

This paper has not been peer reviewed (20/10/2021)

A Goal-Directed Account of Action Slips: The Reliance on Old Contingencies

*Eike K. Buabang¹², *Massimo Köster¹², Yannick Boddez³, Pieter Van Dessel³, Jan De Houwer³, Agnes Moors¹²

*Denotes shared first authorship

¹Research Group of Quantitative Psychology and Individual Differences, KU Leuven, Leuven, Belgium

²Center for Social and Cultural Psychology, KU Leuven, Leuven, Belgium

³Department of Experimental Clinical and Health Psychology, Ghent University, Ghent, Belgium

Author note

Eike K. Buabang  <https://orcid.org/0000-0002-3057-0819>

Massimo Köster  <https://orcid.org/0000-0001-5118-5963>

Yannick Boddez  <https://orcid.org/0000-0002-8067-5969>

Pieter Van Dessel  <https://orcid.org/0000-0002-3401-780X>

Jan De Houwer  <https://orcid.org/0000-0003-0488-5224>

Agnes Moors  <https://orcid.org/0000-0002-5137-557X>

We have no known conflict of interest to disclose. The preregistration, all data, and analysis code can be found [here](#) on the Open Science Framework

Preparation of this article was supported by Research Program G073317N of the Research Foundation - Flanders (FWO) and Grant C14/17/047 of the Research Fund of KU Leuven. Jan De Houwer is supported by Ghent University grant BOF16/MET_V/002.

Correspondence concerning this article should be addressed to Massimo Köster, Research Group of Quantitative Psychology and Individual Differences, KU Leuven, Tiensestraat 102, 3000 Leuven, Belgium. Email: massimo.koester@kuleuven.be

Word count: 8405

A GOAL-DIRECTED ACCOUNT OF ACTION SLIPS

Abstract

People often keep engaging in behaviors that used to be successful in the past but which are knowingly no longer effective in the current situation, so called action slips. Such action slips are often explained with stimulus-driven processes in which behavior is caused by a stimulus-response association and without information about the outcome of the behavior. This process is contrasted with a goal-directed process in which behavior is selected because it is expected to lead to a desired outcome. Failing to act in line with changes in the outcome, is taken as evidence for stimulus-driven processes. Stimulus-driven processes are assumed to get installed after overtraining and to be deployed under poor operating conditions. In line with this, previous research has found that action slips are more likely to occur after extensive training and when under time pressure. We propose an alternative goal-directed explanation according to which action slips are not caused by a stimulus-driven process, but rather by a goal-directed process that relies on old, no longer accurate, outcome information. In the current study, participants learned four stimulus-response-outcome contingencies during a single (i.e., moderate training) or a four-day training schedule (i.e., extensive training). Afterwards two contingencies were reversed. Results show that after extensive training and when under time pressure, participants not only committed more action slips but also reported more old response-outcome contingencies in line with these action slips. This is consistent with the goal-directed explanation that action slips result from a reliance on old, no longer accurate outcome information.

Keywords: action slip, stimulus-driven, habit, goal-directed, operant learning

A GOAL-DIRECTED ACCOUNT OF ACTION SLIPS

A Goal-Directed Account of Action Slips: The Reliance on Old Contingencies

Imagine you have been using the same password to unlock your computer for a year, but you recently changed it because your company asked you to. Yet despite knowing that you have changed it, you occasionally type in your old password again. Behaviors like these, which used to be successful in the past but are no longer effective today, are often called action slips (e.g., Watson et al., 2018). What is characteristic of these behavioral errors is that the behavior has become insensitive to changes in the outcome of the behavior (e.g., entering your old password no longer unlocks the computer), even if the agent knows that the outcome has changed (e.g., you know very well that you have changed the password).

To explain action slips, researchers typically invoke the operation of a stimulus-driven process in which a stimulus activates the association between stimulus and response representations, and this activation in turn causes the response: $S \rightarrow [S-R] \rightarrow R$ (de Wit et al., 2012). In the password example, the opening screen with the request to enter the password (S) activates the link between a representation of this screen with the representation of the behavior of typing in the password ([S-R]), and this in turn causes the entering of the password (R).

The stimulus-driven process is part of a dual-process model in which it is contrasted with a goal-directed process. In a goal-directed process, a response is selected based on the expected utilities of one or more response options. The expected utility of one response option reflects the value of the outcome of the response multiplied by the subjective response-outcome contingency, which is the degree to which a person expects that the response will lead to this outcome (i.e., expectancy). Thus, entering a password based on a goal-directed process involves a representation of the value of the outcome of having access to the computer (O^v) and of the contingency between entering the password and having access to the computer ($R-O^v$). To qualify as a goal-directed process, the expected utility of at least one response option should be processed. If more than one response option is considered, the expected utilities of these options are compared and the option with the highest expected utility should be selected.

The crucial difference between stimulus-driven and goal-directed processes is that the latter relies on representations of outcome knowledge whereas the former does not. One implication is that if changes in the actual outcomes take place, the goal-directed process can take this change into account whereas the stimulus-driven process cannot. Specifically, it is often assumed that, in the goal-directed process, the outcome knowledge gets updated, which

A GOAL-DIRECTED ACCOUNT OF ACTION SLIPS

leads to a change in the ensuing behavior. Thus, if the old password is replaced by a new password, the expectancy of unlocking the computer by typing the old password should turn to zero ($[R_{old}-O^v] = 0$) while the expectancy of unlocking the computer by typing the new password should turn to 1 ($[R_{new}-O^v] = 1$), and the person should now choose typing the new password. The stimulus-driven process, on the other hand, has no outcome knowledge that can be updated, and as a consequence, it cannot lead to an updating of behavior. Thus, if the opening screen of the computer is associated with the response to enter the old password, the person should continue entering the old password, that is, commit an action slip.

Traditional dual-process models endorse a default-interventionist architecture with regard to the interplay between stimulus-driven and goal-directed processes (Evans & Stanovich, 2013; Wood & Runger, 2016). They assume that stimulus-driven processes are the default cause of behavior and that it can occasionally be overridden by a goal-directed process. This architecture is based on assumptions about the ways in which stimulus-driven and goal-directed processes get installed and the conditions under which they get deployed. A common view is that a goal-directed process gets installed during a “moderate” operant conditioning procedure, in which the selection of a response in the presence of a certain stimulus is followed by a valued outcome on a moderate number of occasions. Stimulus-driven processes, on the other hand, can get installed by the mere co-occurrence between a stimulus and a response (i.e., Hebbian learning), but they can also get installed by an “extensive” operant conditioning procedure, in which the reinforcement of the behavior is repeated on an extensive number of occasions. This is supposed to stamp in the stimulus-response association and at the same time to weaken the response-outcome association so that the outcome of the behavior is no longer represented (Dickinson, 1985; Dickinson 2016) or no longer activated (Tricomi, et al., 2009). This type of stimulus-driven process is typically called a habitual process or habit. For instance, having entered a password repeatedly in the past creates a strong association between the opening screen of the computer and the response to enter that password while representations about the value and expectancy of having access to the computer are no longer activated ($[S-R-\emptyset^v]$).

Once stimulus-driven and goal-directed processes are installed, it is further assumed that both processes are deployed under different operating conditions. Stimulus-driven processes are assumed to be computationally simple, involving less complex mental representations, and therefore to proceed automatically. This means that they take place when operating conditions are poor (i.e., when opportunity, capacity, and/or motivation are low; Moors, 2016; Moors & De Houwer, 2006). Goal-directed processes, on the other hand, are

A GOAL-DIRECTED ACCOUNT OF ACTION SLIPS

seen as more computationally taxing, involving more complex mental representations, and therefore to be nonautomatic. This means that they can only intervene to correct stimulus-driven processes when operating conditions are ample (i.e., when opportunity, capacity, and/or motivation are high). Thus, the default-interventionist dual process model predicts a dissociation between the two processes. Poor operating conditions increase reliance on stimulus-driven processes while ample operating conditions allow for the operation of goal-directed processes. This suggests that poor operating conditions increase the likelihood that behavior is caused by a stimulus-driven process and hence that action slips are committed. In the password example, entering the old password should become more likely under conditions such as time pressure, mental load, and if the stakes are low.

Taken together, the default-interventionist dual process model proposes that action slips are caused by a stimulus-driven process. These are more likely to occur if the behavior was repeated and reinforced extensively in the past and when the conditions for deployment of goal-directed processes are poor.

Probably the strongest evidence for the stimulus-driven explanation of action slips to date comes from a study by Hardwick et al. (2019) in which sensitivity to a change in outcome contingencies was tested after an extensive operant conditioning procedure. In an acquisition phase, participants learned how to respond correctly to four visual stimuli. The four stimulus-response-outcome contingencies that were installed in this way were trained either moderately (1 day) or extensively (4 days). Subsequently, there was a contingency reversal phase in which the correct responses for two visual cues were switched while the other two remained the same. The new contingencies were trained by participants until they were able to demonstrate sufficient knowledge of the new contingencies. Finally, in the test phase, participants were presented with the stimuli and asked to respond in line with the new contingencies. Participants were forced to respond at a specific time between 0 and 1200ms. This allowed the researchers to compare responding under poor (i.e., little time) vs. ample (i.e., sufficient time) operating conditions. The results indicated that participants in the extensive training condition only made action slips (i.e., responded in line with the old stimulus-response-outcome contingencies) when they were forced to respond under time pressure (300-600ms), but not when they had more time (> 600ms). In the moderate training condition, action slips were neither observed under time pressure nor when there was more time. This finding is in line with the assumptions of the default-interventionist dual process model that action slips occur after extensive training, when a stimulus-response association had been formed, and when operating conditions were poor (i.e., time was scarce).

A GOAL-DIRECTED ACCOUNT OF ACTION SLIPS

Importantly, this study demonstrates that participants' behavior can be insensitive to a contingency change despite knowing about this change: Participants were able to demonstrate sufficient knowledge of the new contingencies after the contingency reversal but continued to act in line with the old contingencies (i.e., committed action slips) when these had been overtrained and when tested under time pressure. Hardwick and colleagues (2019) have taken the observed insensitivity to changes in the response-outcome contingencies as evidence for the operation of a stimulus-driven process. Moreover, the fact that the insensitivity was only observed after extensive training and when time was scarce was considered to be in line with the ideas that strong stimulus-driven processes get installed via overtraining and that they are especially likely to determine behavior under poor conditions.

Despite the popularity of stimulus-driven explanations for suboptimal behavior such as action slips, recent theoretical and empirical work has challenged the explanatory power of stimulus-driven processes (Buabang et al., 2021; De Houwer et al., 2018; Hogarth, 2020; Hommel & Wiers, 2017; Kruglanski & Szumowska, 2020; Moors et al., 2017). Moors et al. (2017) proposed an alternative dual process model with a parallel-competitive architecture in which stimulus-driven and goal-directed processes can both be automatic. Both processes will therefore often operate in parallel and compete with each other. The model further hypothesizes that the goal-directed process will often win the competition and therefore will determine behavior. Based on this model, these authors suggested that previous evidence for the operation of stimulus-driven processes may be overestimated and that many cases of supposedly stimulus-driven behavior such as action slips can be explained on the basis of goal-directed processes instead. This is possible because there is an asymmetry in the way in which the two processes are diagnosed (De Houwer et al., 2018; Dickinson, 2016). If a change in actual outcome values and response-outcome contingencies leads to a change in behavior, it is concluded that the behavior was caused by a goal-directed process. If no change in behavior occurs, on the other hand, it is concluded that the behavior was caused by a stimulus-driven process. In other words, the presence of a stimulus-driven process is inferred from a null-effect. This is problematic. Behavior may still be goal-directed and fail to change for other reasons.

One reason for the insensitivity of behavior to a change in outcome information regarding one outcome may be that the behavior is still under the control of another outcome that was not changed (Moors et al., 2017). In line with this, De Houwer and colleagues (2018) provided evidence for the idea that some instances of insensitivity to changes in outcome values, which were taken as evidence for stimulus-driven processes (e.g., de Wit et

A GOAL-DIRECTED ACCOUNT OF ACTION SLIPS

al., 2007), could have been directed at an outcome other than the one for which the value was changed. To illustrate this idea, consider the following situation. A person who wants to stop smoking may be advised to satisfy their goal of having a nicotine rush by applying a nicotine plaster, thereby allowing the person to reduce the value of nicotine from smoking cigarettes. However, if the person smokes for other reasons, such as to be able to do something with their hands during a conversation, then satisfying the need for nicotine is unlikely to change the smoking behavior. Thus, the fact that the smoking behavior did not change after changing one of its outcomes does not allow one to conclude that it was stimulus-driven.

Another reason, which we assume to be more likely in the case of action slips like those observed by Hardwick et al. (2019), is that the behavior may sometimes be under the control of the old response-outcome contingency even when people realize that this contingency is no longer valid. So far, researchers have followed the argument that when behavior is insensitive to a response-outcome contingency change despite the fact that the person demonstrates knowledge of this change, the process causing the behavior must be stimulus-driven (e.g., Vaghi et al., 2019). This conclusion seems to rely on the implicit assumption that a change in response-outcome contingencies in the environment results in an adjustment or overriding of the old contingencies with the new contingencies. If old response-outcome contingencies are completely overridden by new ones but the person nevertheless does not act in line with the new ones, it is inferred that the behavior does not rely on response-outcome contingencies but instead on stimulus-response links.

We challenge this assumption. We propose that a contingency change does not lead to the adjustment or overriding of the old contingency, but instead to the installation of a new contingency, while the old, no longer accurate contingency is retained. Based on that, action slips may not occur due to an absence of response-outcome contingency processing, but due to a reliance on information about old response-outcome contingencies (Moors et al., 2017). In the password example, people may enter the old password because they sometimes still rely on their expectation that the old password will give them access to the computer. This proposal is in line with research on extinction in which it is shown that people do not unlearn information (e.g., that a stimulus no longer signals harm) but rather that they learn novel information (e.g., that a stimulus does not signal harm in a certain context) that sits next to the old information (Bouton, 2002).

But how does this alternative account handle the observation that action slips occur more frequently after an extensive operant conditioning procedure and under poor operating conditions? The stimulus-driven account assumes that increased training of an operant

A GOAL-DIRECTED ACCOUNT OF ACTION SLIPS

conditioning procedure (beyond a certain point) results in a progressive *weakening* of the represented response-outcome contingency until a representation of the outcome is either absent (Dickinson, 1985; Dickinson 2016) or no longer activated (Tricomi, et al., 2009). At this point, behavior causation can occur without the activation of outcome knowledge.

According to the goal-directed account, by contrast, increased training of an operant conditioning procedure results in a progressive *strengthening* of the represented response-outcome contingency, which increases the likelihood that information about that contingency is retrieved. If, after the overtraining, the actual response-outcome contingencies change, the representation of the old, overtrained contingency is likely to be stronger than the representation of the new, less trained contingency. This relative difference in strength increases the likelihood that the old contingency is retrieved before the new contingency. Coming back to the password example, we assume that entering a password repeatedly strengthens the expectation that entering the password gives access to the computer. If after this extensive training, the password is changed, the contingency between entering the old password and computer access may still be stronger and may come to mind faster than the contingency between entering the new password and computer access. Note that the assumption that increased training leads to a strengthening rather than a weakening of memory traces is in line with what we know about memory in general (Hintzman, 1976).

Both the stimulus-driven and the goal-directed accounts of action slips assume that poor operating conditions, such as a lack of time, limit the amount of information that can be processed. This has the important implication that behavior under time pressure is likely to be determined by the mental representations that are retrieved first. The stimulus-driven account proposes that only stimulus-driven processes operate automatically and hence influence behavior more quickly than goal-directed processes. The alternative, goal-directed account, on the other hand, accepts that goal-directed processes can also operate automatically (Moors et al., 2017; see Pessiglione et al., 2007). Crucially, we propose that representations of old, overtrained contingencies are stronger than representations of newly installed contingencies and are therefore accessed faster and are more likely to determine behavior, especially when time is scarce. Coming back to the password example, time pressure makes it more likely that a person relies on the first contingency that comes to mind. If the expectation that entering the old password will provide access to the computer comes to mind faster, the person is likely to act based on that expectation when pressed for time and thus to enter the old password. In sum, the goal-directed account proposes that action slips are not the result of a stimulus-driven process in which stimulus-response associations drive responding, but

A GOAL-DIRECTED ACCOUNT OF ACTION SLIPS

instead of a goal-directed process in which a person strives to obtain a valued outcome but selects the wrong response based on the retrieval of and reliance on an overtrained but no longer valid response-outcome contingency.

The aim of the current study was to pit the goal-directed account of action slips against the stimulus-driven account. Both accounts predict that an operant conditioning procedure with extensive training will lead to an increase in action slips when operating conditions are poor. Importantly, however, the stimulus-driven and the goal-directed accounts of action slips make different predictions with regard to the retrieval of response-outcome contingencies. The stimulus-driven account predicts that following extensive training, representations of response-outcome contingencies are less likely to be retrieved under time pressure. The goal-directed account predicts the opposite, namely that overtrained response outcome contingencies are more likely to be retrieved under time pressure.

To study how practice and time pressure influence the retrievability of information about response-outcome contingencies, we designed a task consisting of an acquisition phase, a reversal phase, and a test phase (see Figures 1 and 2). In the acquisition phase, participants completed a series of choice trials to learn about the contingencies between stimuli, responses, and outcomes. In each trial, a pair of horizontally arranged doors was presented in one of four different colors. Participants learned for each colored pair of doors (i.e., stimuli) whether choosing the left or right door (i.e., behaviors) would yield a diamond or a rock (i.e., valued and nonvalued outcome). To manipulate the amount of training, participants were assigned to a moderate (1-day) or an extensive training condition (4-day). In the reversal phase, two stimulus-response-outcome contingencies were reversed while the other two remained the same. Thus, for two pairs of colored doors, the response that led to the valued outcome was changed to the opposite side. Hence, on reversed choice trials, the correct choice was opposite to what was correct in the acquisition phase whereas on unreversed choice trials, the correct choice was the same in all phases. Each participant had to perform choice trials until demonstrating sufficient knowledge of the new contingencies. Finally, in the test phase, participants completed choice trials in line with the new contingencies. These trials were intermixed with expectancy trials in which they were asked to report the currently valid response-outcome contingencies. In particular, participants were presented with a colored pair of doors as well as with a simulation of a response (i.e., a choice of the left or right side) and they had to indicate if, according to the current contingencies, the response would result in a diamond or a rock. On unreversed expectancy trials, the correct response was in line with the contingency that was valid during acquisition whereas on reversed

A GOAL-DIRECTED ACCOUNT OF ACTION SLIPS

expectancy trials, this was not the case. Importantly, both trial types had to be executed under time pressure. This allowed us to assess participants' response-outcome expectancies (on expectancy trials) under the same conditions as those under which the responses were given (on choice trials).

Based on the goal-directed account, we made the following concrete predictions for choice and expectancy trials in the test phase. First, we predicted that participants would commit a higher proportion of action slips on reversed compared to unreversed choice trials, and that this difference would be larger for participants in the extensive compared to the moderate condition. This would indicate that behavior becomes more insensitive to a contingency reversal after extensive compared to moderate training under time pressure. Such a result would replicate the main finding of Hardwick et al. (2019) and is also predicted by the stimulus-driven account.

Second, we predicted that participants would make a higher proportion of mistakes on reversed compared to unreversed expectancy trials and that this difference would be larger for participants in the extensive compared to the moderate condition. This would indicate that the information about the old contingency is still retrieved under time pressure when it has been previously overtrained. The stimulus-driven account would not make the same prediction. This account assumes that if participants demonstrated sufficient knowledge of the new contingencies during the reversal phase, the old contingencies must have been overwritten by the new ones. This account would thus not predict that the old contingency is especially likely to be retrieved under time pressure when it was previously overtrained.

Finally, we explored whether action slips (i.e., mistakes on reversed choice trials) are more likely for those trials on which the retrieval of old contingencies (i.e., mistakes on reversed expectancy trials) is more pronounced. An indicator for this would be a positive correlation between mistakes on reversed choice and reversed expectancy trials.

Method

Transparency and openness

We follow the guidelines outlined by JARS (Kazak, 2018) and report how we determined our sample size, data exclusions, all manipulations, and all measures. The method, materials, sample size estimate, inclusion criteria, and analysis plan were pre-registered on the Open Science Framework prior to data collection. This preregistration alongside all data and the analysis code are available at <http://osf.io/c6uh3>. The computerized task was programmed in PsychoPy (Peirce et al., 2019).

A GOAL-DIRECTED ACCOUNT OF ACTION SLIPS

Participants

The final sample consisted of 112 participants (56 per condition, 76 men, Age $M = 24.70$, $SD = 7.03$) collected through the participant pool of Prolific.co. This target sample size was determined with a preregistered power analysis conducted with the MorePower 6.0.4 software (Campbell & Thompson, 2012), which showed that this sample size was required to reliably detect an effect size of $\eta^2p = .068$, obtained in a pilot study¹, with statistical significance defined at the .05 level. To ensure that participants had learned the contingencies sufficiently they were only included in the analysis if they had at least 60% correct responses in the acquisition phase. Furthermore, to ensure sufficient data points per participant in the test phase, participants were also only included in the analysis if they had less than 30% late responses on choice and expectancy trials in the test phase. We recruited participants until we had 56 participants in each condition that met these inclusion criteria. In the moderate training condition, 106 participants completed the study out of which 56 met the criteria. In the extensive training condition, out of the 107 participants that completed the first day, 88 participants completed all four days, out of which 56 met the criteria. Participants in the moderate condition were paid 2.50 pounds for their participation. Participants in the extensive condition were paid 1.70 pounds for their participation on the first three days, and 2.50 pounds on the final day. To incentivize performance all participants were paid an additional bonus of 0.30 pounds upon completion of the experiment.

Procedure

The experiment was divided into an acquisition phase, a contingency reversal phase, and a test phase (Figure 1). Throughout these phases, participants performed two task types, namely a choice task and an expectancy task (Figure 2). To distinguish these two task types, trials for one task had a white frame and those for the other a black frame. The mapping of task type and frame color was randomized per participant and stayed the same throughout the experiment. The acquisition phase consisted only of choice trials in which participants learned to respond in line with four stimulus-response-outcome contingencies. The amount of training was manipulated by assigning participants randomly to a moderate or an extensive training condition at the beginning of the acquisition phase. Participants in the moderate training condition completed a single training session (i.e., 40 trials); those in the extensive training condition completed four training sessions spread out over four consecutive days

¹ The procedure of the pilot study was identical to the procedure of the final study and included 30 participants.

A GOAL-DIRECTED ACCOUNT OF ACTION SLIPS

(i.e., 340 trials in total: 100 trials on each of the first three days, 40 trials on the fourth day). Each training session was divided into five blocks. Thus, participants in the extensive condition had five blocks of 20 choice trials on each of the first three days and five blocks of eight choice trials on the fourth day. Participants in the moderate condition had five blocks of eight choice trials on their only day. On the last day of the extensive training condition and the only day of the moderate training condition, the acquisition phase was followed by a contingency reversal phase with only choice trials. In this phase, half of the contingencies from the acquisition phase were reversed while the other half remained unreversed. In the test phase, participants received a mix of choice and expectancy trials. In this phase, responding on the choice trials was done in line with the contingencies learned during the reversal phase. During the expectancy trials, participants indicated the outcome that they expected on the various types of trials. The design of our study involved a manipulation of the amount of training that participants received for choice trials (moderate vs. extensive) between-subjects and the contingency type (unreversed vs. reversed) within-subjects. A detailed description of each of the three phases (acquisition, contingency reversal, and test-phase) now follows.

Figure 1

Experimental procedure

| Training | Day 1 | Day 2 | Day 3 | Day 4 | | |
|------------------|--------------------|-------------------|-------------------|------------------|---------------------------------|--|
| | Acquisition | | | | Reversal | Test |
| Extensive | 100 choice trials | 100 choice trials | 100 choice trials | 40 choice trials | choice trials (criterion-based) | 48 choice trials 64 expectancy trials |
| Moderate | No training | | | | | |

Note. During the reversal phase, participants performed the task until they responded correctly for each pair of doors on five consecutive trials and contrary to all other phases there was no time pressure.

Acquisition phase. Participants received choice trials in which they learned to respond in line with four stimulus-response-outcome contingencies. Their goal was to collect as many diamonds as possible (see Figure 2 for an illustration). To incentivize performance, participants were promised a bonus payment on the last day of the study if they managed to collect a substantial number of diamonds. All participants who completed the experiment received this bonus regardless of their performance. Participants first saw a screen with the four pairs of colored doors (two red, two blue, two green, and two yellow) and which

A GOAL-DIRECTED ACCOUNT OF ACTION SLIPS

indicated for each pair whether a right or a left response would lead to a diamond (see Figure 1a). After this, on each trial, participants were presented with one pair of same-colored doors, one located on the left the other on the right. Participants chose for the left door by pressing left (on the “I”-key) with the right index finger or for the right door by pressing right (on the “P”-key) with the right ring finger. Each choice was followed by a 500ms visual animation in which a hand knocked on the chosen door. Afterwards, the chosen door opened and participants either received a diamond (i.e., valued outcome) or a rock (i.e., non-valued outcome). This outcome remained on screen for 1000ms. The mapping of doors and outcomes was randomized with the restriction that for two pairs of colored doors the diamonds could be obtained by pressing left and for the other two pairs of doors they could be obtained by pressing right. The order of appearance of the doors was randomized, but each door appeared equally often in each block. At the end of each block, participants were provided with feedback on how many diamonds and how many rocks they had collected until then. Furthermore, a response deadline was installed at 800ms. If this deadline was exceeded, the doors disappeared and participants lost the opportunity to receive a diamond or a rock. They nevertheless had to respond in order to continue, after which they received a feedback message stating “Too slow!”. On choice trials, we registered the number of trials on which a correct response was made, that is, a response that led to a diamond (i.e., valued outcome).

Contingency reversal phase. Participants again received choice trials. These were identical to the choice trials in the acquisition phase except that the responses required to receive diamonds were reversed for two pairs of colored doors. Participants were informed about this change with the help of the screen with the four pairs of colored doors, which indicated for each pair whether a right or left response would lead to a diamond. As in the study by Hardwick et al. (2019), no response deadline was installed in this phase to give participants plenty of opportunity to learn the new contingencies. Further, participants performed the task until they responded correctly for each pair of doors on five consecutive trials (i.e., criterion-based learning). This is the same criterion as used by Hardwick et al. (2019).

Test phase. Participants received choice trials intermixed with expectancy trials. The choice trials were identical to those in the contingency reversal phase except for the following changes. First, participants no longer received feedback, which means that they no longer saw whether they received a diamond or a rock. Testing was done under extinction to avoid any additional learning and thereby potential differences between participants. Nevertheless, participants were told that their responses were still recorded, and that they would receive

A GOAL-DIRECTED ACCOUNT OF ACTION SLIPS

feedback at the end of the study on how many diamonds they had collected. Second, participants were again instructed to respond before the 800ms response deadline. Although they did not receive an error message when they had responded too late, the doors still disappeared after this time and participants still had to respond in order to continue to the next trial.

During the expectancy trials, the goal of the participants was to predict as accurately as possible whether the outcome of a response was a diamond or a rock (see Figure 2 for an illustration). On each trial, participants were presented with a pair of doors and a visual animation of a hand knocking on the left or the right door. Each response illustration lasted 500ms and was identical to the response illustrations displayed on choice trials. Afterwards, the door on which the hand had knocked opened and a question mark appeared inside the opened door. Participants had to indicate whether they expected a diamond or a rock behind the opened door by pressing up (on the “E”-key) with their left index finger or down (on the “X”-key) with their left thumb. The correct keys to indicate whether the expected outcome was a diamond or rock were randomly assigned per participant. The response deadline on expectancy trials was limited to 500ms.² If this deadline was exceeded, the opened door disappeared, and participants lost the opportunity to respond. They nevertheless had to respond in order to continue, after which they received a feedback message stating “Too slow!”. On expectancy trials, we registered the number of trials on which a correct response was made. A response indicating diamond/rock counted as correct if a diamond/rock was indeed assigned to a door. Throughout the test phase, participants performed 48 choice trials intermixed with 64 expectancy trials. The order of appearance of each trial type was randomized and the entire phase was divided into four blocks each consisting of 28 trials (12 choice and 16 expectancy trials). Within each block and each trial type, each pair of doors appeared equally often. Each pair of doors was also equally often paired with left and right response animations on expectancy trials in each block.

Practice phases. To familiarize participants with the trials of the different phases, a number of practice phases were run prior to each phase. Before the acquisition phase,

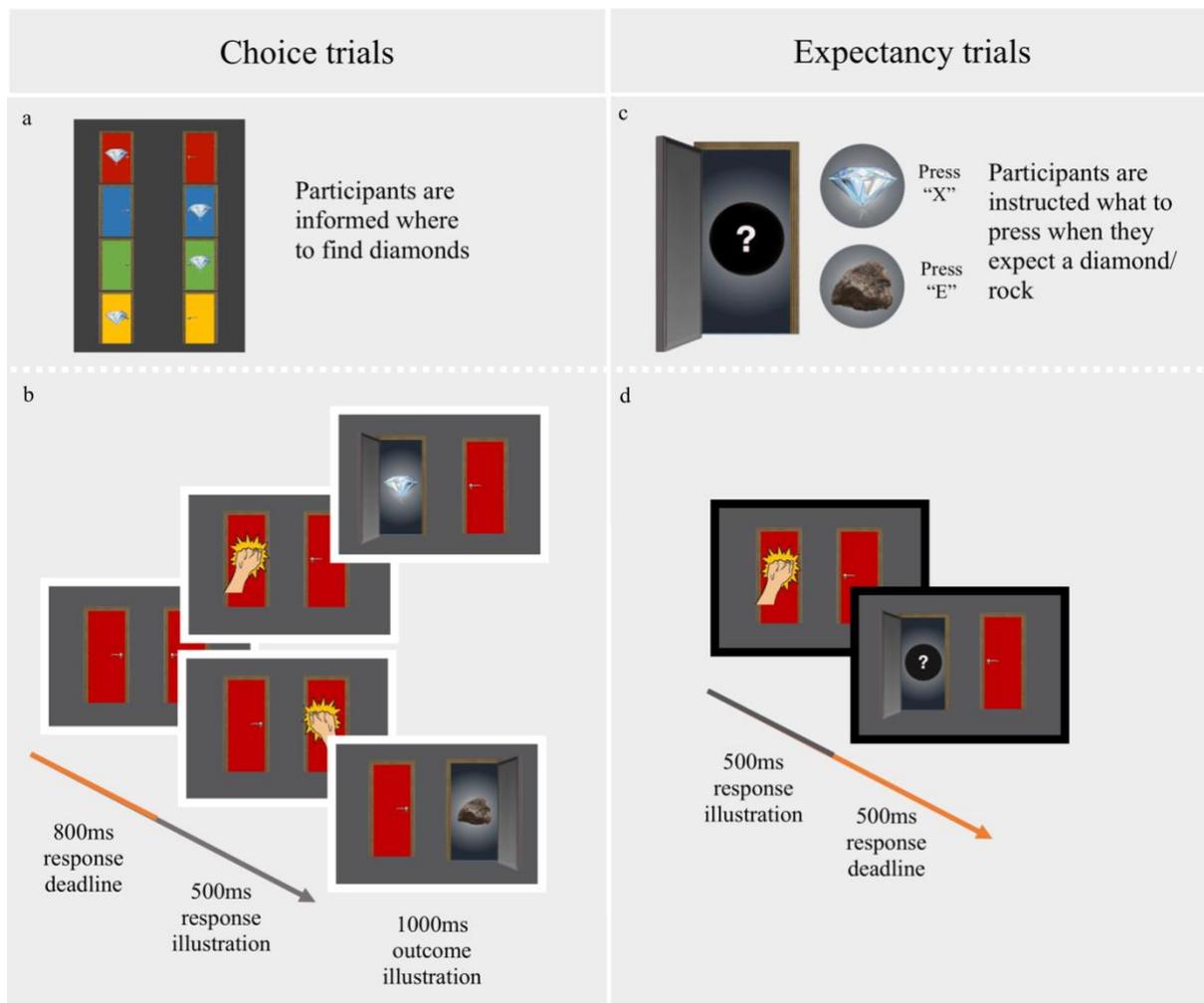
² The response deadline was 800ms on choice trials and 500ms on expectancy trials. The reason for installing a shorter response deadline on expectancy trials was that the 500ms response illustration preceding the response on expectancy trials allowed for additional response preparation, thereby increasing the actual time available on these trials. These response deadlines were determined through additional pilot testing.

A GOAL-DIRECTED ACCOUNT OF ACTION SLIPS

participants completed 12 practice choice trials. Prior to the reversal phase, they completed one practice block of 16 expectancy trials and one block of 28 intermixed trials (12 choice and 16 expectancy trials). The trials of these practice blocks were identical to the trials in the test phase with the only differences that half of the contingencies were not yet reversed and that participants received feedback on their performance. This means that (a) on choice trials they were shown whether their choice had resulted in a diamond or a rock (similar to the acquisition phase) and (b) on expectancy trials, they received feedback about the accuracy of their prediction about the outcome being a diamond or a rock via the messages “correct” or “incorrect”.

Figure 2

Trial types: Choice and expectancy trials



Note. ^aAcross all four pairs of doors, the diamond was always twice on the left and twice on the right side. Choosing the side opposite to the diamond led to a rock. ^bEach choice trial started with a fixation cross and a white or black frame (randomized across task types for each participant). Participants were presented with a pair of doors and had to choose a door

A GOAL-DIRECTED ACCOUNT OF ACTION SLIPS

within the response deadline (800ms). The choice was followed by a visual animation of a hand knocking on the corresponding door. Depending on the chosen door, participants were presented with a diamond or a rock. ^c The allocation of the keys that need to be pressed to indicate that the expected outcome is a diamond or a rock was randomized for each participant. ^d Each expectancy trial started with a fixation cross and a white or black frame (randomized across trial types for each participant). Participants were presented with a pair of doors and the visual animation of a response. The corresponding door opened and participants were presented with a question mark. Participants had to indicate the expected outcome within the response deadline (500ms).

Results

The analyses were conducted using the R software (R Core Team, 2019) package *rstatix* (Kassambara, 2020). For all ANOVAs, the sphericity assumption was checked with Mauchly's test. Corrected results (Greenhouse-Geisser) are reported when the sphericity assumption was violated. For the acquisition phase, only the fourth day of the extensive training condition and the only day of the moderate condition are reported. For this phase, the proportion of correct responses was calculated per block (i.e., across every eight trials). For the test phase, the proportion of mistakes was calculated across all blocks for choice trials (i.e., across all 48 trials) and expectancy trials (i.e., across all 64 trials).

Acquisition phase

To check whether participants learned the contingencies in the acquisition phase sufficiently, we conducted a mixed model ANOVA to analyze the proportion of correct responses with condition (moderate vs. extensive) as between-subjects factor and acquisition block (1 to 5) as within-subjects factor.

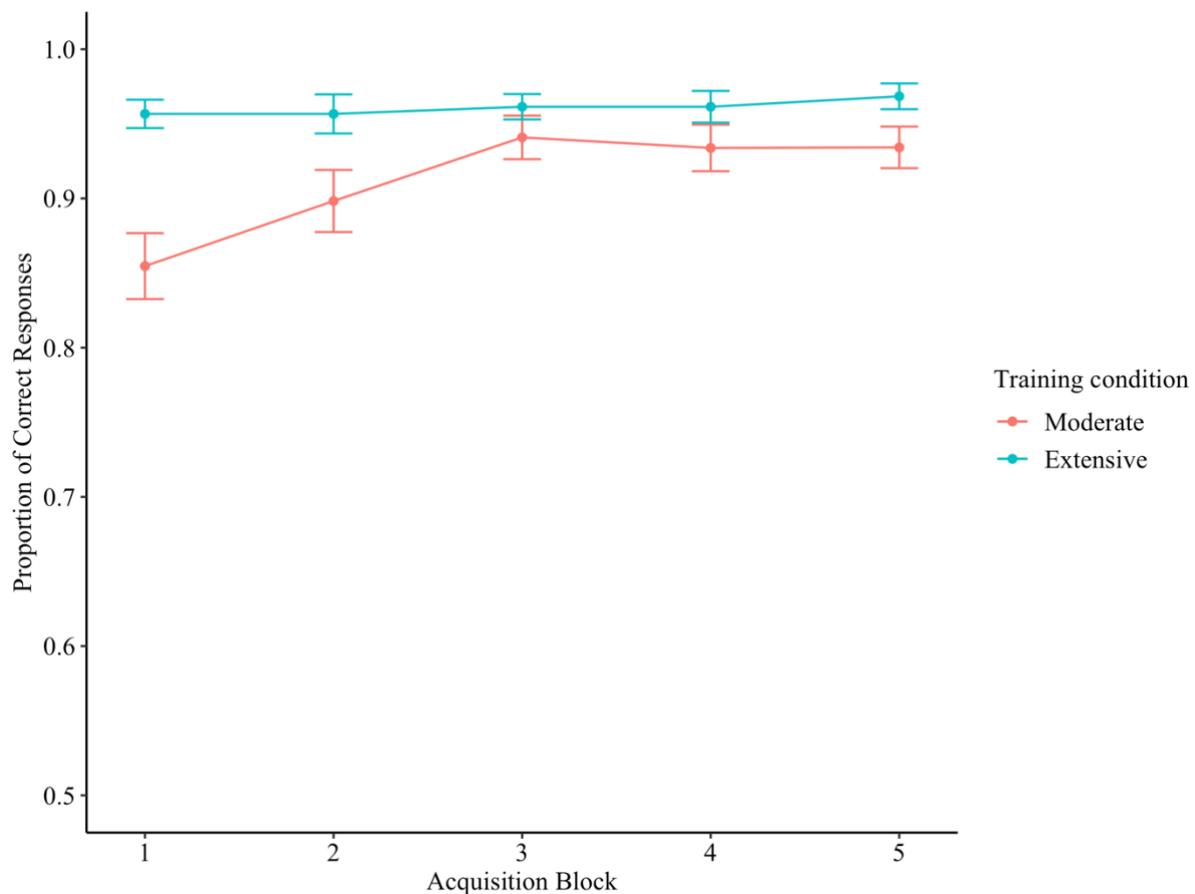
The main effect of condition was significant, $F(1, 110) = 12.24, p < .001, \eta^2_p = .10$, indicating that participants in the extensive training condition ($M = .96, SD = .08$) made more correct responses than those in the moderate training condition ($M = .91, SD = .14$). This suggests that overall, the former participants learned the contingencies better than the latter, which is not surprising given that the former already had received three days of training before the final acquisition phase. The interaction between condition and block was also significant, $F(3.38, 372.14) = 3.88, p = .007, \eta^2_p = .03$. Post-hoc comparisons showed that the moderate training condition made less correct responses than the extensive training condition in the first block (moderate condition: $M = .86, SD = .17$; extensive condition: $M = .96, SD = .07$), $t(110) = -4.24, p < .001$, in the second block, (moderate condition: $M = .90, SD = .16$; extensive condition: $M = .96, SD = .10$) $t(110) = -2.37, p = .019$, and in the fifth block, (

A GOAL-DIRECTED ACCOUNT OF ACTION SLIPS

moderate condition: $M = .93$, $SD = .10$; extensive condition: $M = .97$, $SD = .06$), $t(110) = -2.09$, $p = .039$. The comparisons for Blocks 3 and 4 did not reach significance ($ts < 1.46$). This suggests that the difference in performance between the conditions was larger in early blocks and became smaller over time (see Figure 3). Although participants in the extensive condition performed significantly better than those in the moderate condition, up until the final block, it is worth noting that both groups reached an accuracy of at least 93%.

Figure 3

Acquisition performance per condition (Moderate: Day 1, Extensive Day 4)



Note. Errors bars represent standard errors

Reversal phase

Participants had to make five consecutive correct responses for each stimulus during this phase and there was no time pressure. To check for potential differences in reversal learning performance between the conditions, we conducted a mixed model ANOVA to analyze the number of trials with condition (moderate vs. extensive) as between-subjects factor and contingency type (unreversed vs. reversed) as within-subjects factor. No effects reached significance ($Fs < 1.52$). These results suggest that participants in the moderate ($M =$

A GOAL-DIRECTED ACCOUNT OF ACTION SLIPS

36.60, $SD = 13.70$) and extensive condition ($M = 39.40$, $SD = 18.50$) required a similar number of trials to learn the new contingencies.

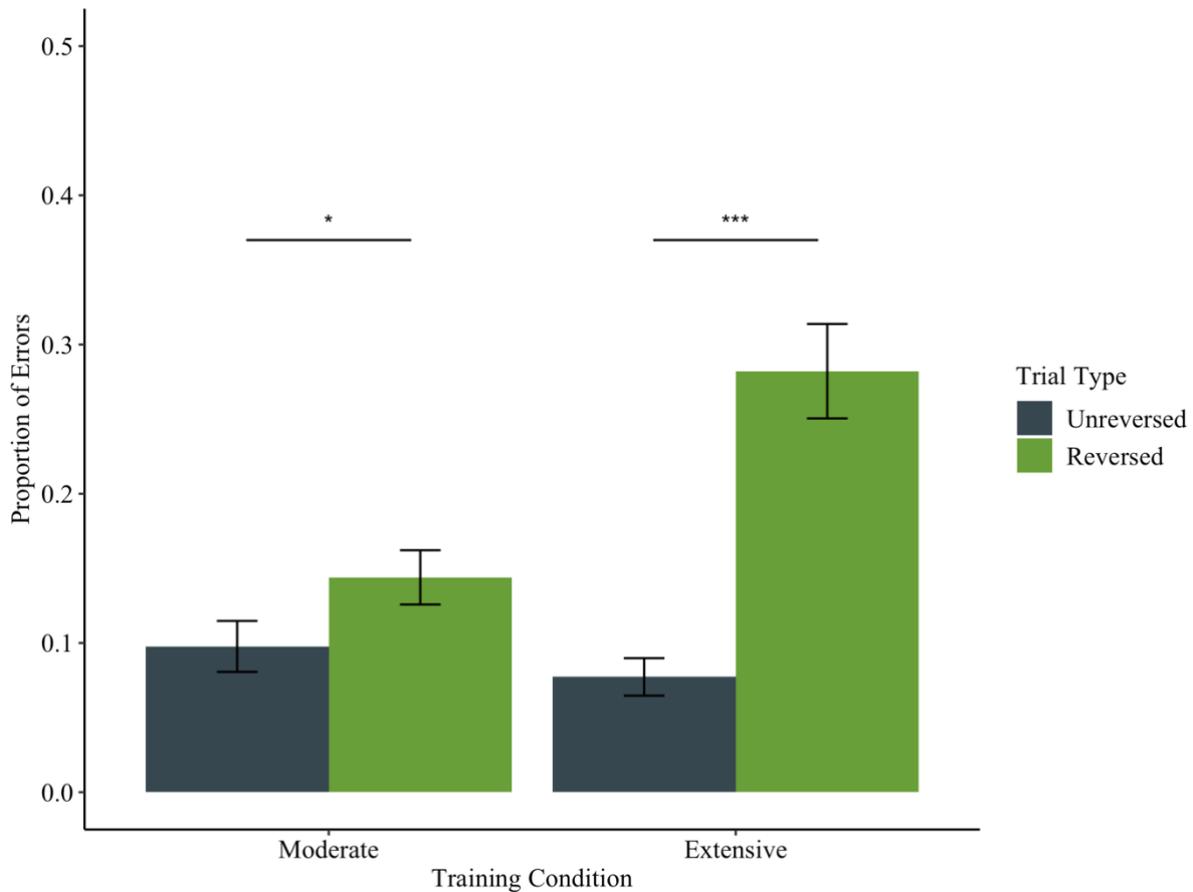
Test phase

Choice trials. To test the first prediction of the goal-directed account, we conducted a mixed model ANOVA to analyze the proportion of errors (i.e., responses other than those required by the contingencies in the test phase) with condition (moderate vs. extensive) as between-subjects factor and contingency type (unreversed vs. reversed) as within-subjects factor.

We found a significant main effect of contingency type, $F(1, 110) = 52.98$, $p < .001$, $\eta^2_p = .33$, suggesting that participants made more mistakes on reversed ($M = .21$, $SD = .20$) than unreversed trials ($M = .09$, $SD = .11$). In line with our prediction, the interaction between condition and contingency type was significant, $F(1, 110) = 21.35$, $p < .001$, $\eta^2_p = .16$, indicating that the proportion of errors on reversed compared to unreversed trials was larger for participants in the extensive condition than in the moderate condition. Post-hoc comparisons indicated that participants in the moderate training condition made more mistakes on reversed ($M = .14$, $SD = .14$) than on unreversed trials ($M = .10$, $SD = .13$), $t(55) = 2.53$, $p = .014$. Participants in the extensive training condition also made more mistakes on reversed ($M = .28$, $SD = .24$) than on unreversed trials ($M = .08$, $SD = .09$), $t(55) = 7.00$, $p < .001$. Crucially, the significant interaction effect that we obtained indicates that the difference between reversed and unreversed trials was larger in the extensive than in the moderate condition (see Figure 4). In sum, these results indicate that action slips are more likely to occur after extensive training compared to moderate training under time pressure, which is in line with previous research (Hardwick et al., 2019).

A GOAL-DIRECTED ACCOUNT OF ACTION SLIPS

Figure 4

Choice task performance

Note. Errors bars represent standard errors

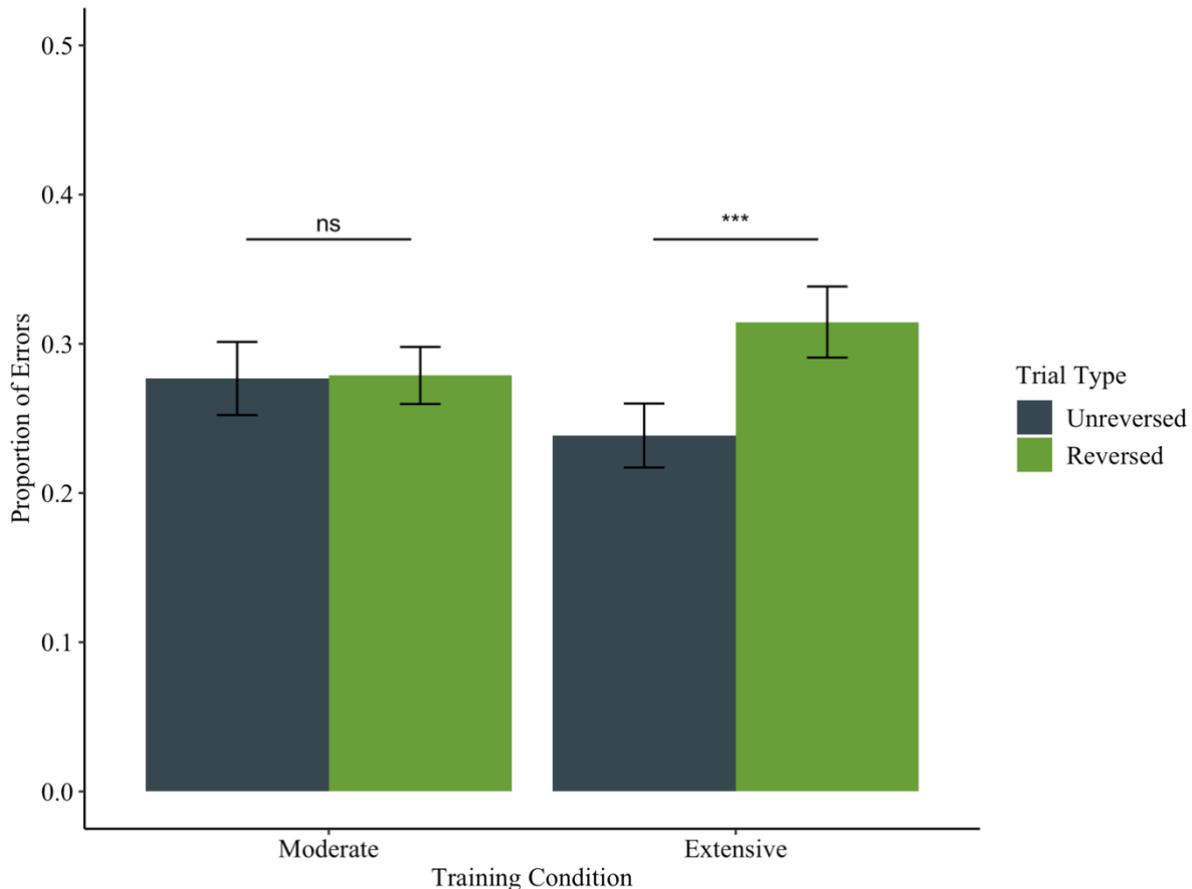
Expectancy trials. To test the second prediction of the goal-directed account, we conducted a mixed model ANOVA to analyze the proportion of errors (i.e., responses other than those in line with the contingencies in the test phase) with condition (moderate vs. extensive) as between-subjects factor and contingency type (reversed vs. unreversed) as within-subjects factor. We found a significant main effect of contingency type, $F(1, 110) = 7.69, p = .007, \eta^2_p = .07$, suggesting that participants made more errors on reversed ($M = .30, SD = .16$) than unreversed trials ($M = .26, SD = .16$). In line with our prediction, the interaction between condition and contingency type was significant, $F(1, 110) = 6.91, p < .001, \eta^2_p = .06$, suggesting that the proportion of errors on reversed compared to unreversed trials was larger for participants in the extensive condition than in the moderate condition. Post-hoc comparisons indicated that in the moderate condition the proportion of errors did not significantly differ between reversed ($M = .28, SD = .14$) and unreversed trials ($M = .28, SD = .18$), $t(55) = 0.11, p = .913$. Crucially, in line with the prediction by the goal-directed account, participants in the extensive training condition made significantly more errors on

A GOAL-DIRECTED ACCOUNT OF ACTION SLIPS

reversed ($M = .32$, $SD = .16$) than on unreversed trials ($M = .24$, $SD = .16$), $t(55) = 3.57$, $p < .001$. As indicated by the post-hoc comparisons, only the extensive condition made more mistakes on reversed compared to unreversed expectancy trials (see Figure 5). In line with the goal-directed account, these results suggest that old contingencies are more likely retrieved after extensive training compared to moderate training under time pressure.

Figure 5

Expectancy task performance



Note. Errors bars represent standard errors

Exploratory analysis

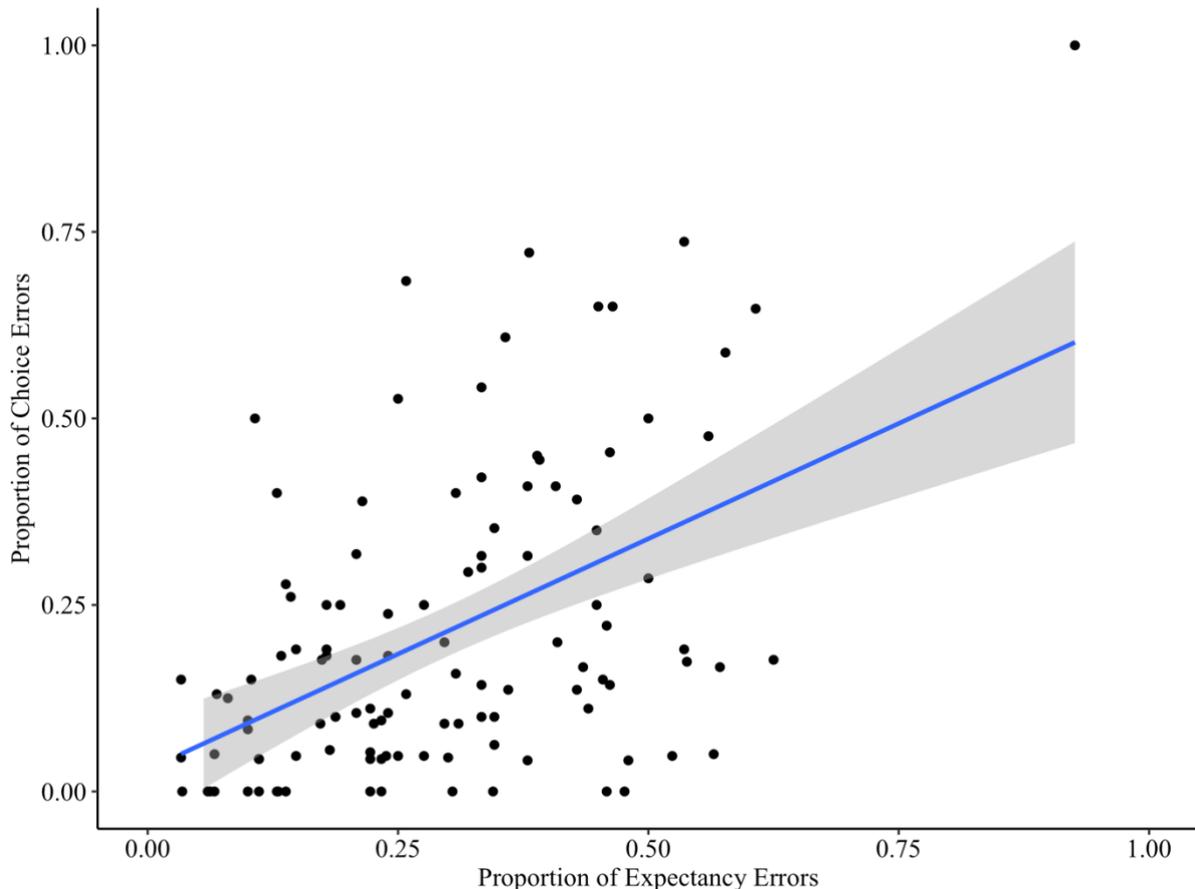
If action slips (errors on reversed choice trials) occur due to a retrieval of information about old contingencies (errors on reversed expectancy trials), participants who make more errors on reversed choice trials should also be more likely to make more errors on reversed expectancy trials. To explore this, we tested if the proportion of mistakes for reversed expectancy trials is positively correlated with the proportion of mistakes on reversed choices trials. This correlation reached significance, $\rho = .49$, 95% CI = .334 to .619, $p < .001$ (see Figure 6), suggesting that participants who were more likely to retrieve old contingencies were also more likely to make action slips. It remains possible, however, that the correlation

A GOAL-DIRECTED ACCOUNT OF ACTION SLIPS

is due to a general ability to perform computerized tasks. If so, we might expect performance on unreversed expectancy trials to be similarly correlated with performance on reversed choice trials. Yet this correlation turned out to be not significant, $\rho = .13$, 95% CI = $-.054$ to $.311$, $p = .162$.

Figure 6

Reversed expectancy-choice task correlation



Note. Confidence band represent 95% confidence interval

Discussion

People sometimes commit action slips, such as entering an old password despite knowing that the password has changed. What is characteristic of action slips is that they do not align with known changes in response-outcome contingencies. Previous research has found that action slips are likely to occur when a behavior has been repeated extensively in the past and when a person has to act under poor operating conditions, such as under time pressure. To explain action slips, advocates of the stimulus-driven account (e.g., Hardwick et al, 2019) invoke the operation of a stimulus-driven process in which behavior is caused by a stimulus-response association that contains no information about the outcomes of this

A GOAL-DIRECTED ACCOUNT OF ACTION SLIPS

response. In the current study, we put forward the alternative proposal that action slips are caused by a goal-directed process that involves representations of old, no longer accurate, contingencies. To test our alternative explanation, we developed a paradigm with the aims of (a) replicating the finding that overtraining of a behavior followed by a contingency change results in action slips under time pressure and (b) testing whether time pressure results in the retrieval of information about old, no longer accurate, but overtrained contingencies. In line with previous research, we found that participants committed more action slips after extensive training and while under time pressure. Our second finding was that under time pressure, participants were especially likely to respond in line with the contingencies that were correct prior to the contingency change when these have been previously overtrained. This finding is in line with the goal-directed account, suggesting that action slips may occur because people sometimes act based on a goal-directed process which involves old, no longer accurate, contingencies. An exploratory analysis furthermore showed that the number of action slips was positively correlated with the extent to which people reported old contingencies. This suggests that participants who committed more action slips were also more likely to report the old contingency that was correct prior to the reversal. This strengthens the goal-directed interpretation of our results.

Action slips are one type of behavior for which stimulus-driven processes have been invoked as the driving mechanism. While many action slips such as typing in the old password may be rather inconsequential, more serious forms of suboptimal behaviors such as addictions, compulsive behaviors, and unhealthy eating have also been explained with stimulus-driven processes (Ersche et al., 2016; Gillan et al., 2016; Pierce-Messick & Corbit, 2021). Following the current study, some of these other types of suboptimal behavior may also or partly be explained by a reliance on old contingencies. Imagine the case of a person who just received the diagnosis that they have diabetes. In the past, the person always drank fresh orange juice in order to reach the outcome of feeling healthy. After the diagnosis, the person gained the novel knowledge that in order to feel healthy, it is better to drink water. Nevertheless, when the person is at a reception and they are fully engaged in a conversation, the person may grab the orange juice rather than the water from the tray.

As such, our findings have widespread implications for understanding and changing behavior. In particular, two promising targets for behavior change may be the storage and retrieval of old contingencies. Our findings suggest that old contingencies are stored even though people no longer report them explicitly. Thus, one fruitful intervention may be to weaken or remove the old contingency from memory, for instance, through reconsolidation

A GOAL-DIRECTED ACCOUNT OF ACTION SLIPS

(Else et al., 2018). It should be noted, however, that recent debates highlight that the boundary conditions of effective reconsolidation are not well known yet and that a number of studies failed to find effective reconsolidation effects (Schroyens, Beckers, & Luyten, 2021).

Our findings further suggest that old contingencies are particularly likely to determine behavior under poor operating conditions. Thus, one approach to limit the influence of old contingencies may be through strengthening the new contingencies, which would make them in turn more likely to be retrieved under such poor conditions. One way to strengthen new contingencies is via repetition. Although the stimulus-driven account also recommends repetition of the new contingencies as a way to override old contingencies, this strategy should only be effective if the repetition is extensive so that a novel stimulus-driven process is installed that can operate under poor operating conditions. The goal-directed account not only takes repetition to strengthen new contingencies but also envisages other ways to strengthen them, for example, by changing their vividness via simulation or imagination training (Papies & Barsalou, 2015).

The results of the current study also hold a key implication for research trying to show that stimulus-driven processes occur under poor operating conditions whereas goal-directed processes happen only under ample operating conditions. Studies on action slips have shown that participants demonstrate sufficient knowledge of the changed outcome contingencies when processing conditions are ample but that they do not act in line with this knowledge when conditions are poor. This has led researchers to conclude that no outcome knowledge can be activated whatsoever under poor conditions. This is problematic because our results show that the use of outcome knowledge under poor conditions is different from that under ample conditions rather than that it is absent. Based on this idea, we argue that in order to investigate whether behavior under poor operating conditions was caused by outcome knowledge or not, it is crucial to make sure that outcome knowledge is assessed under the same conditions as the to-be-explained behavior.

Although our results show that people reported response-outcome contingencies in line with their action slips when tested under the same conditions, we cannot be certain that these contingencies were in fact processed at the time the action slips were made and that they drove these action slips. Participants were instructed to report on their contingency knowledge during the expectancy trials, which may have encouraged them to intentionally process the contingencies. It is possible, however, that at the time at which the action slips were committed, this contingency knowledge was not spontaneously processed and therefore did not cause these action slips. In other words, we cannot not conclude with certainty that

A GOAL-DIRECTED ACCOUNT OF ACTION SLIPS

observing contingency processing in one task means that the same contingency was processed during the other task by making sure that the operating conditions were the same for both tasks. Nevertheless the fact that the observed pattern on expectancy trials matches the observed pattern of action slips on choice trials is strongly convergent with the idea that goal-directed processes can be the cause of the observed action slips.

One element that appears to be in favor of the idea that the contingency knowledge did drive the action slips, is the finding that the retrieval of old contingencies was positively correlated with the number of action slips. While the strong positive correlation aligns well with the goal-directed explanation that action slips stem from a retrieval of and reliance on old contingencies, there may be other factors contributing to this relation. It is, for example, possible that part of the relation can be attributed to a general ability to perform computerized tasks. Participants who made more mistakes on reversed choice trials (i.e., action slips) might have also made more mistakes on reversed expectancy trials (i.e., reporting old contingencies after a reversal) because they were generally worse at performing computerized tasks. However, the fact that the correlation between unreversed expectancy trials and reversed choice trials did not reach significance speaks against this alternative interpretation.

To conclude, people sometimes keep engaging in a behavior that used to be successful in the past, but which is knowingly no longer effective in the current situation. Contrary to the widely held view that such action slips occur when people act in the absence of expectations, the results of the current study suggest that people may perform action slips because they sometimes actually still expect that the behavior would result in the outcome even if only in a flash of a second.

Context paragraph

The research described in this article is embedded in a larger research project on goal-directed causes of suboptimal behaviors. Behaviors can be considered suboptimal if people do not act in line with their explicit goals, such as in addiction and psychopathologies. These types of behavior are typically explained with stimulus-driven or so-called habitual processes, because these processes do not take information about behavioral outcomes into account. In our research project, we outline and test alternative goal-directed explanations for seemingly suboptimal behaviors by considering factors such as hidden goals or, as in this study, the reliance on old outcome information. Factors such as these may make a behavior seem independent of goal pursuit, while it is actually directed at another hidden goal or driven by a no longer accurate expectation on how to reach a certain goal (for an overview see Moors et al., 2017). Identifying these alternative goal-directed causes for phenomena previously

A GOAL-DIRECTED ACCOUNT OF ACTION SLIPS

thought to be caused by stimulus-driven processes has wide implications for behavioral interventions.

A GOAL-DIRECTED ACCOUNT OF ACTION SLIPS

References

- Bouton, M. E. (2002). Context, ambiguity, and unlearning: Sources of relapse after behavioral extinction. *Biological Psychiatry*, 52(10), 976-986.
[https://doi.org/10.1016/s0006-3223\(02\)01546-9](https://doi.org/10.1016/s0006-3223(02)01546-9)
- Buabang, E. K., Boddez, Y., De Houwer, J., & Moors, A. (2021). Don't make a habit out of it: Impaired learning conditions can make goal-directed behavior seem habitual. *Motivation Science*. <https://doi.org/10.1037/mot0000218>
- De Houwer, J., Tanaka, A., Moors, A., & Tibboel, H. (2018). Kicking the habit: Why evidence for habits in humans might be overestimated. *Motivation Science*, 4(1), 50-59. <https://doi.org/10.1037/mot0000065>
- de Wit, S., Niry, D., Wariyar, R., Aitken, M. R., & Dickinson, A. (2007). Stimulus-outcome interactions during instrumental discrimination learning by rats and humans. *Journal of Experimental Psychology: Animal Behavior Processes*, 33(1), 1-11. <https://doi.org/10.1037/0097-7403.33.1.1>
- de Wit, S., Watson, P., Harsay, H. A., Cohen, M. X., Van de Vijver, I., & Ridderinkhof, K. R. (2012). Corticostriatal connectivity underlies individual differences in the balance between habitual and goal-directed action control. *Journal of Neuroscience*, 32(35), 12066-12075. <https://doi.org/10.1523/jneurosci.1088-12.2012>
- Dickinson, A. (1985). Actions and habits: The development of behavioural autonomy. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences*, 308(1135), 67-78. <https://doi.org/10.1098/rstb.1985.0010>
- Dickinson, A. (2016). Instrumental conditioning revisited: Updating dual-process theory. In *Associative learning and cognition. Homage to professor N. J. Mackintosh. In memoriam (1935-2015)* (pp. 177-196). Edicions Universitat Barcelona.
- Elsley, J. W., Van Ast, V. A., & Kindt, M. (2018). Human memory reconsolidation: A guiding framework and critical review of the evidence. *Psychological Bulletin*, 144(8), 797-848. <https://doi.org/10.1037/bul0000152>
- Ersche, K. D., Gillan, C. M., Jones, P. S., Williams, G. B., Ward, L. H., Luijten, M., De Wit, S., Sahakian, B. J., Bullmore, E. T., & Robbins, T. W. (2016). Carrots and sticks fail to change behavior in cocaine addiction. *Science*, 352(6292), 1468-1471.
<https://doi.org/10.1126/science.aaf3700>
- Evans, J. S., & Stanovich, K. E. (2013). Dual-process theories of higher cognition. *Perspectives on Psychological Science*, 8(3), 223-241.
<https://doi.org/10.1177/1745691612460685>

A GOAL-DIRECTED ACCOUNT OF ACTION SLIPS

- Gillan, C. M., Kosinski, M., Whelan, R., Phelps, E. A., & Daw, N. D. (2016). Characterizing a psychiatric symptom dimension related to deficits in goal-directed control. *eLife*, 5. <https://doi.org/10.7554/elife.11305>
- Hardwick, R. M., Forrence, A. D., Krakauer, J. W., & Haith, A. M. (2019). Time-dependent competition between goal-directed and habitual response preparation. *Nature Human Behaviour*, 3(12), 1252-1262. <https://doi.org/10.1038/s41562-019-0725-0>
- Hintzman, D. L. (1976). Repetition and memory. In G. H. Bower (Ed.), *Psychology of learning and motivation* (pp. 47-91). ACADEMIC PRESS.
- Hogarth, L. (2020). Addiction is driven by excessive goal-directed drug choice under negative affect: Translational critique of habit and compulsion theory. *Neuropsychopharmacology*, 45(5), 720-735. <https://doi.org/10.1038/s41386-020-0600-8>
- Hommel, B., & Wiers, R. W. (2017). Towards a unitary approach to human action control. *Trends in Cognitive Sciences*, 21(12), 940-949. <https://doi.org/10.1016/j.tics.2017.09.009>
- Kassambara, A. (2021). rstatix: Pipe-Friendly Framework for Basic Statistical Tests. R package version 0.7.0. <https://CRAN.R-project.org/package=rstatix>
- Kazak, A. E. (2018). Editorial: Journal article reporting standards. *American Psychologist*, 73(1), 1-2. <http://dx.doi.org/10.1037/amp0000263>
- Kruglanski, A. W., & Szumowska, E. (2020). Habitual behavior is goal-driven. *Perspectives on Psychological Science*, 15(5), 1256-1271. <https://doi.org/10.1177/1745691620917676>
- Moors, A. (2016). Automaticity: Componential, causal, and mechanistic explanations. *Annual Review of Psychology*, 67(1), 263-287. <https://doi.org/10.1146/annurev-psych-122414-033550>
- Moors, A., Boddez, Y., & De Houwer, J. (2017). The power of goal-directed processes in the causation of emotional and other actions. *Emotion Review*, 9(4), 310-318. <https://doi.org/10.1177/1754073916669595>
- Moors, A., & De Houwer, J. (2006). Problems with dividing the realm of processes. *Psychological Inquiry*, 17(3), 199-204.
- Papies, E. K., E. K., & Barsalou, L. W., L. W. (2015). Grounding desire and motivated behavior: A theoretical framework and empirical evidence. In *The psychology of desire* (pp. 36-60). Guilford Publications.

A GOAL-DIRECTED ACCOUNT OF ACTION SLIPS

- Peirce, J. W., Gray, J. R., Simpson, S., MacAskill, M. R., Höchenberger, R., Sogo, H., Kastman, E., Lindeløv, J. (2019). PsychoPy2: experiments in behavior made easy. *Behavior Research Methods*. 10.3758/s13428-018-01193-y
- Pessiglione, M., Schmidt, L., Draganski, B., Kalisch, R., Lau, H., Dolan, R. J., & Frith, C. D. (2007). How the brain translates money into force: A neuroimaging study of subliminal motivation. *Science*, 316(5826), 904-906.
<https://doi.org/10.1126/science.1140459>
- Pierce-Messick, Z., & Corbit, L. H. (2021). Problematic eating as an issue of habitual control. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, 110, 110294.
<https://doi.org/10.1016/j.pnpbp.2021.110294>
- R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>
- Tricomi, E., Balleine, B. W., & O'Doherty, J. P. (2009). A specific role for posterior dorsolateral striatum in human habit learning. *European Journal of Neuroscience*, 29(11), 2225-2232. <https://doi.org/10.1111/j.1460-9568.2009.06796.x>
- Vaghi, M. M., Cardinal, R. N., Apergis-Schoute, A. M., Fineberg, N. A., Sule, A., & Robbins, T. W. (2019). Action-outcome knowledge dissociates from behavior in obsessive-compulsive disorder following contingency degradation. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, 4(2), 200-209.
<https://doi.org/10.1016/j.bpsc.2018.09.014>
- Watson, P., Van Wingen, G., & De Wit, S. (2018). Conflicted between goal-directed and habitual control, an fMRI investigation. *eneuro*, 5(4), ENEURO.0240-18.2018.
<https://doi.org/10.1523/eneuro.0240-18.2018>
- Wood, W., & Rünger, D. (2016). Psychology of habit. *Annual Review of Psychology*, 67(1), 289-314. <https://doi.org/10.1146/annurev-psych-122414-033417>