frontiers in Behavioral Neuroscience

ORIGINAL RESEARCH
published: 20 December 2016
doi: 10.3389/fnbeh.2016.00234

Slips of Action and Sequential Decisions: A Cross-Validation Study of Tasks Assessing Habitual and Goal-Directed Action Control

Zsuzsika Sjoerds[1], Anja Dietrich[1], Lorenz Deserno[1,2,3], Sanne de Wit[4,5], Arno Villringer[1,6], Hans-Jochen Heinze[3,7], Florian Schlagenhauf[1,2†] and Annette Horstmann[1,8†]*

[1]*Department of Neurology, Max-Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany,* [2]*Department of Psychiatry and Psychotherapy, Campus Charité Mitte, Charité – Universitätsmedizin Berlin, Berlin, Germany,* [3]*Department of Neurology, Otto-von-Guericke University, Magdeburg, Germany,* [4]*Department of Clinical Psychology, University of Amsterdam, Amsterdam, Netherlands,* [5]*Amsterdam Brain and Cognition, University of Amsterdam, Amsterdam, Netherlands,* [6]*Mind and Brain Institute, Charité and Humboldt University, Berlin, Germany,* [7]*Department of Behavioral Neurology, Leibniz Institute for Neurobiology, Magdeburg, Germany,* [8]*Integrated Research and Treatment Center Adiposity Diseases, Leipzig University Medical Center, Leipzig, Germany*

Instrumental learning and decision-making rely on two parallel systems: a goal-directed and a habitual system. In the past decade, several paradigms have been developed to study these systems in animals and humans by means of e.g., overtraining, devaluation procedures and sequential decision-making. These different paradigms are thought to measure the same constructs, but cross-validation has rarely been investigated. In this study we compared two widely used paradigms that assess aspects of goal-directed and habitual behavior. We correlated parameters from a two-step sequential decision-making task that assesses model-based (MB) and model-free (MF) learning with a slips-of-action paradigm that assesses the ability to suppress cue-triggered, learnt responses when the outcome has been devalued and is therefore no longer desirable. MB control during the two-step task showed a very moderately positive correlation with goal-directed devaluation sensitivity, whereas MF control did not show any associations. Interestingly, parameter estimates of MB and goal-directed behavior in the two tasks were positively correlated with higher-order cognitive measures (e.g., visual short-term memory). These cognitive measures seemed to (at least partly) mediate the association between MB control during sequential decision-making and goal-directed behavior after instructed devaluation. This study provides moderate support for a common framework to describe the propensity towards goal-directed behavior as measured with two frequently used tasks. However, we have to caution that the amount of shared variance between the goal-directed and MB system in both tasks was rather low, suggesting that each task does also pick up distinct aspects of goal-directed behavior. Further investigation of

OPEN ACCESS

Edited by:
Lars Schwabe,
University of Hamburg, Germany

Reviewed by:
Tom Smeets,
Maastricht University, Netherlands
Marc Exton-McGuinness,
University of Birmingham, UK

****Correspondence:***
Zsuzsika Sjoerds
sjoerds.zs@gmail.com

[†]These authors have contributed equally to this work.

Received: *10 October 2016*
Accepted: *28 November 2016*
Published: *20 December 2016*

Citation:
Sjoerds Z, Dietrich A, Deserno L, de Wit S, Villringer A, Heinze H-J, Schlagenhauf F and Horstmann A (2016) Slips of Action and Sequential Decisions: A Cross-Validation Study of Tasks Assessing Habitual and Goal-Directed Action Control. Front. Behav. Neurosci. 10:234. doi: 10.3389/fnbeh.2016.00234

Abbreviations: ANOVA, Analysis of Variance; DSI, Devaluation Sensitivity Index; DSST, Digit Symbol Substitution Test; EEG, Electro Encephalography; IQR, Interquartile Range; ITI, Inter-trial Interval; MB, Model-based; MF, Model-free; ms, Milliseconds; O, Outcome; Q, Choice Values; R, Response; RL, Reinforcement Learning; S, Stimulus; SARSA, State-Act-Reward-State-Act; SD, Standard Deviation; TD, Temporal Difference; TMT, Trail Making Test; VPA, Visual Paired Association Test; WMT, Wiener Matrizen Test.

the commonalities and differences between the MF and habit systems as measured with these, and other, tasks is needed. Also, a follow-up cross-validation on the neural systems driving these constructs across different paradigms would promote the definition and operationalization of measures of instrumental learning and decision-making in humans.

Keywords: goal-directed, habit, model-based, model-free, cross-validation, sequential decision making, slips-of-action, reinforcement learning

INTRODUCTION

Instrumental decision-making requires learning and executing adequate behavior efficiently in relevant situations in order to obtain desired outcomes. Based on an extensive body of animal research, this ability is thought to rely on the functioning of two parallel systems: a reflexive habitual system and a deliberate goal-directed system (Dickinson, 1985; Balleine and Dickinson, 1998a). The habit system is believed to be an evolutionary basal system, suggested to mainly rely on dorsolateral striatal areas (Yin et al., 2004, 2006; Tricomi et al., 2009; Wunderlich et al., 2012a). Habits are "stamped in" by past reinforcements until they are performed in an automatic routine. The habit system is inflexible and suboptimal in changing environments, but it offers the advantage to free up cognitive resources, allowing the allocation of attention to parallel tasks. In contrast, goal-directed behavior has shown to largely involve prefrontal cortical and dorsomedial striatal brain areas (Corbit and Balleine, 2003; Killcross and Coutureau, 2003; Yin et al., 2005a; Valentin et al., 2007; de Wit et al., 2009; but see Jonkman et al., 2009), and is characterized by flexible behavior, which is more easily adaptable in the face of changing contingencies. However, it is thought to be computationally more demanding than the habit system, and the ability to engage the goal-directed system effectively seems to depend on trait factors such as healthy aging (Eppinger et al., 2013; de Wit et al., 2014) and cognitive capacities (Otto et al., 2013; Smittenaar et al., 2013; Schad et al., 2014) or state conditions, including stress (Schwabe and Wolf, 2009; Otto et al., 2013; Radenbach et al., 2015). There is growing evidence that deficient instrumental decision-making based on the dual-systems theory is implicated in multiple disorders (e.g., Gillan et al., 2011; Sjoerds et al., 2013; Horstmann et al., 2015; Voon et al., 2015b; McKim et al., 2016; Reiter et al., 2016). Human behavior, however, might be influenced by a wide variety of unrelated external or internal factors (i.e., social conventions, cultural context or financial situations), rendering it more noisy than e.g., rodent behavior despite the large overlap between human and rodent instrumental systems (Balleine and O'Doherty, 2010). This increases the complexity to measure individual constructs in humans. Together with the need to apply cognitive measurements in pathological samples, this advances the prerequisite to optimize the assessment of these instrumental behaviors.

To adequately assess the degree to which the two proposed systems are used in instrumental choices, it is essential to ensure suitable instruments that objectively assess covert sub-processes contributing to the constructs and that are simultaneously straightforward for intuitive analyses and application in patient samples (Huys et al., 2016). Throughout the past decade, distinct paradigms to study the two systems in humans have been operationalized based on different methodological and historical perspectives of habitual vs. goal-directed behavior (Doll et al., 2012; Dolan and Dayan, 2013). They can be distinguished by whether the paradigm captures the ongoing contingency updating process or largely established behavioral schemata, processes that are not necessarily independent of each other (Gillan et al., 2015). Further variations lie in the focus on central sub-processes underlying habitual and goal-directed choices. For example, goal-directed behavior is complex and involves multiple sub-processes including forward planning, outcome contingency weighting, search processes and abstract inference (Hampton et al., 2006; Abe and Lee, 2011; Daw et al., 2011; Doll et al., 2012). Change in outcome value or contingencies (e.g., outcome devaluation or changes in outcome probabilities) provides the canonical assay of behavioral flexibility as related to the balance between goal-directed vs. habitual control. Frequently used tasks, such as a slips-of-action paradigm (de Wit et al., 2012b) and sequential decision-making paradigm (Daw et al., 2011) assess the ability to rapidly adjust behavior to changes in outcome value.

Classically, the relative involvement of the goal-directed and habitual systems in instrumental choices has been studied in animals by (selective) outcome devaluation procedures (Adams and Dickinson, 1981; Balleine and Dickinson, 1998b), a method that has been adapted to human research (Valentin et al., 2007; Tricomi et al., 2009; Horstmann et al., 2015). Devaluation of an outcome (O) that has been associated with a stimulus (S) will change response (R) behavior when under control of the goal-directed system, as R-O contingencies are represented in the goal-directed system. However, once S-R habitual responding to a stimulus is established, the outcome is no longer taken into account in the choice behavior; therefore, devaluation of the outcome will not immediately influence habitual responding to the stimulus but only after gradual update of the stimulus value after repeated outcome feedback. Following the same line of reasoning, a slips-of-action paradigm was developed (de Wit et al., 2012b), in which participants learn stimulus-reward contingencies. After training, an instructed devaluation phase assesses whether participants can suppress previously learned responses that yield no-longer-valuable outcomes, while continuing to respond for still-valuable outcomes. A failure to do so, as reflected in "slips of action" towards devalued

outcomes, is interpreted as relative reliance on S-R habitual—as opposed to goal-directed—control. Crucially, a devaluation sensitivity index (DSI) is calculated based on the difference in responding between these two trial types, providing a single parameter that represents the relative involvement of the habit vs. goal-directed system in action control. This task has been used extensively to study goal-directed and habitual action control in healthy participants (de Wit et al., 2009, 2012b) following dopamine (de Wit et al., 2012a) and serotonin level reductions (Worbe et al., 2015) and in patient samples, including obsessive-compulsive disorder (Gillan et al., 2011), alcohol dependence (Sjoerds et al., 2013), Gille de La Tourette syndrome (Delorme et al., 2016), Parkinson's Disease (de Wit et al., 2011; O'Callaghan et al., 2013) and autism spectrum disorders (Geurts and de Wit, 2014). It remains unclear, however, how devaluation sensitivity, as assessed with this task, relates to other paradigms assessing goal-directed and habitual control, such as the model-based (MB) and model-free (MF) reinforcement learning (RL) algorithms used during sequential decision making.

RL theory aims to formalize decision-making processes such as goal-directed and habitual learning by describing distinct underlying computational mechanisms. To this end, in addition to analyzing observable behavior such as accuracy, reaction times and win-stay probabilities, generative models are implemented to infer parameters that underlie the observed behavior. One of the frequently used RL models follows the temporal difference (TD) theory (Sutton and Barto, 1998), which is closely linked to habitual learning. It provides a "MF" update rule to learn action values based on past reinforcements. Goal-directed instrumental learning has also been proposed to have a formal counterpart in RL, in a family of algorithms known as "MB" RL (Daw et al., 2005; Rangel et al., 2008; Redish et al., 2008). The MB system uses a model of the environment for flexible forward planning. Resembling the goal-directed system, it contains knowledge on the causal relationship between actions and outcomes. In the context of RL theory, one task to study goal-directed vs. habitual responding with the MB and MF algortihms, respectively is the two-step sequential decision making task (Daw et al., 2011) in combination with computational modeling of decision making using MB and MF algorithms. This two-step task has been extensively used in the past years to study MB vs. MF learning in healthy and diseased samples (Daw et al., 2011; Wunderlich et al., 2012b; Eppinger et al., 2013; Otto et al., 2013; Deserno et al., 2015a,b; Gillan et al., 2015; Radenbach et al., 2015; Voon et al., 2015b; Morris et al., 2016; Reiter et al., 2016; Worbe et al., 2016).

The increasing availability of instruments measuring goal-directed and habitual behavior increases the necessity for cross-validation of different paradigms on the assessment of the two central constructs. Recently, Friedel et al. (2014) have performed a valuable cross-validation study on the goal-directed and habit constructs assessed by a selective devaluation task (Valentin et al., 2007) and the two-step sequential decision-making task (Daw et al., 2011). They found specific cross-correlation between MB choices during sequential decisions and goal-directed behavior after devaluation. This suggests a single framework underlying both task measures, speaking in favor of construct validity of both measurement approaches. However, further comparable research on cross-validation of instrumental decision-making between other tasks is needed. Another recent study directly manipulated MB learning with habitual responding within one paradigm: they used an adjusted two-step sequential decision-making task, including a later phase that provided a DSI (Gillan et al., 2015). By using a median split on this DSI, they defined groups of participants using predominantly goal-directed or habitual responding. They found that MB control during the first phase of the task protected from established habitual responding during the last phase measured by devaluation sensitivity. This further indicates an overlap between MB learning and established goal-directed behavior.

We would like to extend this line of research by cross-validating MB control and goal-directed responding between two different tasks that have been most commonly used in the recent body of literature. We will correlate parameters describing MB and MF control from the two-step sequential decision-making task (Daw et al., 2011) with a DSI from the slips-of-action paradigm (de Wit et al., 2012b; Worbe et al., 2015), which measures the relative balance of goal-directed and habitual choices on a gradual scale. We hypothesize a positive association between the measure of MB behavior and the DSI. In other words, participants who show more MB behavior in the two-step task are expected to be better able to respond selectively for still-valuable outcomes, while suppressing slips of action towards no-longer valuable outcomes in the slips-of-action paradigm. We additionally explore a possible association between the tasks on the habit system, expecting a negative correlation between MF behavior and the DSI. We furthermore assess the role of higher-order cognitive capacities, measured by widely used neuropsychological tests, in the recruitment of the goal-directed/MB and habitual/MF systems.

MATERIALS AND METHODS

Participants

A total of 28 healthy participants (12 females, mean age: 27, see **Table 1**) performed both paradigms. Based on the previous cross-validation study by Friedel et al. (2014), showing effects sizes between 0.5 and 0.7, an a-priori power analysis (G*Power version 3.1.9.2) showed that for the current study a sufficient sample size would lie between $N = 13$ and $N = 34$. Volunteers

TABLE 1 | Sample descriptives.

Descriptive	Mean	SD	Range
Age	27.04	3.415	21–34
Gender (female): *N*, %	*N* = 12	42.90%	
Beck depression inventory (BDI)	3.68	3.422	0–14
Wiener matrizen test (WMT)			
Score	18.96	3.854	11–24
IQ	120.38	12.249	95–136.5
Visual association test (VAT)	12.11	3.891	3–18
Digit symbol substitution test (DSST)	87.05	10.741	60–110

were highly educated non-smokers without indication for major depression as measured with the Beck's Depression Inventory (BDI), cut-off value 18 (Beck et al., 1996). Non-verbal intelligence was assessed with the Wiener Matrizen Test (WMT, The Viennese Matrices Test; Formann and Piswanger, 1979). Visual short-term memory was tested with a Visual Association Test (VAT), a computerized version of the Visual Paired Association (VPA) Test, part of the Wechsler Memory Scale (Wechsler, 1987, 2006). In this test participants have to memorize the combination of shape and color of six different stimulus pairs. Cognitive speed was assessed with the digit-symbol-substitution test (DSST; Wechsler, 1955). These cognitive measures were included to examine their potential relation to performance on each task.

Participants were recruited from the database at the Max Planck Institute for Human Cognitive and Brain Sciences in Leipzig, Germany. All participants were financially compensated for participation with €7-per hour in addition to a monetary reward acquired during the experimental tasks. The study was approved by the Ethics Committee of the University of Leipzig, Germany, and conducted in accordance with the Declaration of Helsinki. Written informed consent was obtained from all participants prior to the study.

Paradigms

Three-Phase Instrumental Learning Task

A simplified version of the slips-of-action paradigm, an instrumental learning task developed by de Wit et al. (2012b) was used, which has been successfully applied in previous studies (e.g., Worbe et al., 2015; Delorme et al., 2016). In the current study, pictures of animals instead of fruit pictures were used. The task consists of three phases, a discrimination training phase to learn S-R-O associations and an outcome devaluation phase and slips-of-action phase to test for the strength of learned S-R-O associations (see **Figure 1**). The slips-of-action phase provides a DSI (for detailed explanation, see below), which encompasses a "balance" measure of relative goal-directed and habitual control. We do report results on the other phases of the task, as is done in all previous studies using the same task (e.g., de Wit et al., 2011, 2014; Geurts and de Wit, 2014; Worbe et al., 2015; Delorme et al., 2016). However, the current study solely aims to assess the parallels between different tasks in measuring relative involvement of goal-directed/MB and habitual/MF control. Therefore the DSI of this task is of main interest for correlational analyses to the current study.

Discrimination training phase

During the first phase, the discrimination training phase, participants learned by trial-and-error to respond (R) with a left or right button press to stimuli (S) in order to gain outcomes (O) that are worth points representing monetary reward. Participants were instructed to earn as many points as possible. A trial started with a box displayed in the middle of the screen, with a picture of an animal printed on the front side. Participants were instructed that the box could be opened with either a left or right button press, but that only one of the two buttons is the correct one, rendering another animal plus a monetary reward in the opened box. When pressing the incorrect button, the box would open, but it would be left empty, without a monetary reward (zero points won). Six different possible stimuli, displayed in a randomized order over the trials, would lead deterministically (i.e., with a 100% contingency) to six different outcomes in the case of a correct response. For three stimuli a right button press would lead to the outcome and a monetary reward, whereas for the other three a left button press would be the correct one to obtain an outcome plus monetary reward. This phase comprised eight blocks and a total of 96 trials. Dividing the task into blocks with randomized stimulus order within each block aided in measuring a learning effect across blocks, and ensured that participants learned all stimuli evenly divided throughout the experiment, instead of randomly seeing only a high amount of repetitions of one stimulus e.g., at the end of the training. Each stimulus was displayed 16 times in order for all participants to adequately learn the S-R-O associations.

Outcome-devaluation test phase

Following the discrimination training phase, an outcome-devaluation test phase assessed the strength of goal-directed R-O associations learned during the training phase. Here, outcomes (again, open boxes with animal icons) were displayed in pairs; one that was previously associated with a left response and one with a right response. In each trial, one of the outcomes was devalued (i.e., it would no longer produce a monetary reward), indicated by a red cross superimposed on the devalued outcome. Now, participants had to use their knowledge of the R-O relationships to (re)direct their choices towards the still-valuable outcome, by pressing the button that had lead to this outcome during the discrimination training phase. This phase was comprised of 36 trials. Participants were not directly given feedback on each trial, but instead were instructed that correct button presses would still earn them points and that they would be shown their total score at the end of the test phase.

During the discrimination training phase and the outcome-devaluation test phase, we assessed the total percentage correct responses (task accuracy). Due to non-normal distribution of all outcome measures (including accuracy in the learning phase), non-parametric testing was used. The Friedman test was applied to test for performance differences across the eight equal blocks of the training-phase, to check for instrumental learning effects over time.

Slips-of-action phase

During the slips-of-action phase, the balance between goal-directed and habitual learning systems was directly assessed and hence, this phase is of main interest for cross-validation with the sequential decision-making task (see below).

This phase was comprised of nine blocks, with a total of 108 trials. At the beginning of each block an instruction screen with six possible outcomes (open boxes with animal icons inside) was shown for 5 s, two of them superimposed with a cross. The cross indicated devaluation of those outcomes, and that responding to the stimulus associated with those outcomes

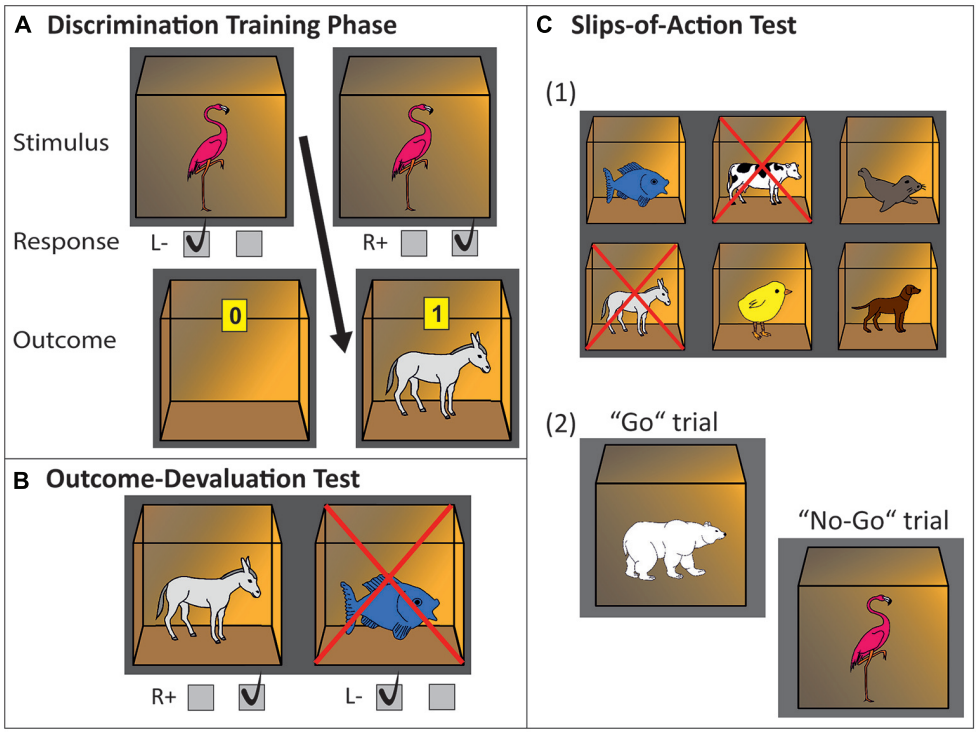

FIGURE 1 | Experimental paradigm, three-phase instrumental learning task. (A) Discrimination training phase. In this example, a flamingo stimulus printed on the front of a closed box indicates that pressing the right key will open the box and will be rewarded with a donkey and points inside of the box. Pressing the left key will not be rewarded (empty open box is revealed). **(B)** Outcome-devaluation test. In this example, two open boxes are presented with a donkey and fish inside. The cross superimposed on the fish signals this outcome is no longer worth any points. The accurate response in this example would be pressing the right key (which yielded the still-valuable donkey outcome during the learning phase). **(C)** Slips-of-action test. **(1)** Participants are first presented with the six outcomes. In this example, donkey and cow are superimposed with a cross, indicating that the response leading to these outcomes will now result in subtraction of points (devaluation). The other animal outcomes are still valuable. **(2)** Afterwards, in rapid succession animal stimuli are presented on the outside of the boxes. Participants are instructed to press the correct key if a stimulus indicates the availability of a still-valuable outcome inside the box ("Go", example: polar bear stimulus signaling fish outcome), but withhold responding if the outcome inside the box has been devalued ("No-Go", example: flamingo stimulus signaling donkey outcome).

would consequentially no longer earn points. After this screen, stimulus pictures were shown in rapid succession. Participants had to respond as fast as possible with a correct button-press to stimuli (closed boxes with an animal icon printed on the front) associated with still-valuable outcomes, and withhold their response for stimuli associated with devalued outcomes. Each stimulus remained on the screen for a fixed 1000 ms, during which the participant had to respond or withhold their response, respectively. The next trial started after an inter-trial interval (ITI) of 1000 ms. As in the outcome-devaluation phase, also in this phase no direct feedback was given, in order to prevent new learning. Instead, the total amount of points was shown at the end of the phase. During each block, each of the six stimuli was shown twice in semi-random order, with the exception that stimuli were never directly repeated. Throughout the nine blocks, each outcome was devalued three times, resulting in 36 trials where the outcome was devalued, and 72 trials with still valuable outcomes.

In this phase, response tendencies through direct S–R associations (related to the habit system) should lead to commission errors on trials showing stimuli associated with the devalued outcomes. Contrarily, successful selective inhibition on the basis of outcome value should be suggestive of dominant goal-directed control through more complex S-R-O associations, which is mediated by anticipation and evaluation of the consequent outcome (see e.g., Gillan et al., 2011; de Wit et al., 2012a,b). We calculated the DSI for the slips-of-action phase by subtracting percentages of responses made toward devalued outcomes from percentages of responses made toward still valuable outcomes, according to the following formula: ((*N* valued responses/*N* total responses) − (*N* devalued responses/*N* total responses)). Following the explanation above, this DSI during the slips-of-action phase is a "balance" measure of relative goal-directed and habitual control, and hence of main interest for the cross validation with the sequential decision-making task.

Baseline test phase

As a control test for general inhibitory impairments, participants also performed a baseline test of inhibitory control. This test closely resembled the slips-of-action phase, except that the decision to respond or withhold could be based directly on stimulus identity as opposed to outcome anticipation. To this end, at the beginning of each block a screen with the six possible

stimuli (closed boxes with animal printed on the outside) was shown, two of them superimposed with a cross. This time participants simply had to withhold their responses for the stimuli that had been superimposed with a cross ("stimulus devaluation"). Again, no direct feedback was given, but they were shown the total amount of earned points at the end of the phase. Importantly, the baseline test controls for outcome-based responding, as it is independent of the outcomes. However, it does not control for S-R associated behavior, thus the test might also be driven by strong S-R associations, which could be indicative of habitual behavior. The order of slips-of-action and baseline test phase was counterbalanced across participants.

Two-Step Sequential Decision-Making Task

A two-step Markov sequential decision-making task (Daw et al., 2011; see **Figure 2**) was used to assess the degree of MB and MF behavioral control. The applied version was identical to previous work from our group (Friedel et al., 2014; Sebold et al., 2014; Deserno et al., 2015a,b; Reiter et al., 2016). The task consisted of 201 two-stage trials. Within each trial, participants made two (stage 1, stage 2) sequential choices out of two stimuli to finally receive a monetary reward after the second stage (**Figure 2A**). At the first stage, participants selected within 2 s one of two stimuli displayed in gray boxes. Responses slower than 2 s were invalid. The chosen stimulus moved to the top middle of the screen and remained displayed during the second stage, while the non-chosen stimulus faded after the choice was made. At the second stage, participants again chose between two stimuli in differently colored pairs of boxes. The position of the stimuli on the left or right side of the screen in both stages was randomized and participants were explicitly instructed that the position of the stimuli was not relevant. The second-stage choice could either be rewarded with 20 eurocents (displaying a coin), or was not rewarded (displaying a zero). Feedback (reward or no reward) after the second choice was delivered in a probabilistic manner following slowly changing Gaussian random walks (see Daw et al., 2011). Participants trained for 55 trials before performing the actual task and were explicitly introduced to the task-structure, similar to Daw et al. (2011). They were informed that they would get financial reimbursement after the task with an amount depending on the total reward received during the task.

Stay probabilities

Crucial to this task is that presentation of second-stage pairs depends probabilistically on first-stage choices: each of the first-stage choices was predominantly associated with one of the two second-stage stimulus pairs (70% → common state) and less with the other (30% → rare state; see **Figure 2B**). These state transition probabilities were fixed during the experiment. A MF agent disregards these transition probabilities and stays

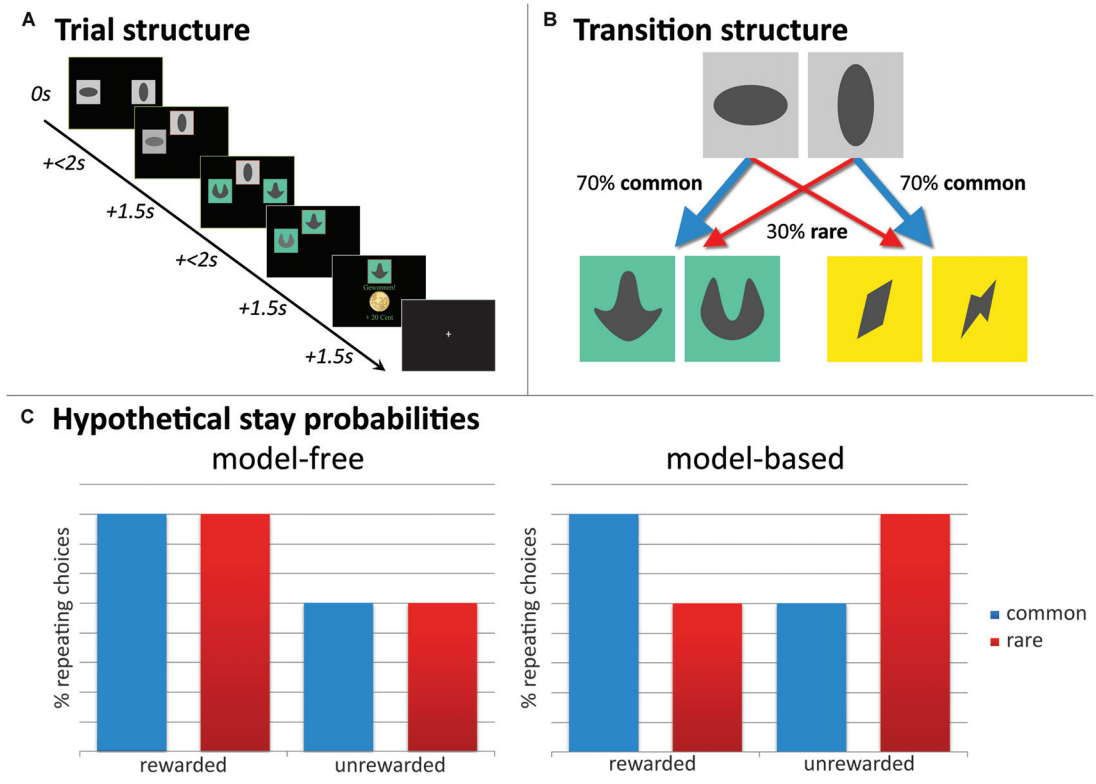

FIGURE 2 | Experimental paradigm, two-step sequential decision-making task. (A) Example of a trial-sequence with timing; **(B)** state-transition probabilities indicating common and rare transitions; **(C)** hypothetical full model-free (MF) and full model-based (MB) choice strategies would result in these choice patterns. Depicted here, stay probability plots for first step choices as a function of reward (reward vs. no reward) and state (common vs. rare). A main effect of reward guides MF choice strategies, whereas MB choice strategies show a reward * state interaction.

with the first-stage actions that have led to a reward after a second-stage choice. This indicates a main effect of reward on stay probabilities at the first stage: the probability that the same action will be repeated in the subsequent trial. Contrary, a MB agent does take into account these transition probabilities and accordingly, contains a "model" of the task. In other words, the MB system increases the chance to switch at the first-stage after a reward was delivered following a "rare" transition, but increases stay behavior at the first-stage after receiving no reward after a "rare" transition. This indicates a reward-by-state interaction effect on stay probabilities. See **Figure 2C** for a hypothetical representation of stay probabilities for a pure MF and pure MB learner, respectively. For task descriptive analyses, individual stay probabilities, as stay-switch behavior was defined as a function of reward (reward vs. no reward) and state (common vs. rare), are subjected to a repeated-measures analysis of variance (ANOVA) with reward and state as within-subjects factors.

Computational modeling

We used computational modeling in the analyses of choice behavior, to deduce covert control strategies in solving the task, based on the MF or MB system. In line with Daw et al. (2011), we used RL models that learn choice values (Q) through prediction errors. To this end we distinguish the three pairs of stimuli in the two stages (first stage: S_A, second stage: S_B, S_C), which are followed by an action a. First, trial-by-trial MF (Q_{MF}) stimulus values were calculated with a State-Act-Reward-State-Act (SARSA) (λ) model as follows:

$$Q_{MF}(S_{i,t+1}, a_{i,t+1}) = Q_{MF}(S_{i,t}, a_{i,t}) + \alpha_i \delta_{i,t} \quad (1)$$

Here i denotes the stage (first-stage: i = 1; second stage: i = 2), and t denotes the trial. In equation 1, δ refers to the trial-by-trial prediction error used to update the stimulus value, weighted by learning rate α. The prediction error is computed as the difference between expected value and obtained reward (r):

$$\delta_{i,t} = r_{i,t} + Q_{MF}(S_{i+1,t}, a_{i+1,t}) - Q_{MF}(S_{i,t}, a_{i,t}) \quad (2)$$

Note that $r_{1,t} = 0$ because no reward is delivered after a first-stage choice, and Q_{MF} ($S_{3,t}$, $\alpha_{3,t}$) = 0 because the task only has two states. First-stage values are additionally updated by a stage-skipping parameter λ, which connects the two stages and allows the reward prediction error at the second stage to modulate first-stage values:

$$Q_{MF}(S_{1,t+1}, a_{1,t+1}) = Q_{MF}(S_{1,t}, a_{1,t}) + \alpha_1 \lambda \delta_{2,t} \quad (3)$$

Next, the MB algorithm learns values by forward planning, and computes first-stage values by merely multiplying the better option at the second stage with the transition probabilities:

$$Q_{MB}(S_A, a_j) = \begin{array}{c} P(S_B|S_A, a_j) max Q_{MF}(S_B, a) \\ + P(S_C|S_A, a_j) max Q_{MF}(S_C, a) \end{array} \quad (4)$$

This simplified approach to MB control is justified because participants are extensively trained on the transition probabilities (also shown in Daw et al., 2011). Finally, these MF and MB decision-values are connected in a hybrid algorithm:

$$Q(S_A, a_j) = \omega Q_{MB}(S_A, a_j) + (1 - \omega) Q_{MF}(S_A, a_j) \quad (5)$$

In this equation ω is a free weighting parameter, which connects the MB and MF values. Therefore, ω represents the relative influence of the MB and MF system that is, other than two separate parameters describing MB and MF choices (see below), a parameter of interest in the correlation with the goal-directed/habit parameters from the slips of action task.

Finally, to connect the calculated values to choices, we used an observation model following the softmax choice rule. This softmax observation model transforms the obtained values into choice probabilities with three parameters: the free inverse temperature parameter (β_i) shows deterministic choices and is allowed to differ between the two stages (β_1 and β_2) and a repetition parameter (ρ) accounting for perseveration of first-stage choices:

$$p\left(a_{i,t} = a|S_{i,t}\right) = \frac{\exp\left(\beta_i \left[Q(S_{i,t}, a) + \rho * rep(a)\right]\right)}{\sum_{a'} \exp\left(\beta_i \left[Q(S_{i,t}, a') + \rho * rep(a')\right]\right)} \quad (6)$$

To connect the choices to the values of the MB and MF system individually, we calculated separate free inverse temperatures for the two systems (β_{MB} and β_{MF}) that specify the degree to which action choices follow from the MB and MF action values respectively. To this end we multiplied the first-stage stochasticity parameter β with ω: $\beta_{MB} = \omega * \beta$ and $\beta_{MF} = (1 - \omega) * \beta$ (see Otto et al., 2013). These two parameters facilitate examination of individual differences in the influence of either the MB or MF system and are therefore used in the correlation analyses with the slips-of-action task.

Bounded parameters were fitted by transformation to a logistic (α, λ, ω) or exponential (β) distribution in order to obtain normally distributed parameter estimates. To infer the maximum-a-posteriori estimate of each parameter for each subject, we set the prior distribution to the maximum-likelihood estimates given the data of all participants, and subsequently used Expectation-Maximization (Huys et al., 2011, 2012).

Correlation Between the Two Paradigms

We were interested whether MB and MF updating was associated with goal-directed/habitual choices. Therefore, the β_{MB} and β_{MF} parameters from the two-step task, describing MB and MF choice behavior respectively, were correlated with the DSI of the slips-of-action phase. The DSI parameter indicates the balance of goal-directed and habitual behavior, and was computed by calculating the difference between percentages of responses made toward valuable outcomes minus percentages of responses made toward devalued outcomes. We expected a positive association between β_{MB} and the DSI, and a negative association between β_{MF} and the DSI. If a strong association of both β_{MB} and β_{MF} with DSI could be found, this could reflect a positive association between the DSI and the balance score of the two-step task, the weighting parameter ω, computed using the modeling approach as described above. Therefore, we also performed a confirmatory

correlation analysis between these "balance" parameters of the two tasks.

In the slips-of-action task, aside from the balance parameter DSI, no separate parameters describe the two individual instrumental systems separately. However, the DSI is calculated based on percentage responses to still valuable outcomes and the percentage slips of actions (i.e., responses for devalued outcomes). A higher amount of slips-of-actions to devalued outcomes is thought to resemble higher S-R habit responding. Therefore, we *post hoc* explored the direction of the association between the two systems by taking the two individual variables of the slips-of-action phase as a rough approximation of goal-directed and habitual behavior, and the MB and MF parameters of the two-step task.

As we had *a priori* hypotheses of a positive association between goal-directed and MB measures from the two tasks (Doll et al., 2012; Friedel et al., 2014; Gillan et al., 2015), we report one-tailed *p*-values. Additionally, we explored a positive association between habitual and MF behavior in the two tasks. Due to the non-normal distribution of the slips-of-action parameters and β_{MF}, we applied the more conservative Spearman correlation coefficient, in line with a previous cross-validation study (Friedel et al., 2014).

RESULTS

For sample description and scores on the general cognitive tests, see **Table 1**.

Three-Phase Instrumental Learning Task

As all variables of the instrumental learning task violated the assumption of normality (Shapiro-Wilk test: p's < 0.05), we report median and interquartile range (IQR) in addition to the average percentages and used non-parametric tests where necessary.

Discrimination Training Phase

All participants showed the expected learning effect, confirmed by a Friedman test that showed a significant increase in percentage of correct responses over the eight blocks ($X^2 = 103.922$, $p < 0.001$; see **Figure 3A**). A non-parametric binomial test shows that by the last block everyone had learned the correct responses to the stimuli significantly above chance level ($P(Y > 50 \mid n = 28, p = 0.5) < 0.001$), with an average percentage correct responses of 97.6% (SD = 4.45; median = 100%; IQR = 6.25) by block 8.

Outcome-Devaluation Test Phase

During the outcome test phase, participants showed an average percentage correct responses of 88.4% (SD = 15.17; median = 94.44; IQR = 10.42), which was also significantly above chance level ($P(Y > 50 \mid n = 28, p = 0.5) < 0.001$; see **Figure 3B**).

Slips-of-Action Phase

The crucial test of this task is the slips of action phase, where competition between outcome-based and stimulus-driven control is tested. Participants responded on average 82.2% (SD = 14.54; median = 85.42; IQR = 19.79) on stimuli that led to still-valuable outcomes. Slips of actions, that is, responding to stimuli that had a devalued outcome, occurred on average in 15.8% (SD = 16.92; median = 9.72; IQR = 13.19) of the devalued trials (see **Figure 3C**, left panel). The calculated DSI was 66.47 (SD = 29.60; median = 75.00; range: 26–95; IQR = 27.08), on average. Three statistical outliers ($z > 2$) had a DSI of around zero or below, mainly due to a low response rate on the stimuli associated with a still valuable outcome. These three participants also showed a deviating response pattern in the other phases (including discrimination training, outcome devaluation, baseline test) compared to the rest of the sample, by z-scores of or above $|2|$. For instance, they showed a response pattern of around chance level on the outcome devaluation

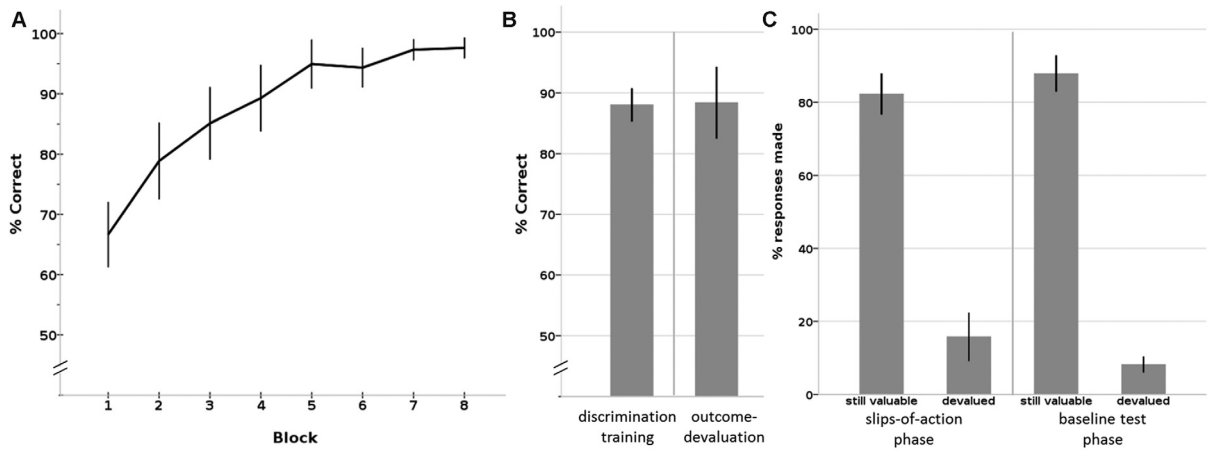

FIGURE 3 | Results of the Instrumental learning task. (A) Instrumental discrimination training phase, displayed as learning over eight blocks. By the eighth block, all participants had learned the correct S-R-O contingencies significantly above chance level. **(B)** Average percentage correct responses on the (total) discrimination training phase (left) and outcome devaluation phase (right). **(C)** Percentage responses to still valuable and devalued trials of the slips-of-action phase (left) and the baseline test phase (right). Error bars: 95% confidence interval.

phase and/or baseline test. As these three participants clearly did not show task participation, per chance by lack of attention or incomprehension of the instructions, we removed them from further correlation analyses.

The DSI showed a moderately positive correlation with other cognitive measures such as general intelligence, as assessed by the WMT and visual short-term memory, as assessed by the VPA (WMT: $\rho_{(25)} = 0.356$, $R^2 = 0.180$, $p = 0.040$; VPA: $\rho_{(25)} = 0.434$, $R^2 = 0.400$, $p = 0.015$). Cognitive processing speed, as measured with the DSST did not show a clear association with devaluation sensitivity ($\rho_{(25)} = 0.274$, $R^2 = 0.050$, $p = 0.108$).

Baseline Test Phase

During the baseline test, participants responded on average to 87.8% of the still-valuable stimuli (SD = 12.81; median = 92.36; IQR = 11.46) and on average to only 8.2% (SD = 5.63; median = 6.94; IQR = 7.64) of the devalued stimuli (see **Figure 3C**, right panel). The difference between %responses to still valuable and devalued trials was 79.61 (SD = 14.11; median = 83.33; IQR = 14.58), on average. This difference score is significantly higher than the difference score (DSI) on the slips-of-action phase (see above; Wilcoxon Signed Rank Test: $Z = -3.019$, $p = 0.002$). This shows that participants had no problems inhibiting their responses after *stimulus devaluation*.

Two-Step Sequential Decision-Making Task

Stay probabilities showed a significant main effect of reward ($F_{(1,27)} = 6.79$, $p = 0.015$), as well as a reward by state interaction ($F_{(1,27)} = 43.76$, $p < 0.001$), but no main effect of state ($F_{(1,27)} = 0.24$, $p = 0.628$; see **Figure 4**). This result replicates previous studies with the same task (Daw et al., 2011; Wunderlich et al., 2012b; Smittenaar et al., 2013; Deserno et al., 2015a; Gillan et al., 2015) and reflects an influence of both rewards and stay transitions on choice behavior. We further quantified this with computational modeling using a hybrid RL model that weights the relative influence of the MF and MB strategies. Distribution of modeling parameters is displayed in **Table 2**.

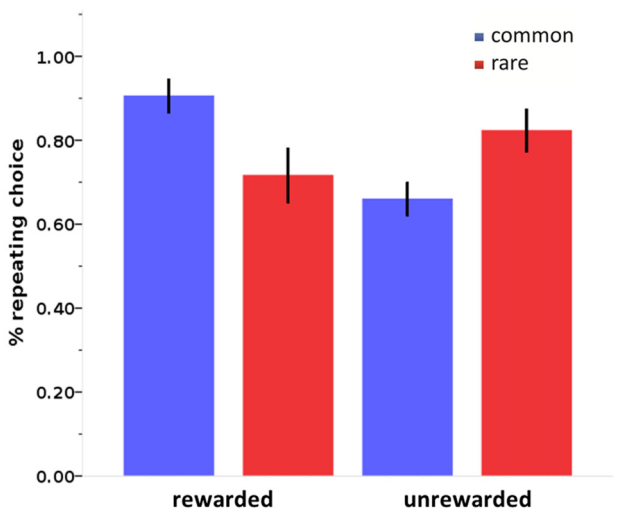

FIGURE 4 | Stay probabilities in the two-step sequential decision-making task. Stay probabilities in the two-step sequential decision-making task show a reward by state interaction. Error bars: 95% confidence interval.

TABLE 2 | Computational modeling parameter estimates.

Parameter	Mean	SD	Quantiles		
			25%	50% (median)	75%
ω	0.68	0.07	0.62	0.7	0.73
β_{MB}	5.28	1.80	3.64	5.58	6.51
β_{MF}	2.39	0.89	1.67	2.17	2.76
β_2	4.18	1.59	2.86	3.98	5.3
α_1	0.50	0.15	0.39	0.51	0.64
α_2	0.52	0.25	0.33	0.58	0.70
λ	0.52	0.24	0.33	0.52	0.71
ρ	0.13	0.03	0.11	0.13	0.16
–LL	179.41	37.45	155.76	186.19	209.51

Legend. ω: relative influence of model-free (MF) and model-based (MB) values. β: stochasticity of the choices for the first stage, under the MB system (β_{MB}), under the MF system (β_{MF}) and the second stage (β_2). α: learning rate for first (α_1) and second (α_2) stage. λ: reinforcement eligibility parameter (estimated value of the second stage should act as the same sort of MF reinforcer for the first stage choice). ρ: first-stage choice perseveration. –LL: negative log likelihood, indicating relative model-fit.

The parameters of interest of the two-step task also showed a moderate association with other cognitive measures. The balance parameter ω only showed a significantly positive correlation with the VPA and a moderate but non-significant (trendwise) association with WMT and DSST (VPA: $\rho_{(25)} = 0.422$, $R^2 = 0.131$, $p = 0.018$; WMT: $\rho_{(25)} = 0.271$, $R^2 = 0.058$, $p = 0.095$; DSST: $\rho_{(25)} = 0.301$, $R^2 = 0.094$, $p = 0.087$). The MB parameter β_{MB} was only significantly associated with the WMT but not (only trendwise) with the VPA (WMT: $\rho_{(25)} = 0.388$, $R^2 = 0.152$, $p = 0.028$; VPA: $\rho_{(25)} = 0.306$, $R^2 = 0.086$, $p = 0.068$; DSST: $\rho_{(25)} = 0.070$, $R^2 = 0.0036$, $p = 0.378$), whereas the MF parameter β_{MF} was not correlated with any of the other cognitive measures (WMT: $\rho_{(25)} = -0.010$, $R^2 = 0.028$, $p = 0.482$; VPA: $\rho_{(25)} = -0.142$, $R^2 = 024$, $p = 0.250$; DSST: $\rho_{(25)} = -0.212$, $R^2 = 0.181$, $p = 0.172$).

Construct Validity: Correlation Between the Two Paradigms

We tested how the parameters of the two-step task describing the individual influence of the two systems on choice behavior were related to devaluation sensitivity. The slips-of-action phase DSI correlated positively with β_{MB} of the two-step task ($\rho_{(25)} = 0.431$, $R^2 = 0.055$, $p = 0.016$), surviving Bonferroni correction for the three correlations of interest that were performed, but not with β_{MF} ($\rho_{(25)} = 0.172$, $R^2 = 0.003$, $p = 0.205$; see **Figure 5**). Next, we tested if a found association between the DSI and individual system parameters of the two-step was reflected in the balance parameter of the two-step task. We see a positive, albeit non-significant, relation between the balance parameters of the two tasks: ω of the two-step task and DSI of the slips-of-action phase correlated positively however, this was non-

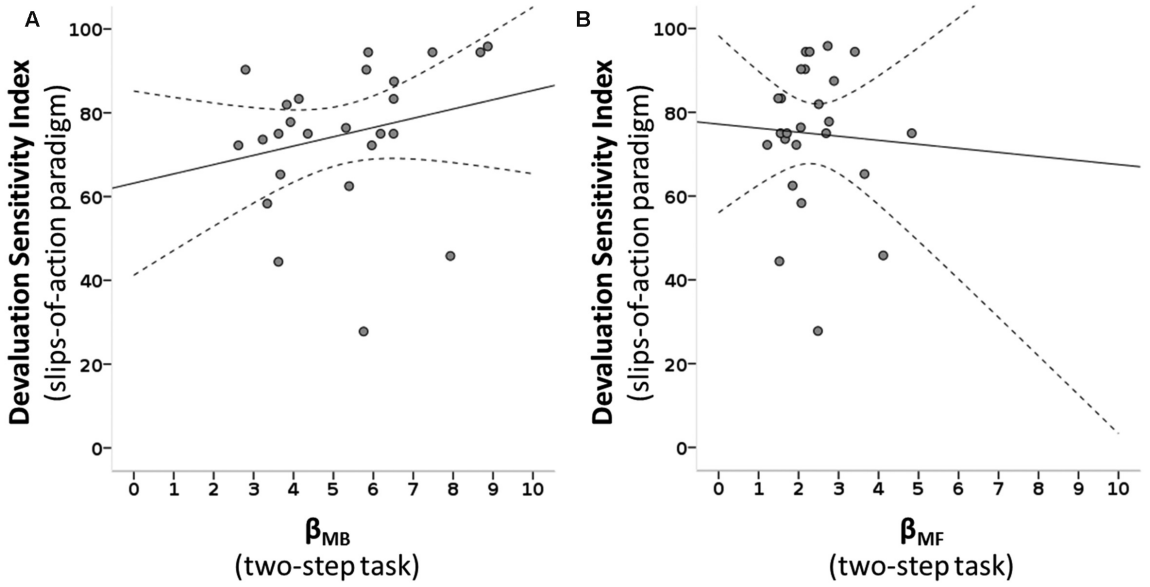

FIGURE 5 | Scatterplots of MB and MF choice values of the two-step task and devaluation sensitivity of the slips-of-action paradigm. (A) A positive correlation is seen between the MB parameter of the two-step task (β_{MB}) and the devaluation sensitivity index (DSI) of the slips-of-action task: $\rho_{(25)} = 0.431$, $R^2 = 0.055$, $p = 0.016$. **(B)** No significant association is seen between the MF parameter of the two-step task (β_{MF}) and the DSI: $\rho_{(25)} = 0.172$, $R^2 = 0.003$, $p = 0.205$. Dotted lines: 95% confidence interval.

significant, but showed only a trend ($\rho_{(25)} = 0.285$, $R^2 = 0.035$, $p = 0.083$).

Following the significant association between devaluation sensitivity and MB control we *post hoc* explored which of the two variables in the slips-of-action phase that contribute to the DSI score (% responses to valued and devalued trials), drove this significant association between β_{MB} and the DSI. The % responses on still valuable trials was positively associated with the MB variable ($\rho_{(25)} = 0.477$, $R^2 = 0.091$, $p = 0.008$), whereas % responses to devalued trials (slips-of-action) was not ($\rho_{(25)} = -0.123$, $R^2 = 0.012$, $p = 0.279$).

As goal-directed behavior in both tasks consistently correlated with an independent measure of visual short-term memory (VPA), we performed a *post hoc* mediation analysis (PROCESS Macro; Hayes, 2013) with bias corrected bootstrap confidence intervals to further elaborate a possible mediation factor in the three-way association. We entered VPA score as a mediator (M) in models with the DSI of the slips-of-action task as outcome variable (Y), and the MB parameter (β_{MB}) of the two-step task as independent variable (X). This model was significant ($R^2 = 0.373$, $F_{(1,23)} = 6.556$, $p = 0.006$), whereas the inverse model with X = DSI and $Y = \beta_{MB}$ was not ($p = 0.137$). Interestingly, this suggests a direction, where MB learning (β_{MB}) is a predictor for devaluation sensitivity, and not vice-versa. The direct effect between X and Y seemed to decrease when entering the mediators in the model (c'-path: $p = 0.730$). We tested the change from c to c' with the conservative Sobel's test, showing a moderate effect size, but no significance (c-c': $k^2 = 0.186$, $Z = 1.301$, $p = 0.193$, 95% CI = [0.03–0.44]). Note that although the Sobel test result is not significant per the p-value, the confidence interval does not include zero, which would lend support to the interpretation that there is a moderate effect size. Therefore, this mediation analysis points toward a partial mediation of visual short-term memory on the association between devaluation sensitivity and the individual MB parameter β_{MB}.

DISCUSSION

The aim of the current study was to cross-validate instrumental behavior from the goal-directed and habit systems assessed by two frequently used tasks. To this end, we correlated parameters assessing the involvement of the two systems from an instrumental learning task with an instructed devaluation slips-of-action phase (de Wit et al., 2007, 2012b) and a two-step sequential decision-making task (Daw et al., 2011). Partly conforming to our hypothesis, we see that MB control in the two-step task is moderately associated with goal-directed behavior in the slips-of-action paradigm, as β_{MB} correlated with the DSI. This effect of the DSI seemed mainly driven by responding to still valuable trials, and not by responding to devalued trials. An association between MB control and devaluation sensitivity was also partly captured by a moderate, albeit only trendwise significant, correlation between DSI and the balance parameter of the two-step task (ω), which assesses a relative involvement of the MB and MF systems in choice behavior. MF control did not seem to be significantly associated with devaluation sensitivity, which could have attenuated the association between the two balance parameters including ω of the two-step task. Ergo, we find a very moderate cross-validation

between these tasks on the assessment of behavior within the goal-directed system, whereas behavior within habit-like systems did not seem overtly related between the tasks.

A moderate correlation between MB learning and goal-directed devaluation sensitivity seems in agreement with a common framework to describe goal-directed behavior, as suggested by Dolan and Dayan (2013) and would indeed support common definitions of (aspects of) goal-directed behavior between the different task operationalizations. Comparably, in a previous construct validity study, Friedel et al. (2014) also found MB and goal-directed behavioral control to be positively correlated between the two-step sequential decision-making task (Daw et al., 2011) and a selective devaluation paradigm (Valentin et al., 2007), respectively. Furthermore, MB learning has been positively associated with goal-directed responding within one paradigm (Gillan et al., 2015). This is in line with the results in the current study, where we associated MB control with devaluation sensitivity between two separate paradigms. This suggests that computational accounts of MB control mirror, at least partly, one of the many aspects of (established) goal-directed behavior as measured with a selective devaluation paradigm.

However, we do have to caution that the amount of shared variance between both tasks was rather low. Moreover, it seems that MB behavior predicts performance on valued trials rather than responses towards devalued items, as it was mainly responding to still valuable trials on the slips-of-action phase that drove the association between devaluation sensitivity and MB behavior. Together, this suggests that each task does pick up distinct additional aspects of goal-directed behavior. It could be conceivable that the DSI more captures sensitivity to outcome value, whereas the two-step task (additionally) seizes sensitivity to outcome contingency; both part of the definition of goal-directed behavior. These distinct aspects of goal-directed behavior may be differently processed in the brain (however, for a study in primates, see Izquierdo et al., 2004). Nonetheless, considering the association between MB behavior and responses to still valuable items specifically, a part of the variance that is shared between the two constructs in this study might also be driven by performance more in general.

Interestingly, and in this line, both goal-directed and MB parameters were positively related to higher-order cognitive measures including visual short-term memory. Goal-directed behavioral control has been repeatedly shown to rely on higher-order cognitive measures, an effect most pronounced with working-memory capacity (Eppinger et al., 2013; Otto et al., 2013; Schad et al., 2014; Culbreth et al., 2016). Working memory capacity has even been shown to influence effects of detrimental environmental factors, such as stress on MB control (Otto et al., 2013). Although we did not directly measure working-memory capacity, we did have information on neurocognitive capacities in other cognitive domains. An exploratory mediation analysis indicated that short-term memory partly mediated the correlation between goal-directed and MB behavior in the two tasks, indicating that the tendency to be MB/goal-directed in each of the tasks depends on higher-order cognitive capacities, which could be part of the explanation of a moderate overlap between the two constructs.

Although negative results should be interpreted carefully, we would like to comment on the complete absence of a significant association between MF learning and habitual behavior in the two tasks, even when including the three participants that were regarded outliers on task behavior. We would like to discuss three possible explanations: (1) the assessments of the habitual system in the two tasks are unrelated; (2) habitual responding is the predominant mode of control leading to little variability, and thus correlation between paradigms; and (3) alternatively, the explanation might not lie at the level of construct, but in the (very goal-directed) sample tested. We discuss these explanations more in detail below.

The first intuitive explanation is that the differently measured aspects of habitual behavior are unrelated, either at the level of the two used paradigms, or more general at the level of construct definition. It is possible that the degree of MF control is not directly related to the propensity to form habits, but that the formation of action sequences might explain habitual actions, as suggested before Dezfouli and Balleine (2013). Contributing to this might be the distinction between assessing ongoing updating processes during the two-step vs. amount of slips of action as the expression of previously learned S-R associations. This might specifically be crucial for the assessment of the habit system. The acquisition vs. expression of habitual control is thought to be represented in distinct neural systems (Liljeholm et al., 2015; although integrative views have also been proposed), advancing the belief that behavioral acquisition vs. expression of habits is distinctively assessed. The two-step task has changing reward contingencies throughout the task and measures an ongoing (MF) TD learning process without reaching an asymptote, forming MF habit-like behavior by repeating previously rewarded choices without considering the task structure. A habit is thought to represent an automatized end-point of learning, while TD learning is, although slowly, still sensitive to changes in the environment. Conversely, the slips-of-action task evaluates the degree to which habitual behavior is expressed during an extinction test probing previously deterministically learned S-R associations. Therefore, it is possible that expressed behavior during an ongoing (MF) learning process differs from behavior observed during testing of established S-R associations. In comparison, the study by Gillan et al. (2015) associated MB and MF control within the same task with an instructed devaluation test assessing goal-directed and habitual behavior. In line with our findings they found that the degree of MB learning was also associated with devaluation sensitivity after the learning phase. MF learning however, was not associated with devaluation sensitivity, comparable to our results.

Second, an absent association between the tasks on the assessment of the habitual system might also reflect the robustness of the habitual/MF system, forming a predominant default mode of response (Wood and Rünger, 2016). The variability in the balance between the two decision-making systems might be predominantly driven by variability in the MB system, thus allowing cross-validation between paradigms in

the goal-directed system specifically, but not the habit system. The ubiquity of the goal-directed system (but not the habit system) has been acknowledged previously (Doll et al., 2012). However, it remains to be established how variability vs. stability in both systems constitutes a balance between goal-directed and habitual behavior (Lee et al., 2014). A default mode of response from the habit system would explain why an imbalance between the two systems in e.g., addiction seems to be driven by decreased goal-directed behavior as opposed to increased habitual responding (Sebold et al., 2014), although the opposite has also been described (Gillan and Robbins, 2014). Devaluation insensitivity after overtraining would then result from impaired goal-directed control instead of heightened habit formation. Of course, the strongest evidence for the notion that not either one, but a balance between two systems determines outcome devaluation sensitivity, comes from animal research, where a double dissociations is described: animals with lesions to dorsolateral striatum and infralimbic cortex are perpetually goal-directed, even after extensive training, whereas animals with dorsomedial striatal and prelimbic cortical lesions are habitual even after only minimal training (Yin et al., 2004, 2005b, 2006). It remains possible that the currently availably human tasks do not offer an adequate translation from the animal paradigms, and that the two tasks under scrutiny may not be optimally tailored to assess the contribution of habits in instrumental behavior. Importantly, neither of the paradigms compared in this study assesses full "end-stage" habits, which are typically manifest only after extensive overtraining (Colwill and Rescorla, 1988; Dickinson, 1994, 2015). Training in the current tasks lasts ten to 20 min, specifically the training phase of the instrumental learning task takes 16 encounters of every possible S-R-O association, which is more than the minimally needed amount to establish stable stimulus-based R-O associations, but less than some other studies (e.g., Tricomi et al., 2009). It therefore seems likely that these tasks fail to induce full end-stage habits. This complicates the discussion of habits for these tasks, and could further explain a lack of commonalities between the MF/habit constructs of the tasks. This end-stage phenomenon might not apply to the goal-directed system, as it is by definition more flexible and updated continuously, even after overtraining. The question further rises how these tasks under scrutiny relate to other tasks used to measure habit strength or related constructs, such as S-R instrumental learning tasks employing overtraining, skill learning tasks, spatial navigation tasks, the weather prediction task measuring implicit habit-like learning, and many others (Knowlton et al., 1994; Salmon and Butters, 1995; Gluck et al., 2002; Boettiger and D'Esposito, 2005; Marchette et al., 2011; Wood and Rünger, 2016).

As a third possibility for the absent association, it should be noted that the currently tested young and highly educated sample shows relatively dominant goal-directed (MB) behavior, which could (further) contribute to the fact that the current study only captures correlations between the tasks on the goal-directed systems. The average weighting parameter ω from the two-step task lies around 0.70, which is high compared to previous studies using the two-step task in healthy samples (Schad et al., 2014; Deserno et al., 2015a,b; Voon et al., 2015a,b; Morris et al., 2016; Worbe et al., 2016) and quantitatively indicates high involvement of the MB system. A predominant involvement of the MB system within this highly educated sample could lead to low variability or even a bottom-effect in the habitual system, rendering it harder to capture correlations between the MF/habit systems. Indeed, we see lower variability in the MF system than in the MB system, expressed by a lower variance of the β_{MF} parameter (see **Table 2**). Interestingly, and affirmatively, on the slips-of-action phase we see a relatively low percentage of responses to devalued trials (∼15% slips-of-actions), compared with existing literature, where the percentage slips-of-actions in healthy samples on averages lies around ∼30%–50% (Delorme et al., 2016; Ersche et al., 2016), indicating that in the current sample the habitual mode of control is relatively low in the slips-of-action phase. In line with the study by Gillan et al. (2015), this could be directly related to the relatively high involvement of the MB system during learning, as they have shown that MB learning protects against forming habits.

A limitation to this study is that due to the correlational associations between the tasks, we cannot elaborate on possible causal relationships between MB/MF learning and the expression of goal-directed/habitual responding during a devaluation test. The *post hoc* mediation analysis we performed did suggest directionality between MB control in the two-step task and devaluation sensitivity in the slips-of-action paradigm. This matches the directional association between MB control and devaluation sensitivity as reported by Gillan et al. (2015). However, in the current study set-up we can only refrain from further elucidations on this directionality.

In conclusion, the current study partly confirms a common framework between assessments of goal-directed and MB behavior, but we do not find such commonalities amongst the MF and habit system. However, the evidence is not strong: it should be acknowledged that the effects were only moderate, and the found associations explained only a small part of the shared variance, indicating there are still different aspects of goal-directed/MB behavior being picked up by the two tasks. Future studies should further elucidate these aspects, and the role of MF learning in forming habits. Moreover, systematic cross-validation on neural correlates of both instrumental decision-making systems is needed to further promote definition and adequate assessment of goal-directed and habitual behavior in healthy and diseased samples.

AUTHOR CONTRIBUTIONS

FS, AH and SW initiated the study. FS, AH, SW, AD and LD designed the study. AD and LD collected the data. LD and ZS performed the analyses. AD, LD, FS, AH and ZS interpreted the results. ZS drafted the article. ZS, AD, LD, SW, AV, H-JH, FS and AH read and corrected versions of the manuscript.

ACKNOWLEDGMENTS

The authors would like to thank M. Pössel, T. Wilbertz and K. Hudl for their assistance in recruitment and data

acquisition. We thank C. Edwards for proofreading the final version of the manuscript. This work was supported by the Max Planck Society, by grants from the German Research Foundation awarded to FS (Deutsche Forschungsgemeinschaft (DFG) SCHL 1969/2-2 as part of FOR1617) and AH (CRC 1052 "Obesity Mechanisms", subproject A05), and by a grant from Netherlands Organization for Scientific Research awarded to ZS (Rubicon 2014/05563/ALW). The work of AH is funded within the framework of the IFB Adiposity Diseases, Federal Ministry of Education and Research (BMBF), Germany, FKZ: 01E01001. AD is funded by a research grant from the FAZIT Foundation (http://www.fazit-stiftung.de). SW is supported by a Vidi grant 452-13-006 of the Netherlands Organization for Scientific Research (NWO).

REFERENCES

Abe, H., and Lee, D. (2011). Distributed coding of actual and hypothetical outcomes in the orbital and dorsolateral prefrontal cortex. *Neuron* 70, 731–741. doi: 10.1016/j.neuron.2011.03.026

Adams, C. D., and Dickinson, A. (1981). Instrumental responding following reinforcer devaluation. *Q. J. Exp. Psychol. B* 33, 109–121. doi: 10.1080/14640748108400816

Balleine, B. W., and Dickinson, A. (1998a). Goal-directed instrumental action: contingency and incentive learning and their cortical substrates. *Neuropharmacology* 37, 407–419. doi: 10.1016/s0028-3908(98)00033-1

Balleine, B. W., and Dickinson, A. (1998b). The role of incentive learning in instrumental outcome revaluation by sensory-specific satiety. *Anim. Learn. Behav.* 26, 46–59. doi: 10.3758/bf03199161

Balleine, B. W., and O'Doherty, J. P. (2010). Human and rodent homologies in action control: corticostriatal determinants of goal-directed and habitual action. *Neuropsychopharmacology* 35, 48–69. doi: 10.1038/npp.2009.131

Beck, A. T., Steer, R. A., and Brown, G. (1996). *Manual for the Beck Depression Inventory-II.* San Antonio, TX: Psychological Corporation.

Boettiger, C. A., and D'Esposito, M. (2005). Frontal networks for learning and executing arbitrary stimulus-response associations. *J. Neurosci.* 25, 2723–2732. doi: 10.1523/JNEUROSCI.3697-04.2005

Colwill, R. M., and Rescorla, R. A. (1988). The role of response-reinforcer associations increases throughout extended instrumental training. *Anim. Learn. Behav.* 16, 105–111. doi: 10.3758/bf03209051

Corbit, L. H., and Balleine, B. W. (2003). The role of prelimbic cortex in instrumental conditioning. *Behav. Brain Res.* 146, 145–157. doi: 10.1016/j.bbr.2003.09.023

Culbreth, A. J., Westbrook, A., Daw, N. D., Botvinick, M., and Barch, D. M. (2016). Reduced model-based decision-making in schizophrenia. *J. Abnorm. Psychol.* 125, 777–787. doi: 10.1037/abn0000164

Daw, N. D., Gershman, S. J., Seymour, B., Dayan, P., and Dolan, R. J. (2011). Model-based influences on humans' choices and striatal prediction errors. *Neuron* 69, 1204–1215. doi: 10.1016/j.neuron.2011.02.027

Daw, N. D., Niv, Y., and Dayan, P. (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nat. Neurosci.* 8, 1704–1711. doi: 10.1038/nn1560

Delorme, C., Salvador, A., Valabrègue, R., Roze, E., Palminteri, S., Vidailhet, M., et al. (2016). Enhanced habit formation in Gilles de la Tourette syndrome. *Brain* 139, 605–615. doi: 10.1093/brain/awv307

Deserno, L., Huys, Q. J. M., Boehme, R., Buchert, R., Heinze, H., Grace, A. A., et al. (2015a). Ventral striatal dopamine reflects behavioral and neural signatures of model-based control during sequential decision making. *Proc. Natl. Acad. Sci. U S A* 112, 1595–1600. doi: 10.1073/pnas.1417219112

Deserno, L., Wilbertz, T., Reiter, A., Horstmann, A., Neumann, J., Villringer, A., et al. (2015b). Lateral prefrontal model-based signatures are reduced in healthy individuals with high trait impulsivity. *Transl. Psychiatry* 5:e659. doi: 10.1038/tp.2015.139

Dezfouli, A., and Balleine, B. W. (2013). Actions, action sequences and habits: evidence that goal-directed and habitual action control are hierarchically organized. *PLoS Comput. Biol.* 9:e1003364. doi: 10.1371/journal.pcbi.1003364

Dickinson, A. D. (1985). Actions and habits: the development of behavioural autonomy. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 308, 67–78. doi: 10.1098/rstb.1985.0010

Dickinson, A. D. (1994). "Instrumental conditioning," in *Animal Learning and Cognition*, ed. N. J. Mackintosh (San Diego, CA: Academic Press), 45–79.

Dickinson, A. D. (2015). "Instrumental conditioning," in *Encyclopedia of Psychopharmacology*, eds I. P. Stolerman and L. H. Price (Berlin, Heidelberg: Springer), 823–828.

Dolan, R. J., and Dayan, P. (2013). Goals and habits in the brain. *Neuron* 80, 312–325. doi: 10.1016/j.neuron.2013.09.007

Doll, B. B., Simon, D. A., and Daw, N. D. (2012). The ubiquity of model-based reinforcement learning. *Curr. Opin. Neurobiol.* 22, 1075–1081. doi: 10.1016/j.conb.2012.08.003

Eppinger, B., Walter, M., Heekeren, H. R., and Li, S.-C. (2013). Of goals and habits: age-related and individual differences in goal-directed decision-making. *Front. Neurosci.* 7:253. doi: 10.3389/fnins.2013.00253

Ersche, K. D., Gillan, C. M., Jones, P. S., Williams, G. B., Ward, L. H. E., Luijten, M., et al. (2016). Carrots and sticks fail to change behavior in cocaine addiction. *Science* 352, 1468–1471. doi: 10.1126/science.aaf3700

Formann, A. K., and Piswanger, K. (1979). *Wiener Matrizen-Test WMT. Ein Rasch-skalierter Sprachfreier Intelligenztest [The Viennese Matrices Test: A Rasch-Scaled Culture-Fair Intelligence Test].* Weinheim: Beltz.

Friedel, E., Koch, S. P., Wendt, J., Heinz, A., Deserno, L., and Schlagenhauf, F. (2014). Devaluation and sequential decisions: linking goal-directed and model-based behavior. *Front. Hum. Neurosci.* 8:587. doi: 10.3389/fnhum.2014.00587

Geurts, H. M., and de Wit, S. (2014). Goal-directed action control in children with autism spectrum disorders. *Autism* 18, 409–418. doi: 10.1177/1362361313477919

Gillan, C. M., Otto, A. R., Phelps, E. A., and Daw, N. D. (2015). Model-based learning protects against forming habits. *Cogn. Affect. Behav. Neurosci.* 15, 523–536. doi: 10.3758/s13415-015-0347-6

Gillan, C. M., Papmeyer, M., Morein-Zamir, S., Sahakian, B. J., Fineberg, N. A., Robbins, T. W., et al. (2011). Disruption in the balance between goal-directed behavior and habit learning in obsessive-compulsive disorder. *Am. J. Psychiatry* 168, 718–726. doi: 10.1176/appi.ajp.2011.10071062

Gillan, C. M., and Robbins, T. W. (2014). Goal-directed learning and obsessive-compulsive disorder. *Philos. Trans. R. Soc. B Biol. Sci.* 369:20130475. doi: 10.1098/rstb.2013.0475

Gluck, M. A., Shohamy, D., and Myers, C. (2002). How do people solve the 'weather prediction' task? individual variability in strategies for probabilistic category learning. *Learn. Mem.* 9, 408–418. doi: 10.1101/lm.45202

Hampton, A. N., Bossaerts, P., and O'Doherty, J. P. (2006). The role of the ventromedial prefrontal cortex in abstract state-based inference during decision making in humans. *J. Neurosci.* 26, 8360–8367. doi: 10.1523/jneurosci.1010-06.2006

Hayes, A. F. (2013). *Introduction to Mediation, Moderation and Conditional Process Analysis*. New York, NY: Guilford.

Horstmann, A., Dietrich, A., Mathar, D., Pössel, M., Villringer, A., and Neumann, J. (2015). Slave to habit? Obesity is associated with decreased behavioural sensitivity to reward devaluation. *Appetite* 87, 175–183. doi: 10.1016/j.appet.2014.12.212

Huys, Q. J. M., Cools, R., Gölzer, M., Friedel, E., Heinz, A., Dolan, R. J., et al. (2011). Disentangling the roles of approach, activation and valence in instrumental and pavlovian responding. *PLoS Comput. Biol.* 7:e1002028. doi: 10.1371/journal.pcbi.1002028

Huys, Q. J. M., Eshel, N., O'Nions, E., Sheridan, L., Dayan, P., and Roiser, J. P. (2012). Bonsai trees in your head: how the pavlovian system sculpts goal-directed choices by pruning decision trees. *PLoS Comput. Biol.* 8:e1002410. doi: 10.1371/journal.pcbi.1002410

Huys, Q. J. M., Maia, T. V., and Frank, M. J. (2016). Computational psychiatry as a bridge from neuroscience to clinical applications. *Nat. Neurosci.* 19, 404–413. doi: 10.1038/nn.4238

Izquierdo, A., Suda, R. K., and Murray, E. A. (2004). Bilateral orbital prefrontal cortex lesions in rhesus monkeys disrupt choices guided by both reward value and reward contingency. *J. Neurosci.* 24, 7540–7548. doi: 10.1523/JNEUROSCI.1921-04.2004

Jonkman, S., Mar, A. C., Dickinson, A. D., Robbins, T. W., and Everitt, B. J. (2009). The rat prelimbic cortex mediates inhibitory response control but not the consolidation of instrumental learning. *Behav. Neurosci.* 123, 875–885. doi: 10.1037/a0016330

Killcross, S., and Coutureau, E. (2003). Coordination of actions and habits in the medial prefrontal cortex of rats. *Cereb. Cortex* 13, 400–408. doi: 10.1093/cercor/13.4.400

Knowlton, B. J., Squire, L. R., and Gluck, M. A. (1994). Probabilistic classification learning in amnesia. *Learn. Mem.* 1, 106–120.

Lee, S. W., Shimojo, S., and O'Doherty, J. P. (2014). Neural computations underlying arbitration between model-based and model-free learning. *Neuron* 81, 687–699. doi: 10.1016/j.neuron.2013.11.028

Liljeholm, M., Dunne, S., and O'Doherty, J. P. (2015). Differentiating neural systems mediating the acquisition vs. expression of goal-directed and habitual behavioral control. *Eur. J. Neurosci.* 41, 1358–1371. doi: 10.1111/ejn.12897

Marchette, S. A., Bakker, A., and Shelton, A. L. (2011). Cognitive mappers to creatures of habit: differential engagement of place and response learning mechanisms predicts human navigational behavior. *J. Neurosci.* 31, 15264–15268. doi: 10.1523/JNEUROSCI.3634-11.2011

McKim, T. H., Bauer, D. J., and Boettiger, C. A. (2016). Addiction history associates with the propensity to form habits. *J. Cogn. Neurosci.* 28, 1024–1038. doi: 10.1162/jocn_a_00953

Morris, L. S., Kundu, P., Dowell, N., Mechelmans, D. J., Favre, P., Irvine, M. A., et al. (2016). Fronto-striatal organization: defining functional and microstructural substrates of behavioural flexibility. *Cortex* 74, 118–133. doi: 10.1016/j.cortex.2015.11.004

O'Callaghan, C., Moustafa, A. A., Wit, S., Shine, J. M., Robbins, T. W., Lewis, S. J. G., et al. (2013). Fronto-striatal gray matter contributions to discrimination learning in Parkinson's disease. *Front. Comput. Neurosci.* 7:180. doi: 10.3389/fncom.2013.00180

Otto, A. R., Raio, C. M., Chiang, A., Phelps, E. A., and Daw, N. D. (2013). Working-memory capacity protects model-based learning from stress. *Proc. Natl. Acad. Sci. U S A* 110, 20941–20946. doi: 10.1073/pnas.1312011110

Radenbach, C., Reiter, A. M. F., Engert, V., Sjoerds, Z., Villringer, A., Heinze, H.-J., et al. (2015). The interaction of acute and chronic stress impairs model-based behavioral control. *Psychoneuroendocrinology* 53, 268–280. doi: 10.1016/j.psyneuen.2014.12.017

Rangel, A., Camerer, C., and Montague, P. R. (2008). A framework for studying the neurobiology of value-based decision making. *Nat. Rev. Neurosci.* 9, 545–556. doi: 10.1038/nrn2357

Redish, A. D., Jensen, S., and Johnson, A. (2008). A unified framework for addiction: vulnerabilities in the decision process. *Behav. Brain Sci.* 31, 415–437; discussion 437–487. doi: 10.1017/s0140525X0800472X

Reiter, A. M. F., Deserno, L., Wilbertz, T., Heinze, H.-J., and Schlagenhauf, F. (2016). Risk factors for addiction and their association with model-based behavioral control. *Front. Behav. Neurosci.* 10:26. doi: 10.3389/fnbeh.2016.00026

Salmon, D. P., and Butters, N. (1995). Neurobiology of skill and habit learning. *Curr. Opin. Neurobiol.* 5, 184–190. doi: 10.1016/0959-4388(95)80025-5

Schad, D. J., Jünger, E., Sebold, M., Garbusow, M., Bernhardt, N., Javadi, A.-H. H., et al. (2014). Processing speed enhances model-based over model-free reinforcement learning in the presence of high working memory functioning. *Front. Psychol.* 5:1450. doi: 10.3389/fpsyg.2014.01450

Schwabe, L., and Wolf, O. T. (2009). Stress prompts habit behavior in humans. *J. Neurosci.* 29, 7191–7198. doi: 10.1523/JNEUROSCI.0979-09.2009

Sebold, M., Deserno, L., Nebe, S., Schad, D. J., Garbusow, M., Hägele, C., et al. (2014). Model-based and model-free decisions in alcohol dependence. *Neuropsychobiology* 70, 122–131. doi: 10.1159/000362840

Sjoerds, Z., Wit, S., van den Brink, W., Robbins, T. W., Beekman, A. T. F., Penninx, B. W. J. H., et al. (2013). Behavioral and neuroimaging evidence for overreliance on habit learning in alcohol-dependent patients. *Transl. Psychiatry* 3:e337. doi: 10.1038/tp.2013.107

Smittenaar, P., FitzGerald, T. H. B., Romei, V., Wright, N. D., and Dolan, R. J. (2013). Disruption of dorsolateral prefrontal cortex decreases model-based in favor of model-free control in humans. *Neuron* 80, 914–919. doi: 10.1016/j.neuron.2013.08.009

Sutton, R. S., (Ed.) and Barto, A. G. (1998). *Reinforcement Learning: An Introduction, IEEE Transactions on Neural Networks*. (Vol. 9) Cambridge, MA: MIT Press.

Tricomi, E., Balleine, B. W., and O'Doherty, J. P. (2009). A specific role for posterior dorsolateral striatum in human habit learning. *Eur. J. Neurosci.* 29, 2225–2232. doi: 10.1111/j.1460-9568.2009.06796.x

Valentin, V. V., Dickinson, A., and O'Doherty, J. P. (2007). Determining the neural substrates of goal-directed learning in the human brain. *J. Neurosci.* 27, 4019–4026. doi: 10.1523/jneurosci.0564-07.2007

Voon, V., Baek, K., Enander, J., Worbe, Y., Morris, L. S., Harrison, N. A., et al. (2015a). Motivation and value influences in the relative balance of goal-directed and habitual behaviours in obsessive-compulsive disorder. *Transl. Psychiatry* 5:e670. doi: 10.1038/tp.2015.165

Voon, V., Derbyshire, K., Rück, C., Irvine, M. A., Worbe, Y., Enander, J., et al. (2015b). Disorders of compulsivity: a common bias towards learning habits. *Mol. Psychiatry* 20, 345–352. doi: 10.1038/mp.2014.44

Wechsler, D. (1955). *Wechsler Adult Intelligence Scale Manual.* New York, NY: Psychological Corporation.

Wechsler, D. (1987). *Manual for Wechsler Memory Scale-Revised.* San Antonio, TX: The Psychological Corporation.

Wechsler, D. (2006). *Wechsler-Intelligenztest für erwachsene: WIE; Übersetzung und Adaption der WAIS-III*, eds M. von Aster, A. Neubauer, and E. Horn (New York, NY: Harcourt Test Services).

de Wit, S., Barker, R. A., Dickinson, A. D., and Cools, R. (2011). Habitual versus goal-directed action control in Parkinson disease. *J. Cogn. Neurosci.* 23, 1218–1229. doi: 10.1162/jocn.2010.21514

de Wit, S., Corlett, P. R., Aitken, M. R. F., Dickinson, A., and Fletcher, P. C. (2009). Differential engagement of the ventromedial prefrontal cortex by goal-directed and habitual behavior toward food pictures in humans. *J. Neurosci.* 29, 11330–11338. doi: 10.1523/JNEUROSCI.1639-09.2009

de Wit, S., Niry, D., Wariyar, R., Aitken, M. R. F., and Dickinson, A. D. (2007). Stimulus-outcome interactions during instrumental discrimination learning by rats and humans. *J. Exp. Psychol. Anim. Behav. Process.* 33, 1–11. doi: 10.1037/0097-7403.33.1.1

de Wit, S., Standing, H. R., Devito, E. E., Robinson, O. J., Ridderinkhof, K. R., Robbins, T. W., et al. (2012a). Reliance on habits at the expense of goal-directed control following dopamine precursor depletion. *Psychopharmacology (Berl)* 219, 621–631. doi: 10.1007/s00213-011-2563-2

de Wit, S., Watson, P., Harsay, H. A., Cohen, M. X., van Vijver, I., and Ridderinkhof, K. R. (2012b). Corticostriatal connectivity underlies individual differences in the balance between habitual and goal-directed action control. *J. Neurosci.* 32, 12066–12075. doi: 10.1523/JNEUROSCI.1088-12.2012

de Wit, S., van Vijver, I., and Ridderinkhof, K. R. (2014). Impaired acquisition of goal-directed action in healthy aging. *Cogn. Affect. Behav. Neurosci.* 14, 647–658. doi: 10.3758/s13415-014-0288-5

Wood, W., and Rünger, D. (2016). Psychology of habit. *Annu. Rev. Psychol.* 67, 289–314. doi: 10.1146/annurev-psych-122414-033417

Worbe, Y., Palminteri, S., Savulich, G., Daw, N. D., Fernandez-Egea, E., Robbins, T. W., et al. (2016). Valence-dependent influence of serotonin depletion on model-based choice strategy. *Mol. Psychiatry* 21, 624–629. doi: 10.1038/mp.2015.46

Worbe, Y., Savulich, G., Wit, S., Fernandez-Egea, E., and Robbins, T. W. (2015). Tryptophan depletion promotes habitual over goal-directed control of appetitive responding in humans. *Int. J. Neuropsychopharmacol.* 18:pyv013. doi: 10.1093/ijnp/pyv013

Wunderlich, K., Dayan, P., and Dolan, R. J. (2012a). Mapping value based planning and extensively trained choice in the human brain. *Nat. Neurosci.* 15, 786–791. doi: 10.1038/nn.3068

Wunderlich, K., Smittenaar, P., and Dolan, R. J. (2012b). Dopamine enhances model-based over model-free choice behavior. *Neuron* 75, 418–424. doi: 10.1016/j.neuron.2012.03.042

Yin, H. H., Knowlton, B. J., and Balleine, B. W. (2004). Lesions of dorsolateral striatum preserve outcome expectancy but disrupt habit formation in instrumental learning. *Eur. J. Neurosci.* 19, 181–189. doi: 10.1111/j.1460-9568.2004.03095.x

Yin, H. H., Knowlton, B. J., and Balleine, B. W. (2005a). Blockaof NMDA receptors in the dorsomedial striatum prevents action-outcome learning in instrumental conditioning. *Eur. J. Neurosci.* 22, 505–512. doi: 10.1111/j.1460-9568.2005.04219.x

Yin, H. H., Ostlund, S. B., Knowlton, B. J., and Balleine, B. W. (2005b). The role of the dorsomedial striatum in instrumental conditioning. *Eur. J. Neurosci.* 22, 513–523. doi: 10.1111/j.1460-9568.2005.04218.x

Yin, H. H., Knowlton, B. J., and Balleine, B. W. (2006). Inactivation of dorsolateral striatum enhances sensitivity to changes in the action-outcome contingency in instrumental conditioning. *Behav. Brain Res.* 166, 189–196. doi: 10.1016/j.bbr.2005.07.012

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.