**Supplementary material: computational modelling**

***Choice of models***

To quantify latent cognitive processes, various cognitive-computational models were fitted independently to participants’ choices (i.e. predictions of ‘left’ or ‘right’) and observed outcomes (i.e. prediction correct or incorrect) for the volatile and the cued block of the prediction task, respectively, separately for individuals with a diagnosis from the schizophrenia spectrum (SZ group) and those without psychiatric diagnoses (HC group). The models included a simple win-stay-loose-shift model (WSLS), four different Reinforcement Learning models (RL), and two variants of a Hidden Markov Model (HMM) – all chosen to allow for the fact that participants might employ different strategies of solving the task. According to the WSLS model (Worthy and Todd Maddox, 2014) individuals would change their predictions each time they received a negative feedback on the previous trial, but stay with their previous prediction if it turned out to be correct. Note that in the prediction task, feedback was not explicit but reflected in whether the outcome on a given trial indeed corresponded to a participant’s prediction. In contrast to the simplistic WSLS model, where choices (predictions) are merely based on the outcome of the previous trial, the RL models imply that prediction errors and choice values (i.e. values for left- and right-tilted) are integrated over a longer timescale. Four versions of RL models were tested, reflecting different assumptions of how this information is integrated. The Rescorla-Wagner model (RLRW; Rescorla and Wagner, 1972) assumes value updating only for the action chosen (prediction of the left or the right-tilted Gabor patch) on a given trial based on a trial’s prediction error weighted by a constant learning rate. In contrast, learning rates are allowed to differ for positive (prediction correct) vs. negative feedback (prediction incorrect) in the reward-punishment model (RLRP; den Ouden et al., 2013), accounting for the fact that participants might learn differently from both types of feedback. In the counterfactual updating model (RLCF; Gläscher et al., 2008), values for both available actions are updated concurrently. This might account better for the anti-correlated task structure, where the probabilities for the left- and the right-tilted stimuli are inversely related. Given that participants’ rate of learning might change during the course of a task block, an additional model based on Pearce and Hall (RLPH; Pearce and Hall, 1980)) was fitted with an adaptive learning rate. Lastly, the HMM (Schlagenhauf et al., 2014), a Bayesian inference model, assumes a higher-order representation of the task structure that accounts for the instability of the task environment. Here, participants are expected to choose ‘left’ or ‘right’ depending on whether they believe to be in a hidden state where either the left- or the right-tilted stimulus constitutes the majority. States beliefs are inferred and updated on each trial, depending on the history of choice-outcome pairs as well as the estimated transition probability , which is the assumed probability for the two states of ‘majority stimulus is left’ and ‘majority stimulus is right’ to change. Thus, indicates a participant’s perceived volatility of the task environment. To allow for the fact that positive (prediction correct) and negative (prediction incorrect) feedback may affect participant’s belief updating differently, a model of the HMM where they were allowed to differ (HMMRP) was contrasted against one where they were not (HMM).

All models are described in detail below. In addition, given that changes between risk conditions were announced in the cued task block, additional variants of all models were specified which incorporated choice value and state probability resets at each announced change point in this task block. In tables S1 – S4, these models are indicated by the suffix ‘\_reset’.

***Computational models***

*(I) Win-Stay-Loose-Shift model*

The Win-Stay-Lose-Shift model (WSLS; Worthy and Todd Maddox, 2014) assumes that if participants were rewarded for their choice on a given trial (i.e. their prediction of either ‘left’ or ‘right’ turned out to be correct), they continue to choose this option on the next trial (‘win-stay’). Similarly, if they were not rewarded (i.e. their prediction of either ‘left’ or ‘right’ turned out to be incorrect), they are expected to change their prediction and choose the other, previously non-selected option (‘lose-shift’). With the two possible actions *A* and *B*, the value of ‘staying’ with a choice after a ‘win’ is then calculated as:

Likewise, the value of ‘switching’ to the other choice option after a ‘loss’ is then:

*(II) Standard Rescorla-Wagner model*

In the Rescorla-Wagner reinforcement learning model (RLRW; Rescorla and Wagner, 1972), a constant learning rate drives the trial-wise value updates for the chosen option. For each trial , the value of the current choice is defined by the value and the prediction error (the difference between ‘reward’ and expected value) of the previous trial , weighted by the learning rate :

The prediction error, , is calculated as:

Importantly, ‘rewards’ in the prediction task were defined as correct predictions and assigned a value of +1. Since the experience of ‘rewards’ (i.e. correct predictions) and ‘punishments’ (i.e. incorrect predictions, assigned a value -1) might impact learning differently, a variant of the model was fitted where learning rates for rewards and punishments were allowed to differ (model: RLRP; den Ouden et al., 2013).

*(III) Counterfactual Reinforcement Learning model*

Given the anti-correlated task structure, values for both the chosen and the unchosen option may be updated simultaneously. To account for that, a counterfactual updating model (RLCF; Gläscher et al., 2008) was fitted to the data. The formula for the value update was the same as in (II), applied to both the chosen and the unchosen option, where a counterfactual prediction error was used for the unchosen option (uc):

*(IV) Pearce-Hall model*

In the Pearce-hall model, the learning rate was dynamic (RLPH; Pearce and Hall, 1980). Value updating was similar to the RLRW model (see II), but the learning rate varied across trials, depended on the previous prediction error:

*(V) Hidden Markov Model*

The Hidden Markov Model (HMM; Schlagenhauf et al., 2014) assumes that participants base their choices (i.e. predict either the left or the right Gabor patch) on their beliefs about the current task state, which in turn are modulated by the probability for those states to reverse (transition probability). Here, the different task states describe states where either the left stimulus (‘state L’) or the right patch is more common (‘state R’).

Participants are expected to infer the belief distribution over the different states from their observations of action-reward pairs (i.e. the combination of their prediction and the consequent ‘reward’, i.e. a correct or an incorrect prediction): . A participant’s estimation of such an action-outcome pair is then represented by the hidden state variable . In a transition matrix, the prior belief over the current state is calculated based on the posterior belief from the previous trial modulated by the transition probability , a free parameter between 0 and 1:

The probability of observing an outcome reflective of a given latent state depends further on the parameters and . Here, is the probability with which a reward (i.e. a positive feedback in terms of a correct prediction) indicates that the true latent state indeed corresponds to the selected chosen option, i.e. the prediction made. Conversely, is the probability with which a ‘punishment’ (i.e. a negative feedback in terms of an incorrect prediction) indicates that the true latent state does *not* correspond to the chosen option. The probability of observing a particular outcome given a particular state is then updated as:

Here, and were free parameters, initialized to lie between 0.5 and 1. Similar to the RLRP model (see II), this allowed for different effects of ‘rewards’ (positive feedback) and ‘punishments’ (negative feedback) in the model updating process. This version of the HMM is subsequently referred to as HMMRP. An additional version of the model (subsequently referred to as HMM) was fitted where positive and negative feedback were treated equally, with .

For both the HMMRP and the HMM, the probability of prior to any outcome observation is calculated for a given trial from the state transition probability (see above) and the posterior probability of :

After the outcome has been observed, the posterior probability of is updated based on the prior and the outcome :

*Softmax action selection*

Values were translated into choice probabilities for options *L* and *R* with a softmax action selector for all models (I) – (IV):

Here, the slope of the sigmoid function and the stochasticity (randomness) of the choices is determined by , the inverse temperature. For the HMM (see V), state probabilities were used in the softmax function in place of values. Further, in this model the inverse temperature parameter was not included in order to reduce non-identifiable parameter estimation.

***Model comparison***

A Hierarchical Bayesian Analysis (HBA; Gelman et al., 2013) was adopted from the hBayesDM package (Ahn et al., 2017) and implemented in the Stan language in R (RStan; Carpenter et al., 2017), to estimate model parameters. Models were fitted separately for both task blocks and both groups (SZ and HC). To compare modes regarding their goodness-of-fit to explain the observed data whilst accounting for model complexity, leave-one-out cross validation was conducted by using the log-likelihood evaluated at the posterior simulations. The results are reported with the leave-one-out information criterion (LOOIC; and corresponding effective number of parameters) in the tables below (S1 – S4), where lower LOOIC values indicate better model fit.

Table S1.

*First, volatile task block: Model fit for SZ (n = 30)*

|  |  |  |
| --- | --- | --- |
| **Model** | **LOOIC** | **no. of parameters** |
| WSLS | 5746 (25) | 1 |
| RLRW | 5077 (42) | 2 |
| RLCF | 5024 (47) | 2 |
| RLRP | 5012 (55) | 3 |
| RLPH | 5071 (58) | 4 |
| HMM | 4973 (48) | 2 |
| **HMMRP** | **4813 (71)** | **3** |

Table S2.

*First, volatile task block: Model fit for HC (n = 30)*

|  |  |  |
| --- | --- | --- |
| **Model** | **LOOIC** | **no. of parameters** |
| WSLS | 5262 (27) | 1 |
| RLRW | 4332 (53) | 2 |
| RLCF | 4275 (49) | 2 |
| RLRP | 4154 (56) | 3 |
| RLPH | 4332 (62) | 4 |
| HMM | 4120 (45) | 2 |
| **HMMRP** | **3923 (68)** | **3** |

Table S3.

*Second, cued task block: Model fit for SZ (n = 30)*

|  |  |  |
| --- | --- | --- |
| **Model** | **LOOIC** | **no. of parameters** |
| WSLS | 5669 (26) | 1 |
| RLRW | 5146 (42) | 2 |
| RLRW\_reset | 4406 (50) | 2 |
| RLCF | 4940 (51) | 2 |
| RLCF\_reset | 4369 (48) | 2 |
| RLRP | 5020 (61) | 3 |
| RLRP\_reset | 3948 (60) | 3 |
| RLPH | 5152 (51) | 4 |
| RLPH\_reset | 4406 (51) | 4 |
| HMM | 4902 (44) | 2 |
| HMM\_reset | 4315 (47) | 2 |
| HMMRP | 4254 (79) | 3 |
| **HMMRP\_reset** | **3826 (70)** | **3** |

Table S4.

*Second, cued task block: Model fit for HC (n = 30)*

|  |  |  |
| --- | --- | --- |
| **Model** | **LOOIC** | **no. of parameters** |
| WSLS | 5627 (27) | 1 |
| RLRW | 5237 (40) | 2 |
| RLRW\_reset | 4907 (44) | 2 |
| RLCF | 4384 (43) | 2 |
| RLCF\_reset | 4299 (42) | 2 |
| RLRP | 5133 (62) | 3 |
| RLRP\_reset | 4240 (53) | 3 |
| RLPH | 5241 (61) | 4 |
| RLPH\_reset | 4378 (54) | 4 |
| HMM | 4839 (44) | 2 |
| HMM\_reset | 4225 (43) | 2 |
| HMMRP | 4502 (63) | 3 |
| **HMMRP\_reset** | **4088 (69)** | **3** |

***Model recovery***

We conducted model recovery analyses (Wilson and Collins, 2019; Crawley, Zhang et al., 2020) using simulated data from 40 participants, following the exact task structure of the experiment. Simulated data were simulated from each candidate model using the posterior parameter from the healthy control group in the volatile block, and then fitted to all candidate models with the same hierarchical Bayesian model fitting procedure described above. Results showed that all candidate models can be properly identified and recovered (Fig. S1).

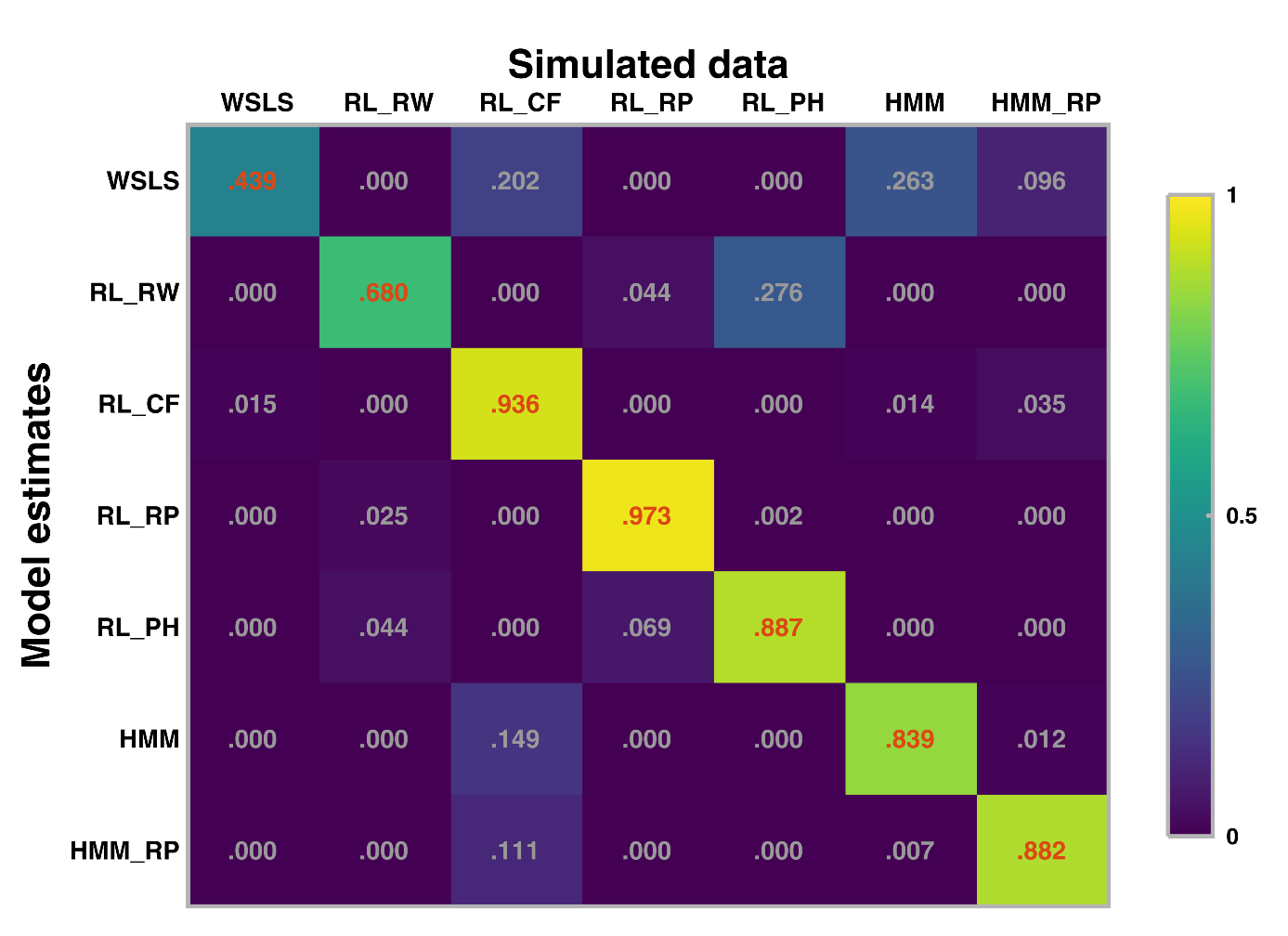
****

Fig. S1 Model recovery. Data from 40 synthetic participants were simulated with each of our candidate models. Color indicates model weights calculated with Bayesian model averaging using Bayesian bootstrap (higher model weight value indicates higher probability of the candidate model to have generated the observed data).

**Supplementary material: additional analyses**

***Choice switches and feedback***

To test for the effect of feedback on the proportion of switches, proportion of switches after positive (correct prediction) and after negative feedback (incorrect prediction), respectively, was calculated for each task block (volatile and cued). A linear mixed-effects model was applied to test to what extent proportion of switches differed by type of feedback on the previous trial (positive vs. negative), block (cued vs. volatile), group (SZ vs. HC), and their respective interactions. Random factors were defined as random intercepts for feedback type nested within blocks nested within participants. Results revealed a negative main effect of positive feedback, indicating that participants switched less after their prediction was correct (*b* = -0.36, *t* = -9.33, *p* < .001). However, this effect did not differ by block or group (see Table S5).

Table S5.

*Linear mixed-effects model results for proportion of choice switches*

|  |  |  |  |
| --- | --- | --- | --- |
| IV | *b* | *t* | *p* |
| Block | -0.03 | -0.79 | .435 |
| Feedback | -0.36 | -9.33 | <.001 |
| Group | 0.04 | 0.78 | .440 |
| Block\*Feedback | 0.05 | 0.92 | .360 |
| Block\*Group | -0.06 | -1.11 | .270 |
| Feedback\*Group | 0.03 | 0.53 | .594 |
| Block\*Feedback\*Group | -0.03 | -0.41 | .686 |

*Notes:* IV = independent variable, Block = contrast of the second, cued task block to the first, volatile task block; Feedback = contrast of positive vs. negative feedback; Group = contrast of the SZ (schizophrenia) to the HC (controls) group

***Entropy (uncertainty) and behavior***

To test whether entropy on trial *t* would predict reaction times (RTs) on trial *t+1*, a linear mixed-effects model was constructed, including a random intercept per participant. For each participant, RTs larger than the participant-specific mean + 2.5\* the participant-specific standard deviation were set to NA and the respective trials removed from analyses. Furthermore, reaction time values of the first trials of each block and participant were omitted, as were entropy values for the respective last trials. Additionally, these analyses excluded the 160 trials of the second task block for the participant who aborted after block one, and the 160 trials of the first task block for the participant where the winning model could not be fitted to this block’s data.

To approximate a normal distribution of residuals, reaction times were cube root transformed (as the typically applied log transformation produced inferior results).

Results showed that higher entropy on trial *t* was associated with significantly prolonged reaction times on trial *t+1* (*b* = 0.02, *t* = 4.15, *p* < .001). This finding is consistent with the idea that higher uncertainty in decision-making paradigms slows down reaction times (see e.g. Volz et al., 2005).

***Symptoms and behavior***

In an exploratory analysis, measures of behavioral task performance (accuracy and choice switches) were correlated with negative and positive symptom severity (assessed with the PANSS). For the most part, this revealed no significant associations for performance assessed within, or averaged across the two task blocks (see Table S6). The only exception was a positive correlation between negative symptoms and accuracy within the cued task block ( = 0.41, *p* = .026). However, this effect did not survive Bonferroni correction for multiple comparisons (αadj = .004).

Table S6.

*Spearman correlations between task performance and symptoms*

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | volatile block | | | |  | cued block | | | |  | block average | | | |
|  | PANSS-P | | PANSS-NvdGaag | |  | PANSS-P | | PANSS-NvdGaag | |  | PANSS-P | | PANSS-NvdGaag | |
|  |  | *p* |  | *p* |  |  | *p* |  | *p* |  |  | *p* |  | *p* |
| accuracy | -0.09 | .630 | 0.18 | .344 |  | -0.10 | .604 | 0.41 | .026 |  | -0.10 | .598 | 0.28 | .122 |
| switches | 0.08 | .670 | -0.15 | .428 |  | -0.08 | .678 | -0.23 | .219 |  | 0.08 | .675 | -0.18 | .334 |

***Pupil dilation and latent variables: extended models by task block***

The linear mixed-effects models linking pupil dilation to the latent variables entropy and Bayesian surprise, respectively, were extended to include task block. This revealed no significant block related effects, while the main effect of entropy and the entropy by group interaction remained largely unchanged. However, the main effect of Bayesian surprise on pupil dilation was no longer significant (see Table S7).

Table S7.

*Linear mixed-effects model results for pupil dilation*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Latent variable (HMMRP) | IV | *b* | *t* | *p* | *R2M* | *R2C* |
| ***Entropy*** |  |  |  |  | 0.01 | 0.19 |
|  | Group | -0.05 | -1.32 | .193 |  |  |
|  | Entropy | 0.03 | 5.63 | <.001 |  |  |
|  | Block | 0.00 | 0.02 | .986 |  |  |
|  | Group\*Entropy | -0.02 | -2.11 | .035 |  |  |
|  | Group\*Block | 0.01 | 0.26 | .796 |  |  |
|  | Entropy\*Block | -0.01 | -1.59 | .111 |  |  |
|  | Group\*Entropy\*Block | 0.00 | 0.43 | .668 |  |  |
| ***Bayesian surprise*** |  |  |  |  | 0.01 | 0.19 |
|  | Group | -0.05 | -1.33 | .190 |  |  |
|  | Bayesian surprise | 0.01 | 1.76 | .079 |  |  |
|  | Block | 0.00 | 0.01 | .995 |  |  |
|  | Group\*Bayesian surprise | 0.01 | 0.68 | .495 |  |  |
|  | Group\*Block | 0.01 | 0.27 | .788 |  |  |
|  | Bayesian surprise\*Block | 0.00 | -0.32 | .747 |  |  |
|  | Group\*Bayesian surprise\*Block | -0.01 | -1.39 | .165 |  |  |

*Notes:* Models were fitted separately for entropy and Bayesian surprise as the predictive latent HMMRP variable, both were z-scored per task block and participant; pupil dilation = square root transformed maximum baseline-corrected pupil dilation during outcome presentation (based on the z-scored pupil trace per participant and block); IV = independent variable; Group = contrast of the SZ (schizophrenia) to the HC (controls) group; random effects were specified for block nested within participant; *R2m* = marginal *R2*, i.e*.* proportion of variance explained by the fixed effects alone; *R2c* = conditional *R2*, i.e. proportion of variance explained by both the fixed and random effects (*R2m* and *R2c* based on Nakagawa and Schielzeth, 2013).

**References (suppl.):**

Ahn, W.Y., Haines, N., Zhang, L., 2017. Revealing Neurocomputational Mechanisms of Reinforcement Learning and Decision-Making With the hBayesDM Package. Comput Psychiatr 1, 24-57.

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., . . . Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software, 76*(1).

Crawley, D., Zhang, L., Jones, E. J., Ahmad, J., Oakley, B., San José Cáceres, A., ... & EU-AIMS LEAP group, 2020. Modeling flexible behavior in childhood to adulthood shows age-dependent learning mechanisms and less optimal learning in autism in each age group. *PLoS Biology*, *18*(10), e3000908.

den Ouden, Hanneke E.M., Daw, Nathaniel D., Fernandez, G., Elshout, Joris A., Rijpkema, M., Hoogman, M., Franke, B., Cools, R., 2013. Dissociable Effects of Dopamine and Serotonin on Reversal Learning. Neuron 80(4), 1090-1100.

Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A., Rubin, D.B., 2013. Bayesian data analysis. CRC press.

Gläscher, J., Hampton, A.N., O'Doherty, J.P., 2008. Determining a Role for Ventromedial Prefrontal Cortex in Encoding Action-Based Value Signals During Reward-Related Decision Making. Cerebral Cortex 19(2), 483-495.

Pearce, J.M., Hall, G., 1980. A model for Pavlovian learning: Variations in the effectiveness of conditioned but not of unconditioned stimuli. Psychological Review 87(6), 532-552.

Rescorla, R.A., Wagner, A.R., 1972. A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement., in: Black, A.H., Prokasy, W.F. (Eds.), Classical Conditioning II: Current Research and Theory. Appleton Century Crofts, New York, pp. 64–99.

Schlagenhauf, F., Huys, Q.J.M., Deserno, L., Rapp, M.A., Beck, A., Heinze, H.-J., Dolan, R., Heinz, A., 2014. Striatal dysfunction during reversal learning in unmedicated schizophrenia patients. NeuroImage 89, 171-180.

Wilson, R. C., & Collins, A. G., 2019. Ten simple rules for the computational modeling of behavioral data. *elife*, *8*, e49547.

Volz, K.G., Schubotz, R.I., Cramon, D.Y.v., 2005. Variants of uncertainty in decision-making and their neural correlates. Brain Research Bulletin 67(5), 403-412.

Worthy, D.A., Todd Maddox, W., 2014. A comparison model of reinforcement-learning and win-stay-lose-shift decision-making processes: A tribute to W.K. Estes. Journal of Mathematical Psychology 59, 41-49.