (Pairwise comparison of proportions with Holm-Bonferroni method for adjustment, $p < 0.01$).

In our previous study [6] we analysed the learning process, examining the learning curve of the same three monkeys used in the present study with a unique reward size equal to the medium reward of the present study in two observational learning conditions and an individual learning condition (Unique Medium Reward condition, UMR). Here, we compared for each monkey, the learning curves obtained in the present study (SR, MR and LR conditions) with the learning curve obtained in the individual learning condition of our previous study (UMR) in which the testing occurred previously. For all three monkeys (Fig. 2 left, red curve), the learning curve in the UMR condition in our previous study corresponds to the learning curve in the LR condition in the present study. For monkey S and monkey D, we did not find a significant difference between the performance in the UMR and the LR condition (Pairwise comparison of proportions with Holm-Bonferroni method for adjustment, Monkey S, $p = 0.07$, $\chi^2 = 3.268$, Monkey D, $p = 0.83$, $\chi^2 = 0.045$), although there is a significant difference between the performance in the UMR and the MR condition (Pairwise comparison of proportions with Holm-Bonferroni method for adjustment, Monkey S, $p < 2.2e\text{-}16$, $\chi^2 = 113.29$; Monkey D, $p < 0.0008$, $\chi^2 = 11.43$) and between the UMR and the SR condition (Pairwise comparison of proportions with Holm-Bonferroni method for adjustment, Monkey S, $p < 2.2e\text{-}16$, $\chi^2 = 377.37$; Monkey D, $p < 2.2e\text{-}16$, $\chi^2 = 175.07$). For monkey M, we found the same tendency. Indeed, the performance was significantly different between the UMR and the LR condition (Pairwise comparison of proportions with Holm-Bonferroni method for adjustment, $p < 0.0009$, $\chi^2 = 11.11$), but the performances in the MR and SR conditions were even lower and the differences with the performance in the UMR condition more significant (Pairwise comparison of proportions with Holm-Bonferroni method for adjustment, UMR vs MR, $p < 2.2e\text{-}16$, $\chi^2 = 95.92$; UMR vs SR, $p < 2.2e\text{-}16$, $\chi^2 = 325.26$).

In the current study, we used a modified version of the Object-In-Place task [7] to test whether in a learning task a different amount of reward could influence the learning speed and rate. From the learning curves (Fig. 2 left), it appears that the reward size influenced the performance (i.e., the higher the reward level, the higher the performance in those trials). Moreover, the reward size played a crucial role in affecting the learning rate as early as the second run of each session. Interestingly, this effect was then constant through the six runs and not limited to the second run, leading to a highly significant difference in the overall performance between the three reward conditions. These results suggest two conclusions. On one hand, a single run is sufficient to elicit a stronger memory for the highly rewarded problems. On the other hand, this effect is maintained through time, at least for the duration of the six runs, in which the monkey's performance in the highly rewarded problems is kept higher compared to problems with lower rewards. In the first run, the animals were not provided with any indication regarding the reward size until the appearance of the feedback, that is after the choice was made, and nevertheless the reward effects were already apparent in the learning curves starting from the second run. That task feature gave us the opportunity to capture the turning point in which the reward size could affect the learning process, that is after the monkey's choice during the first run, at the appearance of the feedback and during the reward delivery (or its absence in case of error trials). It is important to notice that the entire scene is still present on the screen; the combination of feedback and reward/absence of reward with a specific scene allowed the occurrence of rapid learning. Our results suggest that reward size can influence directly the attention toward specific scenes at this later moment of the trial after the monkey's choice, leading to a better performance. Finding the learning effect on the second run also means that the knowledge and expectation of a greater reward size is not required during the choice to promote more rapid learning. This difference in the learning rate of the animals could offer a tool in electrophysiological studies to investigate the neural substrates of fast learning processes.

Some works have already discussed the enhancement of performance as a direct consequence of increased attention. When more difficult stimuli are mixed with easier stimuli in a matching to sample task with primates, the performance for those stimuli is usually lower than when they are tested alone [8]. Increased attention toward a specific stimulus due to the higher difficulty of the task may result in an increased discrimination of that stimulus and therefore in an enhanced behavioural performance in general. Studies with human subjects have also helped to shed light on the link between attention and reward. In an icon-learning task, subjects were instructed to learn a stimulus response association when some icons and distractors were presented on the left or on the right side of the screen, while the reward size and the attention were varied across trials [9]. Subjects learned the association only for those icons presented on the side of the screen in which they were instructed to pay attention, but the reward size did not affect the learning rate. Despite that result, a slightly significant effect of the reward size on the subject's performance was found when they were further tested in a subsequent task with the same associations they had previously learned. This small influence of the reward size could be explained by the nature of the reward itself (a system of points earned), which may have not been particularly effective. Indeed, the importance of reward size has been indicated by another work using monetary gain [10], in which the effect of the manipulation of the reward level in a visual discrimination task was tested. Rejecting stimuli in the test phase was more difficult when they were associated with higher reward in the training phase and vice versa for those stimuli that were associated with lower reward. It emerges from these studies that attentional processes are affected by the past reward experiences. The information regarding how successful a particular action or choice was, is stored and it can influence future actions and choices. In our task, the effect of the reward size on learning rate is strongly evident in all of the three monkeys and it could be explained by an attentional boost occurring after the choice is made already in the first run of each session.

Another important result of our study is the similarity between the performance of the monkeys in the highly rewarded condition and their performance in the same task in which only one size of the reward was provided, equivalent to the medium size of the current task [6]. Indeed, we expected to observe an increase of the monkey's performance when a larger reward was expected and a decrease in their performance when a lower reward was expected. Instead, we observed a readjustment of the monkey's performance. They performed the task in the large reward condition as they performed it previously for a medium reward and their accuracy decreased in the other conditions. This result could indicate the importance of context in which a choice is performed. Indeed, when only one amount of reward is present, every trial elicits a maximal value for the individual and differently, when different amounts are utilized, this value is updated and in relative terms, the value of the medium reward is downgraded. The maximal value for the monkey becomes the large reward and their attention is maximized in those trials. Indeed, the motivational value of a reward can be relative, and it is usual to observe negative and positive contrast effects, also referred to as undershooting and overshooting effects [11–16]. A positive contrast effect is observed when the performance is enhanced in a large reward condition after a shift from a small reward condition compared to a large reward condition in a pre-shift period. Instead, a negative contrast effect is observed when the performance is diminished when shifted from larger to smaller reward comparing to the same small reward in a pre-shift period [11,12]. In our case, the paradigm was different because we did not shift from one reward size to another one in blocks, but rather we proposed to the monkey three reward conditions instead of one. However, comparing the performance of the animals in the medium reward condition, we observed a diminished accuracy that could be interpreted as a negative contrast effect due to the presence of a larger reward. More generally, we observed a re-evaluation of the benefit of the action depending on the reward context in which the monkey is acting. On the one hand, the encoding of the reward or its expectation can be context independent (i.e. same encoding between different tasks [17] or between different effort-level conditions [18]) but on the other hand, several studies described context-dependent reward effects in which a reward is represented in terms of its relative value in various structures of the so-called reward circuit, such as the striatum [19], the amygdala [20,21], the orbitofrontal cortex [21–23], the ventromedial prefrontal cortex [24] and the dopaminergic neurons of the midbrain [25]. Another possible explanation to account for the context effect is the increase in task complexity for the introduction of different reward size as a new variable making the task more demanding for the medium reward condition compared to our previous study [6]. However, against this interpretation we should consider at least the fact that the failure to integrate the reward size information should not affect directly the performance because it was not a task requirement.

In conclusion our results suggest that the relative reward value, determined by the presence of other possible reward sizes, can influence the learning of the association between action and outcome and show the close link between reward in relative terms and learning.

Author contributions

A.G. designed the study. S.N. and L.F. conducted the experiment. S.N. and L.F. analyzed the data. A.G., S.N., L.F. and E.B. wrote the paper.

Conflict of interest

All authors have no conflict of interest to declare.

Declarations of interest

None

Acknowledgments

This work was supported by the ERC 2014 grant (European Research Council, 648734-HUMO).

References

[1] R.S. Sutton, A.G. Barto, Reinforcement Learning: An Introduction, (1998).

[2] B. Fischer, D. Wegener, Emphasizing the "positive" in positive reinforcement: using nonbinary rewarding for training monkeys on cognitive tasks, J. Neurophysiol. 120 (2018) 115–128.

[3] W. Schultz, Multiple reward signals in the brain, Nat. Rev. Neurosci. 1 (2000) 199–207.

[4] T. Minamimoto, G. La Camera, B.J. Richmond, Measuring and modeling the interaction among reward size, delay to reward, and satiation level on motivation in monkeys, J. Neurophysiol. 101 (2008) 437–447.

[5] L. Stanisor, C. van der Togt, C.M.A. Pennartz, P.R. Roelfsema, A unified selection signal for attention and reward in primary visual cortex, Proc. Natl. Acad. Sci. 110 (2013) 9136–9141.

[6] L. Ferrucci, S. Nougaret, A. Genovesio, Macaque monkeys learn by observation in the ghost display condition in the object-in-place task with differential reward to the observer, Sci. Rep. 9 (2019) 401.

[7] D. Gaffan, Scene-specific memory for objects: a model of episodic memory impairment in monkeys with fornix transection, J. Cogn. Neurosci. 6 (1994) 305–320.

[8] H. Spitzer, R. Desimone, J. Morant, Increased attention enhances both behavioral and neuronal performance, Science 240 (1988) 338–340.

[9] D. Vartak, D. Jeurissen, M.W. Self, P.R. Roelfsema, The influence of attention and reward on the learning of stimulus-response associations, Sci. Rep. 7 (2017) 1–12.

[10] C. Della Libera, L. Chelazzi, Learning to attend and to ignore is a matter of gains and losses, Psychol. Sci. 20 (2009) 778–784.

[11] L.P. Crespi, Quantitative variation of incentive and performance in the white rat, Am. J. Psychol. 55 (1942) 467–517.

[12] R.W. Black, Shifts in magnitude of reward and contrast effects in instrumental and selective learning: a reinterpretation, Psychol. Rev. 75 (1968) 114–126.

[13] W.M. Cox, A review of recent incentive contrast studies involving discrete-trial procedures, Psychol. Rec. 25 (1975) 373–393.

[14] C.F. Flaherty, Problems in the Behavioural Sciences, No. 15. Incentive Relativity, Cambridge University Press, New York, NY, US, 1996.

[15] B.A. Williams, Varieties of contrast: a review of incentive relativity, J. Exp. Anal. Behav. 68 (1997) 133–141.

[16] J.B. Engelmann, G. Hein, Contextual and Social Influences on Valuation and Choice, 1st ed., Elsevier B.V., 2013.

[17] E. Marcos, S. Nougaret, S. Tsujimoto, A. Genovesio, Outcome modulation across tasks in the primate dorsolateral prefrontal cortex, Neuroscience 371 (2018) 96–105.

[18] S. Nougaret, S. Ravel, Modulation of tonically active neurons of the monkey striatum by events carrying different force and reward information, J. Neurosci. 35 (2015).

[19] H.C. Cromwell, O.K. Hassani, W. Schultz, Relative reward processing in primate striatum, Exp. Brain Res. 162 (2005) 520–525.

[20] M.A. Bermudez, W. Schultz, Reward magnitude coding in primate amygdala neurons, J. Neurophysiol. 104 (2010) 3424–3432.

[21] A.R. Saez, A. Saez, J.J. Paton, B. Lau, C.D. Salzman, Distinct roles for the amygdala and orbitofrontal cortex in representing the relative amount of expected reward, Neuron 95 (2017) 70–77.

[22] L. Tremblay, W. Schultz, Relative reward preference in primate orbitofrontal cortex, Nature 398 (1999) 704–708.

[23] C. Padoa-Schioppa, X. Cai, The orbitofrontal cortex and the computation of subjective value: consolidated concepts and new perspectives, Ann. N. Y. Acad. Sci. 1239 (2011) 130–137.

[24] R. Abitbol, M. Lebreton, H. Guillaume, B.J. Richmond, S. Bouret, M. Pessiglione, Neural mechanisms underlying contextual dependency of subjective values: converging evidence from monkeys and humans, J. Neurosci. 35 (2015) 2308–2320, https://doi.org/10.1523/JNEUROSCI.1878-14.2015.

[25] P.N. Tobler, C.D. Fiorillo, W. Schultz, Adaptive coding of reward value by dopamine neurons, Science 307 (2005) 1642–1645.