UNIVERSITY OF AMSTERDAM

UvA-DARE (Digital Academic Repository)

(Mal)Adaptive Learning After Switches Between Object-Based and Rule-Based Environments

Schaaf, J.V.; Xu, B.; Jepma, M.; Visser, I.; Huizenga, H.M.

[Link to publication](#)

ORIGINAL PAPER

(Mal)Adaptive Learning After Switches Between Object-Based and Rule-Based Environments

Jessica V. Schaaf[1] · Bing Xu[2] · Marieke Jepma[1] · Ingmar Visser[1,3,4] · Hilde M. Huizenga[1,3,4]

Abstract

In reinforcement-learning studies, the environment is typically object-based; that is, objects are predictive of a reward. Recently, studies also adopted rule-based environments in which stimulus dimensions are predictive of a reward. In the current study, we investigated how people learned (1) in an object-based environment, (2) following a switch to a rule-based environment, (3) following a switch to a different rule-based environment, and (4) following a switch back to an object-based environment. To do so, we administered a reinforcement-learning task comprising of four blocks with consecutively an object-based environment, a rule-based environment, another rule-based environment, and an object-based environment. Computational-modeling results suggest that people (1) initially adopt rule-based learning despite its suboptimal nature in an object-based environment, (2) learn rules after a switch to a rule-based environment, (3) experience interference from previously-learned rules following a switch to a different rule-based environment, and (4) learn objects after a final switch to an object-based environment. These results imply people have a hard time adjusting to switches between object-based and rule-based environments, although they do learn to do so.

Introduction

In reinforcement-learning studies, the environment is usually object-based (also called non-generalizable). That is, people have to choose between pictures of objects, and these objects themselves are predictive of a reward. For example, choosing the picture of the cat generally leads to a reward whereas choosing the picture of the house does not. In a rule-based environment (also called generalizable), people have to

choose between pictures that comprise multiple dimensions (e.g., color, pattern, and shape), and these dimensions are predictive of a reward. For example, choosing squared stimuli generally leads to a reward whereas choosing round stimuli does not. So in this example, the rule is "all squared stimuli are correct", irrespective of their color and pattern. Learning in object-based environments has often been studied in the reinforcement-learning literature (Rescorla & Wagner, 1972; Sutton & Barto, 2018); learning in rule-based environments has recently received more attention (Balcarras & Womelsdorf, 2016; Ballard et al., 2018; Collins et al., 2014; Farashahi et al., 2017, 2020; Geana & Niv, 2014; Leong et al., 2017; Niv et al., 2015; Radulescu et al., 2016; Wilson & Niv, 2012; Wunderlich et al., 2011). In the current study, we build on these two literatures by examining how people learn (1) in an object-based environment, and how they adapt their behavior (2) following a switch from an object-based environment to a rule-based environment, (3) following a switch from a rule-based environment to a rule-based environment governed by another rule (e.g., from "all squared stimuli are correct" to "all striped stimuli are correct"), and (4) following a switch back from a rule-based environment to an object-based environment. We investigate

✉ Jessica V. Schaaf
j.v.schaaf@uva.nl

1. Department of Psychology, University of Amsterdam, Nieuwe Achtergracht 129-B, 1018 WS Amsterdam, The Netherlands

2. Department of Child and Adolescent Psychiatry/Psychology, Erasmus MC, University Medical Center Rotterdam, Wytemaweg 80, 3000 CB Rotterdam, The Netherlands

3. Yield, Research Institute for Child Development and Education, Nieuwe Achtergracht 129-B, 1018 WS Amsterdam, The Netherlands

4. ABC, Amsterdam Brain and Cognition Centre, Nieuwe Achtergracht 129-B, 1018 WS Amsterdam, The Netherlands

these questions by inspecting choice accuracy and by comparing object-based and rule-based computational reinforcement-learning models and parameter estimates governing the best-fitting models.

In object-based models, the value of an object is updated when this object is chosen. For example, the value of the picture of the cat increases when choosing this picture led to a reward. In contrast, in rule-based models, the values of stimulus features are updated. For example, the values of "red", "striped", and "square" increase when choosing the red striped square led to a reward. Because of this feature-value updating, learning about the red striped square also informs learning about other red, striped, and squared stimuli. This aspect of rule-based learning makes learning faster in rule-based environments.

With respect to our first question, Farashahi and colleagues (Farashahi et al. 2017, 2020) investigated how people learn in an environment in which stimuli comprise multiple dimensions but no single dimension is predictive of a reward, that is, an object-based environment. These studies showed that people tend to learn rules in the beginning of the task, but adopt object-based learning at the end of the task. However, these tasks were complex as pairs changed across trials within a block and some options were thus correct in one pairing, but incorrect in another pairing. Outside the reinforcement-learning literature, it is known that complex tasks promote the use of simple rules (e.g., Gigerenzer et al., 1999); therefore, task complexity may have induced this rule-based learning. In the current study, we build on these studies by adopting a novel, simpler, task in which pairs were fixed within a block. We hypothesized that in this simpler task, participants would be able to learn that objects were predictive of reward and thus to adopt object-based learning throughout the block.

Concerning our second question, several reinforcement-learning studies showed that people learn rules in a rule-based environment not preceded by an object-based environment (Ballard et al., 2018; Collins et al., 2014; Geana & Niv, 2014; Leong et al., 2017; Niv et al., 2015; Radulescu et al., 2016). In the current study, we build on this literature by investigating whether and how people adapt their behavior following a switch from an object-based to a rule-based environment.[1] Given this previous literature, we hypothesized that people would be able to adopt rule-based learning in a rule-based environment preceded by an object-based environment. Also, we hypothesized that accuracy would be higher in this rule-based environment compared to the

preceding object-based environment because rule-based learning is faster compared to object-based learning.

Regarding our third question, three reinforcement-learning studies investigated how people learn in a rule-based environment in which the rewarding dimension changes across trials (Marković et al., 2015; Wilson & Niv, 2012; Wunderlich et al., 2011). Computational modeling showed that people adopt rule-based instead of object-based learning. In the current study, we extend this literature by testing whether people experience interference from the previously-learned rule. We do this by inspecting the learning parameters in the best-fitting models. We hypothesized that people adopt rule-based learning in the new rule-based environment, but that they would experience interference from the previously-learned rule as previously shown outside the reinforcement-learning literature (Best et al., 2013; Bröder & Schiffer, 2006; Hoffmann et al., 2019; Kämmer et al., 2013). We thus expected that people would initially apply the previous rule and only later would apply the current rule. Also, because of this interference, we hypothesized that accuracy would be lower in the new rule-based environment.

With respect to our fourth question, to our knowledge, no reinforcement-learning studies have yet investigated how people learn following a switch from a rule-based environment back to an object-based environment. Again, in our simple task, we hypothesized that people would adopt object-based learning after a switch to this environment, and that accuracy in this object-based environment would be lower than that in the preceding rule-based environment because object-based learning is slower as compared to rule-based learning.

In the current preregistered study, we investigated these four questions by administering a two-armed probabilistic reinforcement-learning task with three-dimensional stimuli, characterized by one feature on each dimension (e.g., a red (color dimension) striped (pattern dimension) square (shape dimension)). In each of four blocks, either objects or features were probabilistically related to a reward. In the first, object-based, block, objects (hence no single dimension) were predictive of reward. In the second, rule-based, block, a single dimension was predictive of reward. In the third, rule-based, block, another dimension was predictive of reward. And in the fourth, object-based, block, again objects were predictive of reward. This enabled us to investigate how people learned in an object-based environment (block 1), how they adjusted their behavior following a switch to a rule-based environment (block 2 versus 1), how they adjusted their behavior following a switch to a different rule-based environment (block 3 versus 2), and how they adjusted their behavior following a switch back to an object-based environment (block 4 versus 3). We analyzed these data with both regression analyses on accuracy and by fitting computational models (Leong et al., 2017; Niv et al., 2015; Rescorla &

---

[1] Note Farashahi and colleagues (Farashahi et al., 2017, 2020) did investigate learning in object-based and rule-based environments, but did not test whether people adaptively switched between these environments.

Fig. 1 Example of four pairs presented in each block to a fictive participant. On top of each panel, the block type is indicated. In each panel, the rows represent the pairs; the options on the left of the vertical line are the correct options; the options on the right the incorrect options. Stimuli were inspired by Niv et al. (2015)
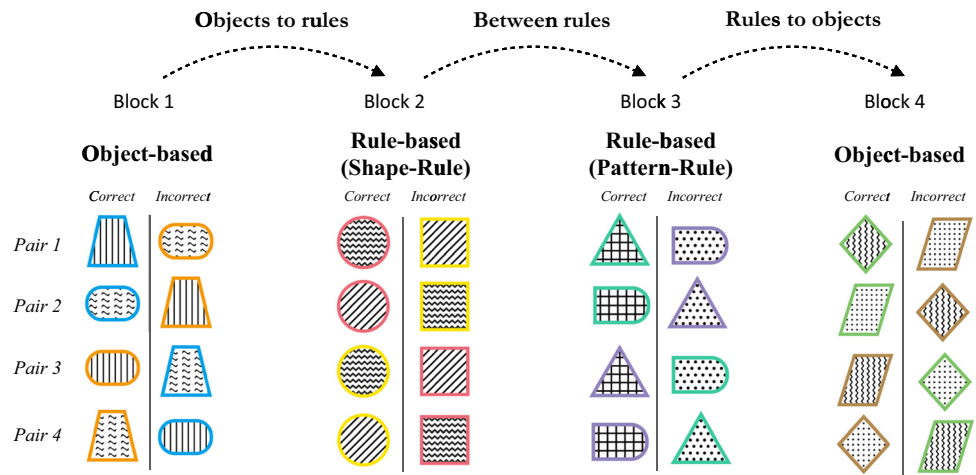
Wagner, 1972; Sutton & Barto, 2018); the latter allowed us to uncover whether participants used object-based or rule-based learning and to investigate learning parameters governing the best-fitting models.

Method

This study was preregistered as *Reinforcement Learning of Rules* on https://osf.io/a2zmp. We followed all preregistered procedures, except for some small deviations which are addressed in Online Resource A. Non-preregistered analyses are considered exploratory and reported as such.

Participants A total of 43 participants were recruited via the University of Amsterdam. According to preregistered criteria, data from one participant were removed because this participant did not finish the task. In addition, data from six participants were removed because they indicated they had used alcohol or recreational drugs on the day of testing or indicated a lack of understanding of the task. The final sample thus consisted of 36 adults (22 female; $M_{age} = 22.3$ (3.36), range: 18–31 years). No included participants had diagnosed psychological or neurological problems, or color blindness. All participants actively consented and received €5 or research credits plus a variable bonus between €0 and €2 ($M_{bonus} = €0.52$) depending on their performance on the task.

Experimental Design Participants performed a two-armed probabilistic reinforcement-learning task. The task comprised two types of blocks. In object-based blocks, the correct option in each pair could be predicted neither by a single dimension nor by a combination of two dimensions; that is, all three dimensions in combination were required to choose the correct option. In rule-based blocks, the correct option in each pair could be predicted by a single dimension.

Participants were randomly assigned to one of two task versions: The pattern-to-shape version or the shape-to-pattern version. All participants started with an object-based block. Hereafter, two rule-based blocks followed. Participants in the pattern-to-shape version were presented with a pattern-rule in the second block and a shape-rule in the third block. In contrast, participants in the shape-to-pattern version were presented with a shape-rule in the second block and a pattern-rule in the third block. In the fourth block, all participants again completed an object-based block.

Reinforcement-Learning Task On each trial, participants were presented with two options; each was characterized by one feature (e.g., "red" or "striped" or "square") on each of three dimensions (i.e., color, pattern, and shape; Fig. 1; Niv et al., 2015). The same dimensions were used across blocks, but features differed between blocks. The two options in each pair differed in all three dimensions. In each block, four pairs were presented, 20 times each (i.e., the average number of trials per game used by Niv et al., 2015; 80 trials in total). The order of the four pairs within a block was determined randomly per four trials to ensure that pairs were presented a maximum of twice in a row. The options were presented on the left and right side of the screen in a counterbalanced order. Which option was correct was determined randomly for each participant and did not change across the trials of a block.

To get acquainted with the task, participants first completed a practice block with 4 pairs, each presented 6 times (24 trials in total). This practice block was an object-based block to prevent biasing participants toward learning rules. The practice block was performed without time limitations to promote accuracy over speed. Participants were instructed that they would play "four game rounds" and that each game round would "last approximately 8 min", but they were unaware of the exact trial numbers. Furthermore, they were told that one of the options in each pair yielded the highest

Fig. 2 Example trial of the reinforcement-learning task
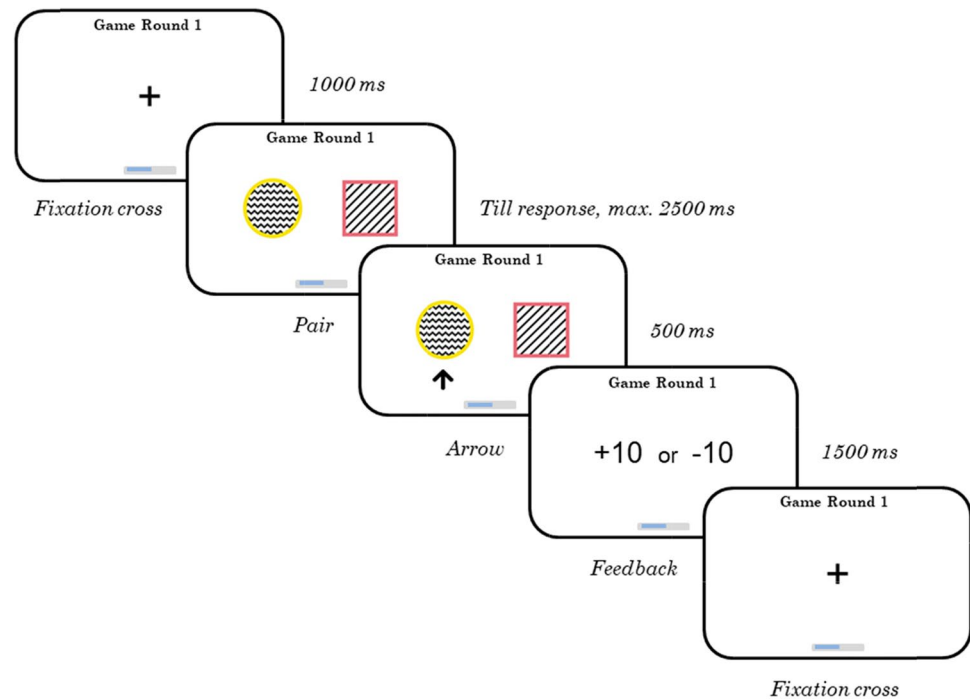
reward (i.e., "Choices for one stimulus lead to more points ('right' choice) than choices for the other stimulus ('wrong' choice)"), and were instructed "to win as many points as possible". As in most reinforcement-learning studies, we used probabilistic feedback. Congruency of the feedback was 75%; that is, in 75% of the cases, participants gained points ($+10$) after a correct choice and lost points (-10) after an incorrect choice, and in 25% of the cases they lost points after a correct choice and gained points after an incorrect choice. Participants were instructed that "the feedback was usually correct, but not always". We fixed the order of congruent versus incongruent feedback across participants to rule out individual differences in task difficulty due to congruency.

As can be seen in Fig. 2, each trial started with a fixation cross (1000 ms), followed by presentation of a pair (RT to max. 2500 ms). Participants chose between the two options by pressing the "z" or "/" key, for the left and the right option respectively, on a qwerty keyboard. After a choice was made, a black arrow was presented below the chosen option (500 ms), followed by the points gained or lost (1500 ms). If a choice was not made within 2.5 s, "Too late!" appeared on the screen (1500 ms) and participants lost 10 points. The next trial was again signaled by a fixation cross. At the top of the screen, the current game round was indicated; at the bottom of the screen, a progress bar kept track of the proportion already-administered trials (Fig. 2).

Procedure All participants were tested individually in a lab cubicle. After signing an informed consent form,

the participant took place behind a computer screen and received on-screen instructions. Then they performed the reinforcement-learning task. At the end of the experiment, (bonus) money or research credits were paid out.

Computational Models

To assess whether people applied object-based or rule-based learning, we fitted computational reinforcement-learning models to participants' choice data in each block separately (all data and code are freely available on https://osf.io/rvcx5/). We considered 5 object-based and 5 rule-based models. The models differed in whether learning rates were dynamic or static across trials, and in whether learning rates were equal or unequal across pairs (for object-based models) or across dimensions (for rule-based models). From these 10 models, we selected the best-fitting model and inspected the parameter estimates in this best-fitting model. Below we only discuss the most-complicated versions of the models. Details on all models, the estimation procedure, and full results are presented in Online Resource B.

Object-Based Versus Rule-Based Learning

In object-based models, people update the value of objects. On each trial t ($T = \{1, \ldots, 80\}$) a pair p ($P = \{1, \ldots, 4\}$) is presented, people update the value Q of the chosen option s ($S = \{1, 2\}$). They do so with a proportion (i.e., learning rate (LR)) of the prediction error (PE), the difference between the observed outcome and the value of the chosen option. In the

most-complicated object-based model, we modeled dynamic learning rates (through initial learning rate *LR1* and decay parameter α) that were allowed to be unequal across pairs:

$$Q(p, s = \text{chosen}, t + 1) = Q(p, s = \text{chosen}, t) + LR(p, t) \times PE(t) \tag{1}$$

with

$$LR(p, t) = LR1(p) \times t^{-\alpha(p)} \tag{2}$$

and

$$PE(t) = \text{Outcome}(t) - Q(p, s = \text{chosen}, t) \tag{3}$$

These object-based models thus implement that people update the value of each chosen object. As a result, object values are updated on a maximum of 20 out of the 80 trials. In rule-based environments, this aspect of object-based learning results in slow learning.

In rule-based models (Leong et al., 2017; Niv et al., 2015), people update the value of features, instead of the value of objects. On each trial, people update the value *V* of each feature f ($F = \{1, 2\}$) on dimension *d* ($D = \{\text{color}, \text{pattern}, \text{shape}\}$) present in the chosen option. Again, this updating is done with a proportion of the prediction error. In the most-complicated rule-based model, we modeled dynamic learning rates in a similar way as in the most-complicated object-based model, allowing them to be unequal across dimensions:

$$V(d, f(s = \text{chosen}), t + 1) = V(d, f(s = \text{chosen}), t) + LR(d, t) \times PE(t) \tag{4}$$

with

$$LR(d, t) = LR1(d) \times t^{-\alpha(d)} \tag{5}$$

and

$$PE(t) = \text{Outcome}(t) - Q(p, s = \text{chosen}, t) \tag{6}$$

The value *Q* of the chosen option *s* is computed by adding the values *V* of the features *f* present in that option; equally weighing each dimension (i.e., 1/3[2]):

$$Q(p, s = \text{chosen}, t + 1) = \sum_{\substack{D = \{color, \\ pattern, shape\}}} \frac{1}{3} \times V(d, f(s = \text{chosen}), t + 1) \tag{7}$$

These rule-based models thus implement that people update the value of all features present in the chosen option. As a result, learning for one pair contributes to learning of other pairs on each of the 80 trials. In rule-based

environments, this aspect of rule-based learning results in faster learning compared to object-based learning.

Model Selection Procedure

Model selection was done by means of the Deviance Information Criterion (DIC; Spiegelhalter et al., 2002), which was transformed into a model weight (Wagenmakers & Farrell, 2004) indicating the probability of each model being the best-fitting model given the model set. The model with the highest weight above 0.9 was referred to as "best-fitting". In case none of the weights exceeded 0.9, and thus no model clearly fitted the data best, we ordered the weights from highest to lowest and coined all models for which the cumulative sum of the weights was above 0.9 as "similar-fitting". We interpreted learning parameters of best-fitting and similar-fitting models.

Regression Models

To test whether people adapted their behavior following switches in learning environments, we preregistered to perform a multilevel logistic regression analysis on participants' choice accuracy and to perform a multilevel linear regression on participants' response times (cf. Online Resource C). However, the multilevel logistic regression analysis failed to converge, and therefore we omitted random effects, thereby yielding a regular logistic regression analysis. For this logistic regression, we included learning condition (*within*; blocks 1 to 4; treated as factor), trial (*within*; linear effect; coded backwards to estimate main effects at the final trial), version (*between*; pattern-to-shape or shape-to-pattern), and their interactions as predictors. We ran three contrasts to assess behavior adaptation from an object-based to a rule-based environment (block 2 versus block 1), from a rule-based environment to a different rule-based environment (block 3 versus block 2), and from a rule-based environment back to an object-based environment (block 4 versus block 3). Details on the preregistered multilevel regression analysis on response times and the results of this analysis are reported in Online Resource C and D.

Results

We applied computational modeling and performed regression analyses to assess whether people used object-based or rule-based learning and whether they adapted their behavior following switches in learning environments. Below, per question, we first discuss computational-modeling results,

---

[2] Because the inverse temperature (see below) scales to the range of *Q*-values, 1/3 instead of 1 is chosen, to ensure inverse-temperature estimates are comparable across models.

Fig. 3 In each panel, black line segments represent binned (4 blocks of 10×8 trials) observed data with 1 SEM; blue line segments represent predicted data for the best-fitting models. In the second block, triangles represent data from pps in the pattern-to-shape version and squares from pps in the shape-to-pattern version
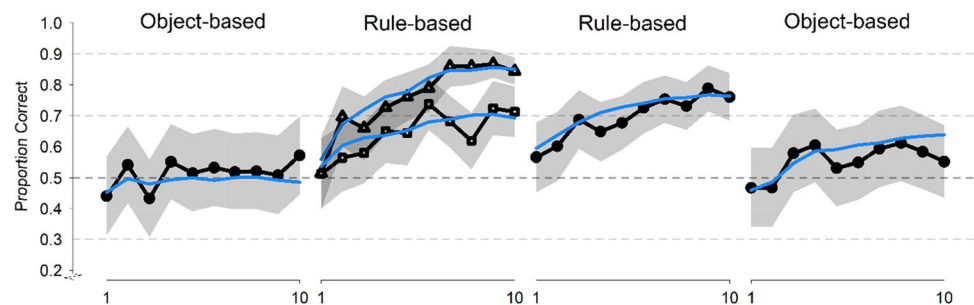
then report exploratory regression analysis on accuracy per block, and finally report preregistered regression analyses comparing accuracy across blocks. In these latter regression analyses, we only report effects including block as we were interested in switching behavior and thus block comparisons. Accuracy data can be found in Fig. 3. Response-time data, and analyses on these data, can be found in Online Resource C and D.

We fitted the computational models aggregating data over both task versions, i.e., pattern-to-shape and shape-to-pattern, except for block 2 because the regression analyses on accuracy revealed a difference between the two versions in this block (see the "Regression Analyses on Accuracy" section). The best-fitting models in each block are discussed below alongside interpretations of the parameter estimates in these models. Parameter estimates and parameter-comparison results can be found in Online Resource E.

Learning in an Object-Based Environment

Computational Modeling

We expected participants to learn objects in the first object-based block; however, our model-comparison results suggested that they adopted rule-based learning. A rule-based model including a dynamic learning rate that was unequal across dimensions fitted the data best. The most-attended dimension, that is, the dimension with the highest learning rate, was "pattern", followed by "color" (Fig. 4, Online Resource E).

Regression Analyses on Accuracy

In the first block, with respect to improvement across trials, accuracy results showed a main effect of trial ($z = -2.51$, $p = 0.01$), but no trial × version interaction ($p = 0.72$). This indicates that participants improved across trials and that this improvement was similar across versions. With respect to the endpoint of learning, accuracy results showed no main effect of version ($p = 0.91$), indicating the endpoint of learning was similar across versions. In an exploratory analysis, we tested whether

participants performed above chance level in the final bin. This proved to be the case ($t(35) = 2.97$, $p = 0.005$), but only slightly so ($M = 0.57$ [0.52; 0.62]).

Together, these computational-modeling and accuracy results indicate that, regardless of its suboptimal nature, participants adopted rule-based learning in an object-based environment. In this rule-based learning, they had a preference for the pattern dimension. Suboptimality of rule-based learning may have resulted in low final accuracy.

### Learning Following a Switch from an Object-Based to a Rule-Based Environment

Computational Modeling

Participants adopted rule-based learning in the second rule-based block. In the pattern-to-shape version (that is, "pattern" is the relevant dimension in the second block), two rule-based models fitted the data similarly: a model with a dynamic learning rate that was equal across dimensions and a model with a static learning rate that was
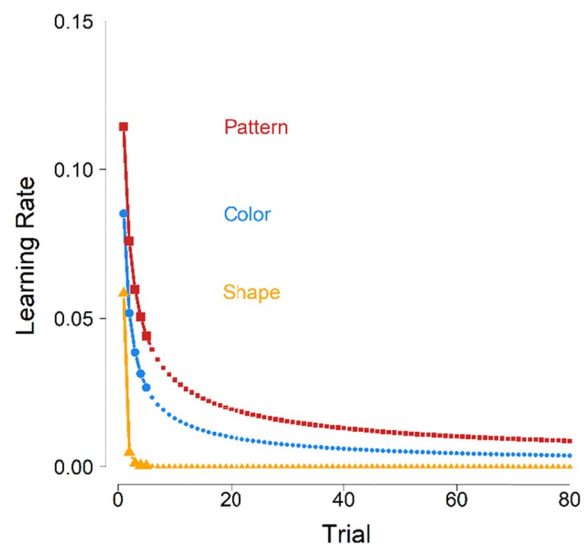
Fig. 4 Estimated learning rates of the best-fitting model in block 1

Fig. 5 Estimated learning rates of the best-fitting models in block 2 in the pattern-to-shape (left) and shape-to-pattern (right) version. In the left panel, the colored lines represent the estimates of the static learning-rate model; the black line of the dynamic learning-rate model
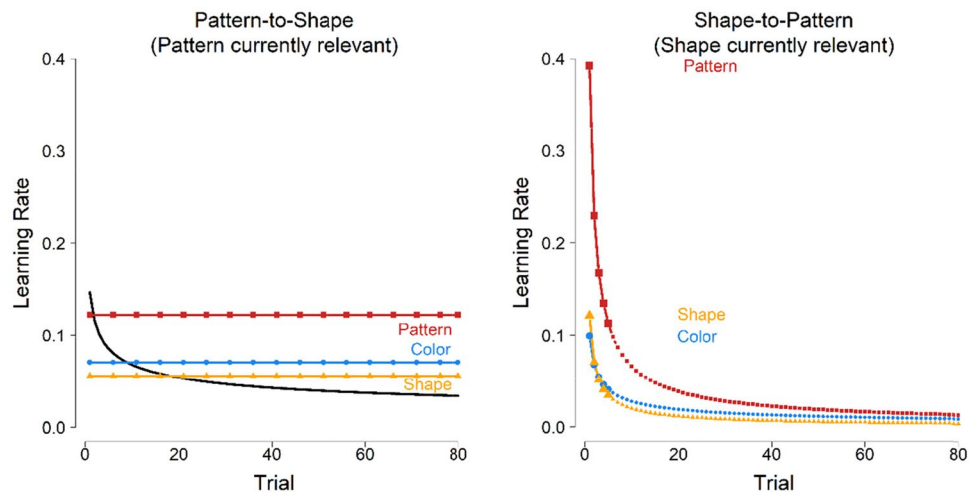
unequal across dimensions. Parameter estimates from the latter model suggested participants mostly attended to the relevant dimension, i.e., "pattern" (Fig. 5, Online Resource E); note this was also the dimension they attended to in the previous block.

In the shape-to-pattern version (that is, "shape" is the relevant dimension in the second block), a rule-based model including a dynamic learning rate that was unequal across dimensions fitted the data best. More specifically, results suggested participants focused on "pattern" instead of the relevant dimension "shape" (Fig. 5, Online Resource E); note again that this was also the dimension they attended to in the previous block.

To further explore the result that in block 2 participants learned the pattern-rule, but not the shape-rule, we split the data in block 2 into four subsets of 20 trials (separately for each version) and inspected for each subset the fit of all computational models and parameter estimates of the best-fitting models (cf. Online Resource F). Results indicated that, in the pattern-to-shape version, in the first 20 trials, participants paid similar attention to all three dimensions, and that they attended to "pattern" after 20 trials. In the shape-to-pattern version, in the first 20 trials, participants attended to "pattern" and "color", in trials 20 to 40, they paid similar attention to all three dimensions, and only in the final 40 trials, they attended to "shape".

Regression Analyses on Accuracy

In the second block, with respect to improvement across trials, accuracy results showed a main effect of trial ($z = -6.69$, $p < 0.001$), and a trial $\times$ version interaction ($z = -2.65$, $p = 0.008$). This indicates that participants improved across trials and that this improvement differed between versions. Follow-up tests in each version showed that participants

that learned the pattern-rule (i.e., pattern-to-shape version; $z = -5.19$, $p < 0.001$) improved faster across trials compared to participants that learned the shape-rule (i.e., shape-to-pattern version; $z = -3.82$, $p < 0.001$). With respect to the endpoint of learning, accuracy results showed a main effect of version ($z = 3.44$, $p = 0.001$), indicating the endpoint of learning was higher for participants that learned the pattern-rule.

Comparing the second block to the first block, with respect to improvement across trials, accuracy results showed a block $\times$ trial interaction ($z = -5.38$, $p < 0.001$) and a block $\times$ trial $\times$ version interaction ($z = -2.41$, $p = 0.02$). This indicates that participants improved faster across trials in the second compared to the first block and that this block $\times$ trial interaction differed between versions. Follow-up tests in each version showed the block $\times$ trial interaction was stronger for participants that learned the pattern-rule in the second block ($z = -5.02$, $p < 0.001$) compared to participants that learned the shape-rule in the second block ($z = -2.36$, $p = 0.02$). With respect to the endpoint of learning, accuracy results showed a main effect of block ($z = 11.5$, $p < 0.001$); moreover, they showed a block $\times$ version interaction ($z = 4.71$, $p < 0.001$). This indicates that the endpoint of learning was higher in the second compared to the first block and that this block effect differed between versions. Follow-up tests in each version showed the block effect was stronger for participants that learned the pattern-rule in the second block ($z = 10.3$, $p < 0.001$) compared to participants that learned the shape-rule in the second block ($z = 5.61$, $p < 0.001$).

Together, these computational-modeling and accuracy results suggest participants applied the pattern-rule rather quickly, whereas it took participants some time to overcome their preference to rely on the pattern dimension and apply the shape-rule in block 2. This difference between rules may have resulted in a smaller difference in accuracy between
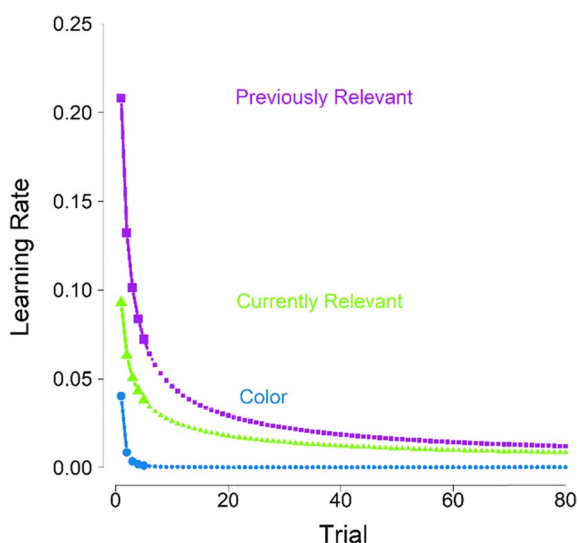
Fig. 6 Estimated learning rates of the best-fitting model in block 3. The previously-relevant and currently-relevant dimensions were "pattern" and "shape" for the pattern-to-shape version respectively, and "shape" and "pattern" for the shape-to-pattern version respectively

blocks 2 and 1 in the shape-to-pattern version compared to the pattern-to-shape version.

Learning Following a Switch from a Rule-Based Environment to a Different Rule-Based Environment

Computational Modeling

Participants adopted rule-based learning in the third rule–based block. A rule-based model including a dynamic learning rate that was unequal across dimensions fitted the data best. Most importantly, participants mostly attended to the previously-relevant dimension (Fig. 6, Online Resource E).

The result in block 3, that participants attended to the dimension that was relevant in block 2, suggests participants experienced interference from the previously-learned rule. To further explore this, we split the data in block 3 into four subsets of 20 trials and inspected in each subset the fit of all computational models and parameter estimates of the best-fitting models (cf. Online Resource G). Results indicated that participants attended to the currently-relevant dimension after 20 to 40 trials.

Regression Analyses on Accuracy

In the third block, with respect to improvement across trials, accuracy results showed a main effect of trial ($z = -4.66$, $p < 0.001$), but no trial × version interaction ($p = 0.91$). This indicates that participants improved across

trials and that this improvement was similar across versions. With respect to the endpoint of learning, accuracy results showed no main effect of version ($p = 0.83$), indicating the endpoint of learning was similar across versions.

Comparing the third block to the second block, with respect to improvement across trials, accuracy results showed no block × trial interaction ($p = 0.08$), but a block × trial × version interaction ($z = 2.87$, $p = 0.004$). This indicates that, in general, trial effects were similar in blocks 3 and 2, but that the block × trial interaction differed between versions. Follow-up tests in each version showed a block × trial interaction in the pattern-to-shape version ($z = 2.98$, $p = 0.003$), but not in the shape-to-pattern version ($p = 0.38$). In the pattern-to-shape version, participants improved slower across trials in the third ($z = -4.62$, $p < 0.001$) compared to the second block ($z = -5.19$, $p < 0.001$). With respect to the endpoint of learning, accuracy results showed a main effect of block ($z = -2.02$, $p = 0.04$); moreover, they showed a block × version interaction ($z = -5.13$, $p < 0.001$). This indicates that, in general, the endpoint of learning was lower in the third compared to the second block and that this block effect differed between versions. Follow-up tests in each version showed a negative block effect (i.e., block 3 < block 2) in the pattern-to-shape version ($z = -4.56$, $p < 0.001$), but a positive block effect (i.e., block 3 > block 2) in the shape-to-pattern version ($z = 2.50$, $p = 0.01$). Note that this means higher final accuracy when "pattern" was the relevant dimension compared to when "shape" was the relevant dimension.

Together, these computational-modeling and accuracy results suggest participants learned the currently-relevant rule when the rule changed but that they experienced interference from the previously-learned rule at the beginning of learning. Based on the accuracy results, this interference seemed especially persistent for participants that switched from a pattern-rule to a shape-rule.

Learning Following a Switch from a Rule-Based Environment Back to an Object-Based Environment

Computational Modeling

Participants adopted object-based learning in the fourth object-based block. An object-based model including a dynamic learning rate that was unequal across pairs fitted the data best (Online Resource E).

Regression Analyses on Accuracy

In the fourth block, with respect to improvement across trials, accuracy results showed a main effect of trial ($z = -3.14$, $p = 0.002$), but no trial × version interaction ($p = 0.61$). This indicates that participants improved across trials and that

this improvement was similar across versions. With respect to the endpoint of learning, accuracy results showed no effect of version ($p = 0.47$), indicating the endpoint of learning was similar across versions. In an exploratory analysis, we tested whether participants performed above chance level in the final bin. This proved not to be the case ($M = 0.55$ [0.48; 0.62]; $t(35) = 1.47$, $p = 0.15$).

Comparing the fourth block to the third block, with respect to improvement across trials, accuracy results showed a block × trial interaction ($z = 3.28$, $p = 0.001$), but no block × trial × version interaction ($p = 0.49$). This indicates that participants improved slower across trials in the fourth compared to the third block and that this block × trial interaction was similar across versions. With respect to the endpoint of learning, accuracy results showed a main effect of block ($z = -8.10$, $p < 0.001$), but no block × version interaction ($p = 0.24$). This indicates the endpoint of learning was lower in block 4 compared to block 3 and that this block effect was similar across versions.

Together, these computational-modeling and accuracy results suggest that participants adopted object-based learning in an object-based environment at the end of the task. Although participants applied the optimal strategy, final accuracy was at chance level.

Discussion

In this study, we examined how people learn in an object-based environment, and how they adapt their behavior following a switch to a rule-based environment, following a switch to a different rule-based environment, and following a switch back to an object-based environment. To do so, we performed regression analyses and applied hierarchical Bayesian computational modeling to uncover whether participants used object-based or rule-based learning and to investigate learning parameters governing the best-fitting models. First, our results showed that people initially adopt rule-based learning despite its suboptimal nature in an object-based environment. Second, they showed that people learn rules after a switch to a rule-based environment. Third, they showed that people experience interference from previously-learned rules following a switch to a different rule-based environment. Fourth, they showed that people learn objects, although poorly, after a final switch to an object-based environment.

We argued that in our task with fixed pairs, people would be able to apply object-based learning in an object-based environment. Unexpectedly, our first main result showed that people adopted rule-based learning. Potentially, our task in which four pairs characterized by three dimensions needed to be learned was still too complex (Collins & Frank, 2012; Schaaf et al., 2019) to apply object-based learning in the

beginning of the task. Future studies in which less pairs need to be learned or less dimensions are used are thus advised to test this explanation. Note that the finding that people tend to search for rules in an object-based environment questions the general assumption in reinforcement-learning studies that people learn objects. Therefore, it may be beneficial if reinforcement-learning studies carefully select stimuli to minimize the tendency to search for rules and preferably use computational modeling to test whether rule-based learning is applied.

Our second main result, that people learned rules in a rule-based environment, extends previous reinforcement-learning findings (Ballard et al., 2018; Collins et al., 2014; Geana & Niv, 2014; Leong et al., 2017; Niv et al., 2015; Radulescu et al., 2016) by showing that people also learn rules when such a rule-based environment is preceded by an object-based environment.

Inspection of parameter estimates in best-fitting computational models allowed us to uncover our third main result, that people experience interference from previously-learned rules. This result is new in the reinforcement-leaning literature but in accordance with findings outside this literature, that is, on deterministic experience-based decision-making (Bröder & Schiffer, 2006; Hoffmann et al., 2019; Kämmer et al., 2013) and categorization (Best et al., 2013).

Our fourth main result, that people learned objects at the end of the task but that learning was minimal, suggests that learning in an object-based manner after learning in a rule-based environment is challenging. In the current design, it is difficult to disentangle different explanations for this result. It may be that, even in the final block, the task was too complex to adequately apply object-based learning or that participants experienced interference from the preceding rule-based block. Future studies adopting a mixed design could help disentangle these explanations, for example, by comparing behavior in a condition in which participants perform an object-based block followed by a rule-based block to a condition in which participants perform two consecutive object-based blocks.

Even though objects were predictive of a reward in the first as well as the fourth block, our main results suggest that participants employed different strategies in these blocks, that is, rule-based learning in the first block and object-based learning in the fourth block. Previous work in the reinforcement-learning literature (Farashahi et al., 2017), but also outside this literature (e.g., Johansen & Palmeri, 2002; Raijmakers et al., 2014), similarly showed that people tend to rely on rule-based strategies during early trials while they rely on object-based strategies during late trials. It may be that, in the first block, participants had too few trials to overcome their rule-based tendency and to apply object-based learning. To test this explanation, future studies are advised to administer more trials in the object-based blocks.

Next to the four main results, we found that people had a preference to rely on the pattern dimension. That is, (i) people incorrectly focused on "pattern" in the first, object-based, block, and (ii) learning the pattern-rule was easier than learning the shape-rule in the second, rule-based, block. What may have induced this saliency of the pattern dimension? Maybe it was due to the fact that the color and shape dimension were intertwined, meaning "color" and "shape" could not be observed independently, whereas this was not the case for the pattern dimension. To test this explanation, we performed an additional free-categorization experiment (cf. Online Resource J) in which we tested whether the pattern dimension was more salient (Schutte et al., 2017) than the other dimensions. We did not find evidence for this explanation, and therefore future studies are needed to replicate and understand this preference for the pattern dimension.

Three potential limitations can be identified. First, we modeled the data by either object-based or rule-based learning. By doing so, computational-modeling results showed that in the first object-based block, participants adopted rule-based learning. However, regression results showed an improvement across trials in this block, something that was not predicted by the best-fitting rule-based model (Fig. 3). One solution is to add object-based models including forgetting (Collins & Frank, 2012) as these models are better able to capture slight improvements across trials and might thus fit the data better. Another solution is to include hybrid models, combining object-based and rule-based learning (Niv et al., 2015), to further pinpoint the role of both approaches in multidimensional environments. It may be, for example, that participants start by applying rule-based learning but rely more on object-based learning as the block progresses (Farashahi et al., 2017).

Second, the computational models we considered, all assume a gradual learning process. However, in multidimensional environments, people might sequentially test whether a dimension is predictive of a reward or not (Choung et al., 2017; Radulescu et al., 2019; Wilson & Niv, 2012). Especially in real-world decision problems, the environment might be too complex to gradually learn the value of all features. Future studies on (mal)adaptive learning after switches between environments could investigate in which situations people adopt hypothesis-testing strategies and how application of these strategies is influenced by, for example, the dimensionality of the environment.

Third, because of identifiability problems, we only considered rule-based models in which learning was allowed to differ between dimensions. It could also be, however, that people weigh the dimensions differently when making a choice, e.g., choosing based on the most-informative dimension (Wilson & Niv, 2012; Wunderlich et al., 2011). To test whether interference in the third block is due to either differential learning or differential weighing in the beginning of this block, future studies could adopt different designs in order to test this distinction, for example, by including EEG measures to make a distinction between differential learning and differential weighing (Leong et al., 2017).

We were the first to assess adaptivity when switching between object-based and rule-based environments in a reinforcement-learning context. As such, we chose to assess how people learned in these environments on a group level. However, our within-subjects design lends itself for additional interesting comparisons. For example, do participants that tend to adopt rule-based learning in the first block also learn the rule quicker in the second block? And do these participants also tend to adopt rule-based learning in the fourth block? To answer these questions, we advise future studies to use mixture modeling in order to obtain individual learning models as opposed to learning models for the complete sample.

Taken together, these results obtained by computational modeling of behavior in a probabilistic reinforcement-learning task indicate that people tend to search for rules, even if they are not present. They also indicate that if rules are present, people are able to learn them, but are impaired when the relevant rule changes. Finally, they indicate that people find switching from learning rules to learning objects challenging. People thus have a hard time adjusting to switches between object-based and rule-based environments, although they learn to do so.

Declarations

Ethics Approval This research was approved by the Ethics Review Board of the Faculty of Behavioral and Social Sciences, University of Amsterdam (2019-DP-10313).

Consent to Participate Informed consent was obtained from all individual participants included in this study.

References

Balcarras, M., & Womelsdorf, T. (2016). A flexible mechanism of rule selection enables rapid feature-based reinforcement learning. *Frontiers in Neuroscience, 10*, 125. https://doi.org/10.3389/fnins.2016.00125

Ballard, I., Miller, E. M., Piantadosi, S. T., Goodman, N. D., & Mcclure, S. M. (2018). Beyond reward prediction errors: Human striatum updates rule values during learning. *Cerebral Cortex, 28*, 3965–3975. https://doi.org/10.1093/cercor/bhx259

Best, C. A., Yim, H., & Sloutsky, V. M. (2013). The cost of selective attention in category learning: Developmental differences between adults and infants. *Journal of Experimental Child Psychology, 116*(2), 105–119. https://doi.org/10.1016/j.jecp.2013.05.002

Bröder, A., & Schiffer, S. (2006). Adaptive flexibility and maladaptive routines in selecting fast and frugal decision strategies. *Journal of Experimental Psychology: Learning Memory and Cognition, 32*(4), 904–918.

Choung, O. H., Lee, S. W., & Jeong, Y. (2017). Exploring feature dimensions to learn a new policy in an uninformed reinforcement learning task. *Scientific Reports, 7*(1), 1–12. https://doi.org/10.1038/s41598-017-17687-2

Collins, A. G. E., & Frank, M. J. (2012). How much of reinforcement learning is working memory, not reinforcement learning? A behavioral, computational, and neurogenetic analysis. *European Journal of Neuroscience, 35*(7), 1024–1035. https://doi.org/10.1111/j.1460-9568.2011.07980.x

Collins, A. G. E., Cavanagh, J. F., & Frank, M. J. (2014). Human EEG uncovers latent generalizable rule structure during learning. *The Journal of Neuroscience, 34*(13), 4677–4685. https://doi.org/10.1523/JNEUROSCI.3900-13.2014

Farashahi, S., Rowe, K., Aslami, Z., Lee, D., & Soltani, A. (2017). Feature-based learning improves adaptability without compromising precision. *Nature Communications, 8*(1), 1768. https://doi.org/10.1038/s41467-017-01874-w

Farashahi, S., Xu, J., Wu, S. W., & Soltani, A. (2020). Learning arbitrary stimulus-reward associations for naturalistic stimuli involves transition from learning about features to learning about objects. *Cognition, 205*(September), 104425. https://doi.org/10.1016/j.cognition.2020.104425

Geana, A., & Niv, Y. (2014). Causal model comparison shows that human representation learning is not Bayesian. *Cold Spring Harbor Symposia on Quantitative Biology, 79*, 161–168. https://doi.org/10.1101/sqb.2014.79.024851

Gigerenzer, G., Todd, P. M., Group, T. A. R. (1999). *Simple heuristics that make us smart*. Oxford University Press.

Hoffmann, J. A., von Helversen, B., & Rieskamp, J. (2019). Testing learning mechanisms of rule-based judgment. *Decision, 6*(4), 305–334.

Johansen, M. K., & Palmeri, T. J. (2002). Are there representational shifts during category learning? *Cognitive Psychology, 45*(4), 482–553. https://doi.org/10.1016/S0010-0285(02)00505-4

Kämmer, J. E., Gaissmaier, W., & Czienskowski, U. (2013). The environment matters: Comparing individuals and dyads in their adaptive use of decision strategies. *Judgment and Decision Making, 8*(3), 299–329.

Leong, Y. C., Radulescu, A., Daniel, R., DeWoskin, V., & Niv, Y. (2017). Dynamic interaction between reinforcement learning and attention in multidimensional environments. *Neuron, 93*(2), 451–463. https://doi.org/10.1016/J.NEURON.2016.12.040

Marković, D., Gläscher, J., Bossaerts, P., O'Doherty, J., & Kiebel, S. J. (2015). Modeling the evolution of beliefs using an attentional focus mechanism. *PLoS Computational Biology, 11*(10), 1–34. https://doi.org/10.1371/journal.pcbi.1004558

Niv, Y., Daniel, R., Geana, A., Gershman, S. J., Leong, Y. C., Radulescu, A., & Wilson, R. C. (2015). Reinforcement learning in multidimensional environments relies on attention mechanisms. *Journal of Neuroscience, 35*(21), 8145–8157. https://doi.org/10.1523/JNEUROSCI.2978-14.2015

Radulescu, A., Daniel, R., & Niv, Y. (2016). The effects of aging on the interaction between reinforcement learning and attention. *Psychology and Aging*. https://doi.org/10.1037/pag0000112.supp

Radulescu, A., Niv, Y., & Daw, N. D. (2019). A particle filtering account of selective attention during learning. *2019 Conference on Cognitive Computational Neuroscience*.

Raijmakers, M. E. J., Schmittmann, V. D., & Visser, I. (2014). Costs and benefits of automatization in category learning of ill-defined rules. *Cognitive Psychology, 69*, 1–24. https://doi.org/10.1016/j.cogpsych.2013.12.002

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning II: Current research and theory* (pp. 64–99). Appleton-Century-Crofts.

Schaaf, J. V., Jepma, M., Visser, I., & Huizenga, H. M. (2019). A hierarchical Bayesian approach to assess learning and guessing strategies in reinforcement learning. *Journal of Mathematical Psychology, 93*, 102276. https://doi.org/10.1016/j.jmp.2019.102276

Schutte, I., Slagter, H. A., Collins, A. G. E., Frank, M. J., & Kenemans, J. L. (2017). Stimulus discriminability may bias value-based probabilistic learning. *PLoS ONE, 12*(5), e0176205. https://doi.org/10.1371/journal.pone.0176205

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society. Series B: Statistical Methodology, 64*(4), 583–616. https://doi.org/10.1111/1467-9868.00353

Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction*. MIT Press.

Wagenmakers, E. J., & Farrell, S. (2004). AIC model selection using Akaike weights. *Psychonomic Bulletin and Review, 11*(1), 192–196. https://doi.org/10.3758/BF03206482

Wilson, R. C., & Niv, Y. (2012). Inferring relevance in a changing world. *Frontiers in Human Neuroscience, 5*, 1–14. https://doi.org/10.3389/fnhum.2011.00189

Wunderlich, K., Beierholm, U. R., Bossaerts, P., & O'Doherty, J. P. (2011). The human prefrontal cortex mediates integration of potential causes behind observed outcomes. *Journal of Neurophysiology, 106*(3), 1558–1569. https://doi.org/10.1152/jn.01051.2010