Title: Species-conserved mechanisms of cognitive flexibility in complex environments

Authors: Florian Bähner[1,2*], Tzvetan Popov[2,4], Selina Hermann[1,2], Nico Boehme[1,2], Tom Merten[1,2], Hélène Zingone[1,2], Georgia Koppe[3], Andreas Meyer-Lindenberg[2†], Hazem Toutounji[5†], Daniel Durstewitz[3,6,7†]

Affiliations:

[1]RG Behavioral Physiology in Psychiatry, Central Institute of Mental Health, Heidelberg University, Medical Faculty Mannheim; Mannheim, Germany.

[2]Department of Psychiatry and Psychotherapy, Central Institute of Mental Health, Heidelberg University, Medical Faculty Mannheim; Mannheim, Germany.

[3]Department of Theoretical Neuroscience, Central Institute of Mental Health, Heidelberg University, Medical Faculty Mannheim; Mannheim, Germany.

[4]Department of Psychology, University of Zurich; Zurich, Switzerland.

[5]School of Psychology, University of Nottingham; Nottingham, UK.

[6]Interdisciplinary Center for Scientific Computing, Heidelberg University; Heidelberg, Germany.

[7]Faculty of Physics and Astronomy, Heidelberg University; Heidelberg, Germany.

*Corresponding author. Email: florian.baehner@zi-mannheim.de.

†These authors contributed equally to this work.

Abstract: Flexible decision making in complex environments is a hallmark of intelligent behavior but the underlying learning mechanisms and neural computations remain elusive. Through a combination of behavioral, computational and electrophysiological analysis of a novel multidimensional rule-learning paradigm, we show that both rats and humans sequentially probe different behavioral strategies to infer the task rule, rather than learning all possible mappings between environmental cues and actions as current theoretical formulations suppose. This species-conserved process reduces task dimensionality and explains both observed sudden behavioral transitions and positive transfer effects. Behavioral strategies are represented by rat prefrontal activity and strategy-related variables can be decoded from magnetoencephalography signals in human prefrontal cortex. These mechanistic findings provide a foundation for the translational investigation of impaired cognitive flexibility.

Main Text:

How do we make choices in complex, changing and novel situations? Cognitive flexibility is the ability to respond to the environment in adaptive ways and deficits in this cognitive domain are observed in several major neuropsychiatric disorders (*1*). However, how correct task rules are identified in ever-changing contexts and how this information is encoded in neural activity is not fully understood (*2–5*). A powerful theoretical framework in this area is reinforcement learning (RL) which explains how choice values are learned to guide decisions and maximize reward (*6*). The discovery of neural substrates of RL parameters such as action values (i.e., the expected reward from taking that action) and reward prediction errors (i.e., the difference between actual and expected reward) highlight their relevance for neuroscience (*2*, *7*). A core assumption of classic RL models is that subjects learn to map each state of the environment to the maximum-value action (*6*). However, the real world is multidimensional, where many sources of information, such as sensory cues, reward history, and working memory of past choices combine to form each state. This results in an exponential growth in the number of states and thus of state-action mappings (i.e., the curse of dimensionality), which would translate into slow and gradual learning (*8*). Consequently, basic RL models struggle to explain hallmarks of flexible behavior such as sudden transitions in performance, rapid learning in complex environments or faster learning with prior experience (*2*, *5*, *9–12*). These shortcomings can be resolved by assuming that, rather than learning general mappings of states to actions, animals learn and test hypothetical rules underlying these mappings, such as *"don't press any lever paired with a bright light, regardless of lever location"*. This reduces dimensionality dramatically since every cue becomes an instance of an abstract task feature, where its position or its relationship with other task features can be ignored. Animals can subsequently test different hypothetical rules against

environmental evidence until they successfully identify the experimenter-defined rule. We call these hypothetical rules *behavioral strategies* in the following to distinguish them from task rules.

Flexible human behavior can indeed be explained by several computational models that are compatible with the above framework, including hierarchical and attention-modulated RL models (*5*, *8*, *10*). However, it is unknown how rats infer task rules in complex environments and whether some of the underlying behavioral and neural mechanisms are conserved across species. Pioneering work has shown that in some contexts, rats also follow behavioral strategies like *win-stay-lose-shift* or *spontaneous alternation* (*13*, *14*). But these findings have not been generalized to cognitive flexibility or empirically contrasted to the prevailing view of classical RL models. Novel behavioral analysis tools in rodents may help to disentangle such complex cognitive processes because they allow to track behavioral parameters which were hard to measure before (e.g., movement patterns) (*15*, *16*). This may help to constrain the space of possible learning mechanisms subjects use to infer task rules (i.e., when several models could account for the same experimental phenomenon equally well) and is a prerequisite to understand which computations are carried out by brain circuits (*15*, *16*).

To address how rodents and humans learn new rules, and whether the underlying mechanisms are preserved across species, we developed a new translational rule-learning task. We used a combination of behavioral, computational, and electrophysiological methods to show that both rats and humans deploy top-down attention to specific task features and follow low-dimensional behavioral strategies to test each feature for relevance rather than learning all possible mappings between environmental states and actions simultaneously. We further emphasize the cross-species link by showing that strategy-related parameters can be decoded from both rat multiple-

single unit and human magnetoencephalography (MEG) recordings in homologous prefrontal regions.

Results

Rats follow behavioral strategies to infer task rules in a novel multidimensional rule-learning task

We developed a novel dual-choice multidimensional rule-learning task to evaluate the hypothesis that rats identify experimenter-defined rules by sequentially testing different low-dimensional behavioral strategies (Fig. 1). A loudspeaker and a cue light were positioned above each lever and in each trial one auditory and visual cue were presented to model multisensory input (*17*). Rats were not pre-trained on any rule and had to learn one out of eight possible rules with deterministic reward feedback: go-click (i.e., follow the auditory cue in the presence of a distracting visual cue), go-silent, go-light, go-dark, go-right, go-left, alternate or win-stay-lose-shift/WSLS (N=125 rats). Thus, four task features could be relevant for reward in every trial: the visual and auditory cues as well as outcome and choice histories. Rats were able to learn each rule, albeit at different learning rates (Table S3). We used change point analysis and behavioral modelling to identify when and how abrupt performance increases during rule learning. These transitions are best characterized as a step change (median (Q_1-Q_3) of 10-90 % rise time: 1 (1-1) trials, N=117).

The null hypothesis based on the classic RL view assumes that rats learn the correct action (e.g., right or left lever press) for each of the 2^4=16 cue combinations (Fig. 1B). We propose an alternative hypothesis where rats use up to eight low-dimensional behavioral strategies (*go-*

light/go-dark, go-click/go-silent, go-right/go-left, alternate or win-stay-lose-shift) to sample potential task-relevant information (i.e., significance of sensory cues, past choices and outcome history).

To test this, we developed a strategy detection algorithm that measures whether a particular strategy is more likely than chance or any other strategy within every possible trial window (Fig. S1). The algorithm identifies trial sequences within each session in which a particular strategy is the winning strategy. In each task rule, animals follow 7 (6-8) out of 8 possible strategies (Table S3). Interestingly, we replicated this finding in a separate cohort (N=19) that was presented with random reward feedback (i.e., none of the four task features were predictive of reward). We defined an empirical measure for learning based on the comparison between the length of the correct strategy sequence and the distribution of sequence lengths of that same strategy in the cohort that received random reward (e.g., we assume learning of the task rule go-silent has occurred if the sequence length of the *go-silent* strategy is sufficiently larger than its typical length in the random reward cohort). The trial at which the empirical learning criterion was reached significantly correlated with the performance change point (Fig. 1D). Moreover, strategies also provide an explanation why performance increases are sudden because the length of correct strategy sequences increases abruptly when rats reach the learning criterion (sequence length before reaching criterion vs. learning criterion: 10 (8-15) vs. 48 (31-83.5) trials; $p=6.12*10^{-21}$, Wilcoxon matched pairs test, N=117). Furthermore, several strategy-associated variables are predictive of learning (Fig. 1E, S2, Table S4). We estimated a linear regression model in which the empirical learning criterion is predicted by the trial number at which the animal followed the correct strategy for the first time, the number of correct strategy sequences before reaching the empirical learning criterion, a perseveration measure (*18*) at the strategy

level, and a categorical regressor indicating the task rule. While not all trials could be assigned to a specific strategy (mean ± SEM: 63.84 ± 1.24 % of trials), adding this measure as a predictor variable to our regression model did not increase R^2 (78.9 %, p=0.43 for this regressor). We found comparable results in rats performing an established operant two-rule set-shifting task (*18*) which also follow behavioral strategies that go along with sudden transitions in performance (Fig. S3). In sum, our data suggest that to infer task rules, rodents test low-dimensional behavioral strategies rather than learning high-dimensional mappings between environmental states and actions.

How rats perceive the world: selective attention mechanisms guide choice and learning

Selective attention focused on different task features is a candidate mechanism for guiding behavioral responses (i.e., the rat focuses on one general task feature at a time instead of a combination of low-level cues), which can explain why rats follow strategies for many trials, even when reward is random. Selective attention also offers a means for reducing dimensionality in complex environments to a manageable subset of task features that can be tested for relevance (*2*, *8*). To test the role of attention in our paradigm, we applied machine learning techniques to screen videotaped sessions for behavioral markers of attention in naïve rats that performed multiple rule switches. We found a strong orientation reaction at cue onset that was influenced by the current strategy. More specifically, we found that the onset and offset trials of each strategy sequence are locked to abrupt changes in head direction at cue onset towards and away from the task feature that is relevant for that strategy (Fig. 2A). This allowed us to obtain a trial-by-trial measure of the rat's attention to each task feature at the time of cue presentation, which we term "attention-at-choice" score. Strategy-specific abrupt changes in attention to a corresponding task

feature were found for all eight strategies (Fig. S4). A similar, but feature-independent, orientation reaction was observed during reward feedback, which we term “attention-at-reward” score. These findings were replicated in the random reward cohort (N=19) where animals had no incentive to follow a specific strategy (Fig. S5). Rats thus follow low-dimensional behavioral strategies even when there is no behavioral advantage in doing so.

To assess whether the two sets of attention scores are relevant for multidimensional rule learning, we first tested whether changes in attention across trials are predictive of learning. Abrupt transitions in attention scores were detected in a large proportion of learned rules using change point analysis (Fig. S6; attention-at-choice: 87.2 %, attention-at-reward: 67 %, N=29 rats learning four different rules). We then estimated a linear regression model to show that the empirical learning criterion is predicted by both attentional change points. Thus, strategy-dependent attention scores based on head movements predict learning with high accuracy (Fig. 2B, Table S4). Second, we incorporated the attention scores into a RL model in different ways to test whether attention modulates behavior during choice alone (attention-at-choice AC model), learning (or reward feedback) alone (attention-at-learning AL model), or both (ACL)(*2*). Attention scores bias learning and choice in the ACL model such that the task feature related to the current strategy has a privileged position. A fourth model where attention scores neither bias choice nor learning (uniform-attention/UA model) corresponds to the null hypothesis that rats learn all possible mappings between the high-dimensional state space and actions. Fitting the parameters of the four RL models to experimental data, followed by cross-validated model comparison revealed that the ACL model was the best at predicting animal behavior in held-out data in 27/29 rats (Fig. 2C)(*17*).

In sum, our data suggest that strategy-specific selective attention has a strong influence on rapid learning by reducing task dimensionality in complex environments, since it allows animals to focus on one task feature at a time instead of dividing attention between the specific cues at every trial.

How rats choose what to do next: strategies are hypothetical rules that are selected based on their value

Attention-modulated RL models highlight the relevance of low-dimensional strategies for behavior but cannot explain how these strategies are selected. Our data show that the current strategy constrains action selection in response to a stimulus and this effect can be explained by the corresponding attentional focus we included as a parameter in our RL models. This suggests a hierarchical action selection process in the sense that strategies are hypothetical rules operating as high-level, abstract actions that are selected first (e.g., *go-light*) and then map specific stimuli (e.g., cue light left) to low-level actions (e.g., left lever press). But how do rats decide which strategy to execute next? In human rule learning, hierarchical action selection makes several empirically testable predictions (*5*) that also apply to our paradigm (Fig. 3A, S7). A core prediction is that strategies are selected based on their value (i.e., action values at the higher hierarchical level). We tested this assumption using two approaches.

First, strategies that have been previously reinforced should have a higher value and thus be chosen earlier in a novel context. We therefore compared strategy selection between two cohorts of rats receiving random reward feedback that were either naïve or previously learned three different experimenter-defined rules (go-dark→ place→ alternate). Indeed, the strategies *go-dark* and *alternate* were chosen earlier in the cohort with prior learning experience (Fig. 3B).

Second, hierarchical models predict positive transfer effects with prior learning experience because strategies are re-selected as a whole instead of relearning all possible state-action mappings (*5*). Indeed, experienced rats that learned the task rules alternate and go-silent, respectively, both probed the correct strategy earlier (first trial with correct strategy sequence *alternate*: 283 ± 37.6 vs. 495.5 ± 65.5, p=0.016; *go-silent*: 638.3 ± 87.4 vs. 1111 ± 127.7, $p=2*10^{-3}$, unpaired t-test) and reached the learning criterion faster (Fig. 3C) as compared to naïve animals learning the respective rules.

The rat prefrontal cortex encodes strategies as abstract rules

Research on hierarchical rule learning in humans suggests that strategies as abstract, hypothetical rules are represented in prefrontal brain activity (*5*). Indeed, we found that behavioral strategies could be decoded from the activity of prelimbic neurons in rats performing rule switches in either the novel multidimensional rule-learning paradigm or a conventional set-shifting task (Fig. 4; 62 sessions from nine rats, 34 (21-38) neurons per session). We compared decoding accuracy for each pair of strategies within an experimental session which was above chance in all cases (83.2 (74.4-89.9) %, 105 pairs, bin size 3 s). Decoding accuracy was highest *prior* to cue onset (-0.9 (-1.5 - +0.75) s) and significantly decreased after cue presentation (Fig. 4B; comparison 3s before vs. after cue onset: $p=1.9*10^{-9}$, Wilcoxon signed rank test). This adds to the evidence that action selection is not tied to specific cues but is rather a top-down influence. Finally, hierarchical RL models predict that strategies should be abstract “high-level” actions and the prefrontal representation of strategies should thus be invariant with respect to the “low-level” motor response in a given trial (*19*). For strategy pairs with sufficient sequence length, we trained a classifier on trials with lever presses on one side and compared decoding accuracy when tested on

trials with lever presses on either the same or opposite side. Decoding accuracy in the 3s before cue onset was lower if sides were different in training and test conditions but still much higher than chance (Fig. 4C). A systematic analysis of additional factors influencing decoding accuracy corroborated those findings (*17*). In sum, our results show that behavioral strategies are encoded in prefrontal activity and this representation has features of an abstract rule (i.e., it is not tied to specific cues or a simple motor action).

Behavioral and MEG findings in humans echo insights from the rodent model

Although our rodent task is inspired by human research, it is unknown whether humans show similar behavioral and neural processes in this experimental context. We therefore analyzed the data of 31 healthy adults that solved an adapted version of the multidimensional rule-learning task during MEG recordings (Fig. 5). Although the number of trials was necessarily lower in the human task version due to limits on MEG session duration, subjects tested most of the behavioral strategies (6 (5-6) out of 8 strategies) and were able to learn 58.1 % of the rules. Across rules, the onset of the first correct strategy sequence correlated with both increases in performance (Fig. 5B) and decreases in reaction time (Fig. 5C). Interestingly, the onset of the first correct strategy sequence occurred before the decrease in reaction time (trial 29 (15-63) vs. 51 (40.5-86.5), $p=1.2*10^{-3}$, Wilcoxon matched pairs test). Faster responses, indicating more confidence about their choices, thus occurred only after subjects tested the correct strategy for the first time. This supports the notion that humans indeed test strategies to identify the correct task rule.

At the neural level, we decoded the relevant task feature tested by a strategy to be able to perform a whole-brain second level analysis (i.e., not all subjects showed the same strategies but sampled the same task features: e.g., *go-click/go-silent* share the same task feature). We used multivariate

pattern searchlight analysis at the source level and found significantly above-chance decoding accuracy in a cortical network prominently including lateral and medial prefrontal as well as cingulate sources (Fig. 5D). As for animals, decoding accuracy was already above chance prior to cue onset (Fig. S8), indicating that attention allocation in humans is also driven by top-down processes related to strategy learning rather than being tied to specific cues. Subjects also completed a neuropsychological test battery, the CANTAB, to link task results to established markers of executive function (Fig. S9). In sum, behavioral and neural findings in humans replicate those in rodents, suggesting that both species address novel task challenges by probing different low-dimensional behavioral strategies in contrast to solving the much higher-dimensional problem of learning all possible cue-action pairs.

Discussion

We developed a novel translational rule-learning task to show that rats are able to infer task rules in multidimensional environments similar to what has been described in humans. We used a combination of behavioral, computational and electrophysiological methods to show that both rats and humans sequentially follow different hypothetical rules - which we call behavioral strategies - to identify the current task rule. The use of low-dimensional behavioral strategies to test task features for relevance contrasts with the prevailing view of RL that predicts independent learning of all possible mappings between environmental states and actions. Strategy-based learning can explain salient experimental findings like sudden transitions in performance and positive transfer effects. Furthermore, strategy-related parameters could be decoded from both rat multiple-single unit and human MEG recordings in functionally similar prefrontal regions. These

findings indicate that learning an abstract task structure based on behavioral strategies is a species-conserved mechanism for cognitive flexibility.

Computational models can explain how humans infer task rules in complex environments (*5*, *8*, *10*). In rodents, however, these learning mechanisms are largely unknown in spite of a wealth of information about task-related neural firing patterns, necessary brain regions and involved neurotransmitters in the context of rule-switching paradigms (*1*, *18*). Since most learning problems can in principle be handled by many different computational mechanisms (*6*) and it is largely an open question which aspects of cognitive flexibility are conserved across species, our novel rule-learning task was designed to reveal how rats infer relevant task structure by directly testing predictions made by different learning models.

An important class of models suggests that attention mechanisms can reduce the perceived complexity of a multidimensional environment by focusing the animal's cognitive resources on a small subset of features, thus allowing for efficient sampling of a simplified state space (*2*, *8*). In line with previous literature (*20*), our video analysis showed head movements related to cue onset and reward feedback and these orientation reactions were strongly modulated by attention. More importantly, abrupt shifts of attention were strategy-specific and similar to what has recently been described in humans, attention-modulated RL models (*2*) best explained task response if empirical measures for attention at both choice and learning were included in the model.

A challenge for attention-modulated RL models is to predict how strategy and associated attention focus are selected (*2*, *8*). In order to understand the mechanisms of strategy selection in rats, we tested experimental predictions of hierarchical RL models, an influential model of human cognitive flexibility (*5*).

These models make core experimental predictions that we tested and confirmed using both behavioral and neural data. First, we showed that strategy selection depends on their (learned) value, i.e., previously successful strategies are preferably selected in a novel context. This facilitates transfer to novel contexts because previously acquired strategies are reselected as a whole instead of relearning of all possible state-action mappings (*5*). Moreover, our data not only show positive transfer effects: strategy perseveration (i.e., negative transfer effects) - which is a strong predictor of learning in our task - is also qualitative evidence for hierarchical processing (*19*). Second, according to hierarchical RL models and other approaches, hypothetical rules should be represented in the human lateral PFC (*5*, *8*), a prediction that has already received experimental support (*21*). We showed that behavioral strategies could be decoded from multiple single-unit recordings of the rat prelimbic PFC, the rodent prefrontal region that is functionally closest to the primate lateral PFC (*1*, *22*) (but see (*23*) for a different view). In the human task version, decoding accuracy was highest for a set of prefronto-parietal regions including dorsolateral PFC. Our results indicate that this neural representation has features of an abstract rule. From a translational view, these findings are especially interesting given that prefrontal neurons in homologous regions have consistently been found to encode task rules in the steady state (i.e., after the task rule has been reliably identified) (*1*, *9*, *24*).

In sum, through a combination of a novel paradigm with computational and electrophysiological analysis, we provide evidence for a species-conserved rule learning mechanism that can explain salient experimental phenomena which are challenging for the prevailing models of RL. Our approach allows us to investigate the “black box” of covert hypothesis testing in rodents and offers a unifying link between different concepts related to cognitive flexibility. While it is clear that human learning in complex environments goes far beyond what we have found in rats (*10*),

we identified a conserved set of computational and neural mechanisms that provides a foundation for the mechanistic study of cognitive inflexibility in animal models of neuropsychiatric disease. In turn, insights from the rodent model may help to better understand impaired cognitive flexibility in clinical populations by combining behavioral, imaging and computational approaches (*6*).

References and Notes

1. N. P. Friedman, T. W. Robbins, The role of prefrontal cortex in cognitive control and executive function. *Neuropsychopharmacology*. **47**, 72–89 (2022).

2. Y. C. Leong, A. Radulescu, R. Daniel, V. DeWoskin, Y. Niv, Dynamic Interaction between Reinforcement Learning and Attention in Multidimensional Environments. *Neuron*. **93**, 451–463 (2017).

3. H. Malagon-Vina, S. Ciocchi, J. Passecker, G. Dorffner, T. Klausberger, Fluid network dynamics in the prefrontal cortex during multiple strategy switching. *Nat. Commun.* **9**, 309 (2018).

4. N. J. Powell, A. D. Redish, Representational changes of latent strategies in rat medial prefrontal cortex precede changes in behaviour. *Nat. Commun.* **7**, 12830 (2016).

5. A. G. E. Collins, M. J. Frank, Cognitive control over learning: creating, clustering, and generalizing task-set structure. *Psychol. Rev.* **120**, 190–229 (2013).

6. Q. J. M. Huys, M. Browning, M. P. Paulus, M. J. Frank, Advances in the computational understanding of mental illness. *Neuropsychopharmacology*. **46**, 3–19 (2021).

7. W. Schultz, W. R. Stauffer, A. Lak, The phasic dopamine signal maturing: from reward via behavioural activation to formal economic utility. *Curr. Opin. Neurobiol.* **43**, 139–148 (2017).

8. A. Radulescu, Y. Niv, I. Ballard, Holistic Reinforcement Learning: The Role of Structure and Attention. *Trends Cogn. Sci.* **23**, 278–292 (2019).

9. D. Durstewitz, N. M. Vittoz, S. B. Floresco, J. K. Seamans, Abrupt transitions between

prefrontal neural ensemble states accompany behavioral transitions during rule learning. *Neuron*. **66**, 438–48 (2010).

10. M. Donoso, A. G. E. Collins, E. Koechlin, Foundations of human reasoning in the prefrontal cortex. *Science (80-.).* **344**, 1481–1486 (2014).

11. R. Bartolo, B. B. Averbeck, Prefrontal Cortex Predicts State Switches during Reversal Learning. *Neuron*. **106**, 1044-1054.e4 (2020).

12. M. P. Karlsson, D. G. R. Tervo, A. Y. Karpova, Network resets in medial prefrontal cortex mark the onset of behavioral uncertainty. *Science*. **338**, 135–9 (2012).

13. J. L. Evenden, T. W. Robbins, Win-Stay Behaviour in the Rat. *Q. J. Exp. Psychol. Sect. B.* **36**, 1–26 (1984).

14. W. N. Dember, H. Fowler, Spontaneous alternation behavior. *Psychol. Bull.* **55**, 412–428 (1958).

15. J. W. Krakauer, A. A. Ghazanfar, A. Gomez-Marin, M. A. MacIver, D. Poeppel, Neuroscience Needs Behavior: Correcting a Reductionist Bias. *Neuron*. **93**, 480–490 (2017).

16. T. D. Pereira, J. W. Shaevitz, M. Murthy, Quantifying behavior to understand the brain. *Nat. Neurosci.* **23**, 1537–1549 (2020).

17. Please see *Supplementary Materials* for detailed methods as well as supplementary results supporting findings reported in the main text.

18. S. B. Floresco, A. E. Block, M. T. L. Tse, Inactivation of the medial prefrontal cortex of the rat impairs strategy set-shifting, but not reversal learning, using a novel, automated

procedure. *Behav. Brain Res.* **190**, 85–96 (2008).

19. M. K. Eckstein, A. G. E. Collins, Computational evidence for hierarchically structured reinforcement learning in humans. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 29381–29389 (2020).

20. P. C. Holland, F. L. Schiffino, Mini-review: Prediction errors, attention and associative learning. *Neurobiol. Learn. Mem.* **131**, 207–15 (2016).

21. A. G. E. Collins, J. F. Cavanagh, M. J. Frank, Human EEG uncovers latent generalizable rule structure during learning. *J. Neurosci.* **34**, 4677–85 (2014).

22. H. B. M. Uylings, H. J. Groenewegen, B. Kolb, Do rats have a prefrontal cortex? *Behav. Brain Res.* **146**, 3–17 (2003).

23. M. Laubach, L. M. Amarante, K. Swanson, S. R. White, What, If Anything, Is Rodent Prefrontal Cortex? *eNeuro*. **5** (2018), doi:10.1523/ENEURO.0315-18.2018.

24. J. D. Wallis, K. C. Anderson, E. K. Miller, Single neurons in prefrontal cortex encode abstract rules. *Nature*. **411**, 953–956 (2001).

25. S. Berg, D. Kutra, T. Kroeger, C. N. Straehle, B. X. Kausler, C. Haubold, M. Schiegg, J. Ales, T. Beier, M. Rudy, K. Eren, J. I. Cervantes, B. Xu, F. Beuttenmueller, A. Wolny, C. Zhang, U. Koethe, F. A. Hamprecht, A. Kreshuk, ilastik: interactive machine learning for (bio)image analysis. *Nat. Methods*. **16**, 1226–1232 (2019).

26. R. C. Wilson, Y. Niv, Inferring Relevance in a Changing World. *Front. Hum. Neurosci.* **5**, 189 (2012).

27. M. Basseville, Detecting changes in signals and systems—A survey. *Automatica*. **24**, 309–

326 (1988).

28. E. Russo, T. Ma, R. Spanagel, D. Durstewitz, H. Toutounji, G. Köhr, Coordinated Prefrontal State Transition Leads Extinction of Reward-Seeking Behaviors. *J. Neurosci.* **41**, 2406–2419 (2021).

29. H. Toutounji, D. Durstewitz, Detecting Multiple Change Points Using Adaptive Regression Splines With Application to Neural Recordings. *Front. Neuroinform.* **12**, 67 (2018).

30. R. A. Rescorla, A. R. Wagner, in *Classical Conditioning. 2. Current Research and Theory.*, P. W. Black AH, Ed. (Appleton-Century-Crofts, New York, 1972), pp. 64–69.

31. K. A. Ludwig, R. M. Miriani, N. B. Langhals, M. D. Joseph, D. J. Anderson, D. R. Kipke, Using a common average reference to improve cortical neuron recordings from microelectrode arrays. *J. Neurophysiol.* **101**, 1679–89 (2009).

32. C. Rossant, S. N. Kadir, D. F. M. Goodman, J. Schulman, M. L. D. Hunter, A. B. Saleem, A. Grosmark, M. Belluscio, G. H. Denfield, A. S. Ecker, A. S. Tolias, S. Solomon, G. Buzsaki, M. Carandini, K. D. Harris, Spike sorting for large, dense electrode arrays. *Nat. Neurosci.* **19**, 634–641 (2016).

33. E. M. Meyers, The neural decoding toolbox. *Front. Neuroinform.* **7**, 8 (2013).

34. T.-P. Jung, S. Makeig, M. J. McKeown, A. J. Bell, T.-W. Lee, T. J. Sejnowski, Imaging Brain Dynamics Using Independent Component Analysis. *Proc. IEEE. Inst. Electr. Electron. Eng.* **89**, 1107–1122 (2001).

35. R. Oostenveld, P. Fries, E. Maris, J.-M. Schoffelen, FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Comput. Intell.*

Neurosci. **2011**, 156869 (2011).

36. R. S. Desikan, F. Ségonne, B. Fischl, B. T. Quinn, B. C. Dickerson, D. Blacker, R. L. Buckner, A. M. Dale, R. P. Maguire, B. T. Hyman, M. S. Albert, R. J. Killiany, An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage*. **31**, 968–80 (2006).

37. J.-M. Schoffelen, A. Hultén, N. Lam, A. F. Marquand, J. Uddén, P. Hagoort, Frequency-specific directed interactions in the human brain network for language. *Proc. Natl. Acad. Sci. U. S. A.* **114**, 8083–8088 (2017).

38. M. S. Treder, MVPA-Light: A Classification and Regression Toolbox for Multi-Dimensional Data. *Front. Neurosci.* **14**, 289 (2020).

39. S. Farashahi, K. Rowe, Z. Aslami, D. Lee, A. Soltani, Feature-based learning improves adaptability without compromising precision. *Nat. Commun.* **8**, 1768 (2017).

40. B. M. Lake, T. D. Ullman, J. B. Tenenbaum, S. J. Gershman, Building machines that learn and think like people. *Behav. Brain Sci.* **40**, e253 (2017).

41. A. M. Owen, A. C. Roberts, C. E. Polkey, B. J. Sahakian, T. W. Robbins, Extra-dimensional versus intra-dimensional set shifting performance following frontal lobe excisions, temporal lobe excisions or amygdalo-hippocampectomy in man. *Neuropsychologia*. **29**, 993–1006 (1991).

Acknowledgments: We thank the lab of Prof. Hamprecht (IWR Heidelberg) and Dr. T. Enkel. Some of the analyses required the use of a high-performance computing cluster (bwUniCluster 2.0, https://www.bwhpc.de/). The authors acknowledge support by the state of Baden-Württemberg through bwHPC. All figures were created using BioRender.com. Figures include color specifications and designs developed by Cynthia Brewer (http://colorbrewer.org/).

Funding: This work was funded by the German Research Foundation (BA 5382/1-1, BA 5382/2-2to FB, DU 354/8-2, DU 354/10-1 to DD), the third CIMH Young Investigator Award and the MACS (Mannheim Advanced Clinician Scientist) Program (FB).

Author contributions:

Experimental design: FB, TP

Investigation: FB, TP, SH, NB, TM, HZ

Methodology: FB, GK, HT, DD

Data analysis: FB, TP, SH, NB, TM, HZ

Computational modelling: HT

Writing – original draft: FB, TP, HT

Writing – review & editing: FB, TP, GK, AML, HT, DD

Visualization: FB, NB

Funding acquisition: FB, DD

Supervision: FB, TP, AML, HT, DD

Competing interests: AML has received consultant fees from the American Association for the Advancement of Science, Atheneum Partners, Blueprint Partnership, Boehringer Ingelheim, Daimler und Benz Stiftung, Elsevier, F. Hoffmann-La Roche, ICARE Schizophrenia, K. G. Jebsen Foundation, L.E.K Consulting, Lundbeck International Foundation (LINF), R. Adamczak, Roche Pharma, Science Foundation, Sumitomo Dainippon Pharma, Synapsis Foundation – Alzheimer Research Switzerland, System Analytics, and has received lectures fees including travel fees from Boehringer Ingelheim, Fama Public Relations, Institut d'investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), Janssen-Cilag, Klinikum Christophsbad, Göppingen, Lilly Deutschland, Luzerner Psychiatrie, LVR Klinikum Düsseldorf, LWL Psychiatrie Verbund Westfalen-Lippe, Otsuka Pharmaceuticals, Reunions i Ciencia S. L., Spanish Society of Psychiatry, Südwestrundfunk Fernsehen, Stern TV, and Vitos Klinikum Kurhessen. All other authors declare that they have no competing interests.

Data and materials availability: All data, code and materials will be made available on heiDATA (https://heidata.uni-heidelberg.de/) upon publication.

Supplementary Materials

Materials and Methods

Supplementary Text

Figs. S1 to S9

Tables S1 to S4

References (*25–41*)

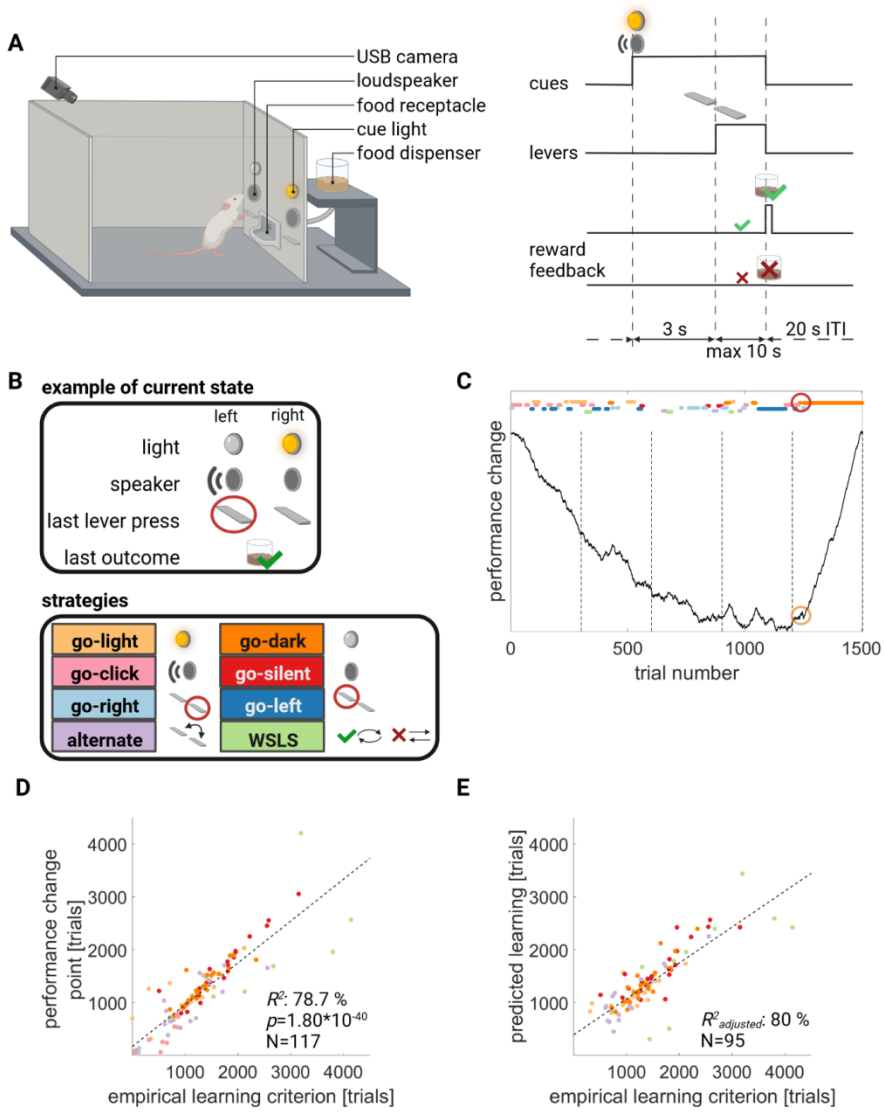

Fig. 1. Rats test behavioral strategies to infer the task rule.

(**A**) In the multidimensional rule-learning task, each trial starts with the presentation of one auditory and one visual cue. 3s later, two levers extend and rats use trial-and-error to learn the correct choice. (**B**) Our core hypothesis is that rats sequentially follow low-dimensional behavioral strategies (*bottom*) to test four task features for relevance (e.g., *go-click*/*go-silent* to test the relevance of the loudspeakers for reward). This is in contrast with the classic notion that

animals learn all possible mappings between cue combinations (see *top* for example) and actions independently. (**C**) Performance change of an individual rat learning the task rule go-dark. Strategy sequences are color-coded, change in performance is visualized using a CUSUM plot. Detected performance change point and empirical learning criterion are marked with circles, dashed vertical lines mark the end of an experimental session. (**D**) High correlation between empirical learning criterion and performance change point, each dot is a rat learning one out of seven possible rules. (**E**) A linear regression model shows that strategy-related parameters significantly predict the learning trial. ITI/inter-trial interval.

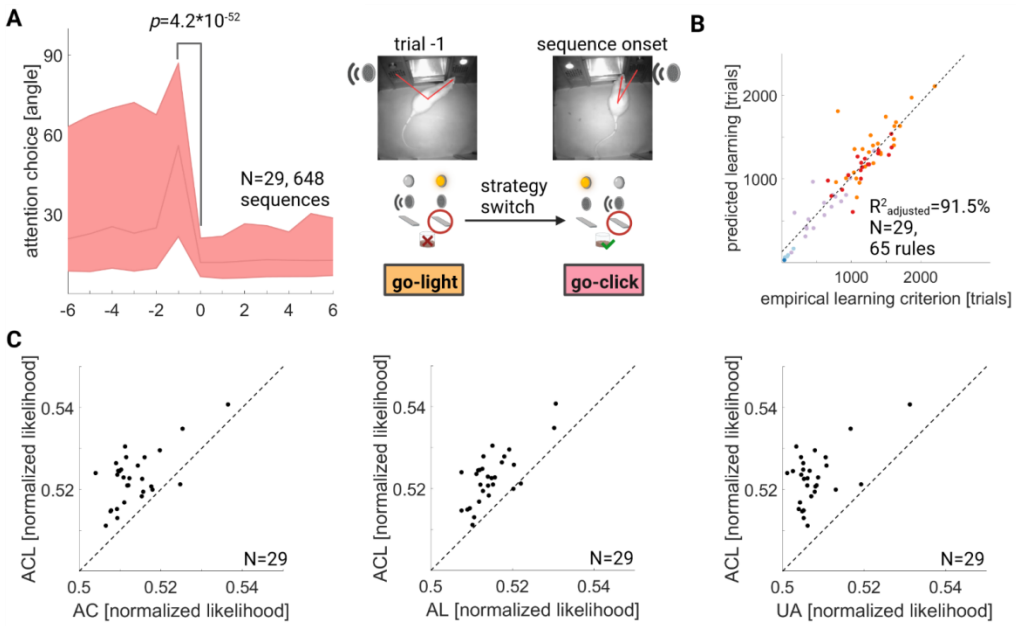

Fig. 2. Attention effects during choice and learning.

(**A**) *Left:* Head direction plot for strategy *go-click* during choice formation. Head angles with respect to sound are aligned to cue onset and abruptly decrease at sequence onset (trial 0). *Right:* example of head angles in trial before and at sequence onset. (**B**) Abrupt changes in attention (during choice formation and reward feedback) across trials predict learning. Each dot represents empirical vs. predicted learning trial, rules are color-coded as in Fig. 1. (**C**) Attention-modulated RL models that include both empirical attention scores at choice and learning (ACL) as compared to models assuming only attention-at-choice (AC) or attention-at-learning (AL) or no attention focus (uniform-attention/UA). Data points above the dotted line indicate that the ACL model explains the data best.

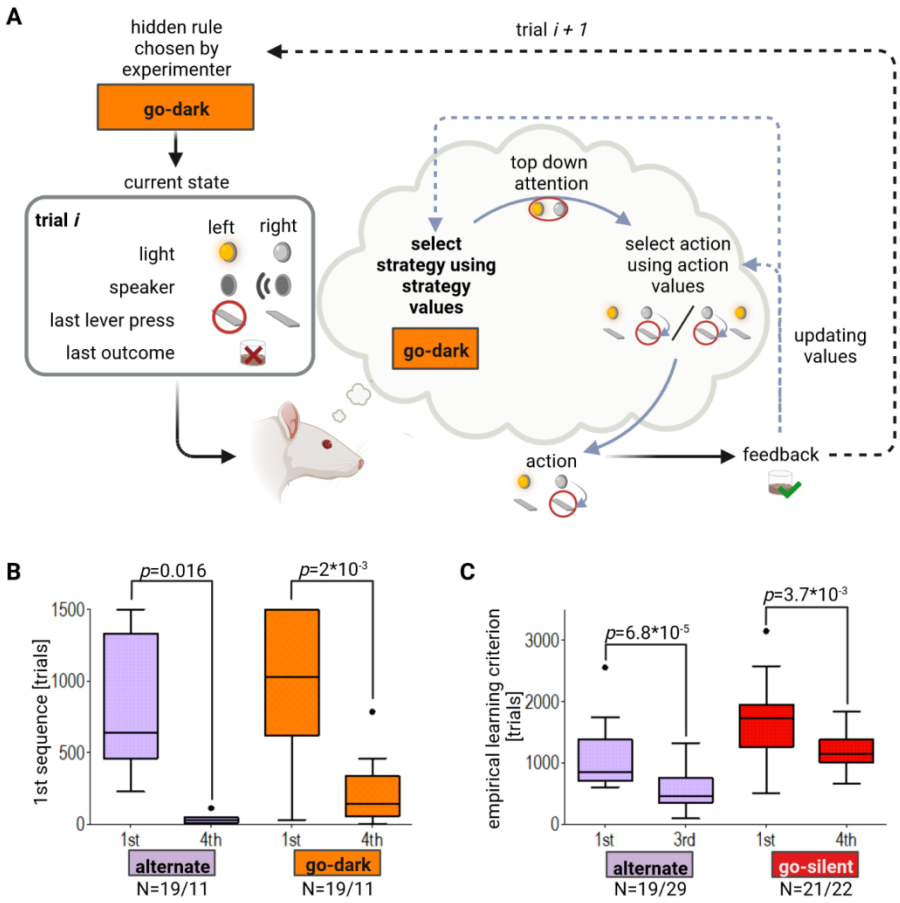

Fig. 3. Behavioral strategies as hypothetical rules.

(**A**) Strategies are selected based on their value as predicted by hierarchical RL models. (**B**) In the absence of a task rule (randomized reward feedback), strategies are selected earlier if they have been previously reinforced (i.e., in experienced rats/4th rule) as compared to a naïve cohort (1st rule). (**C**) Hierarchical models also predict positive transfer effects. Indeed, experienced rats learning the task rules alternate/go-silent reached the criterion faster as compared to naïve animals.

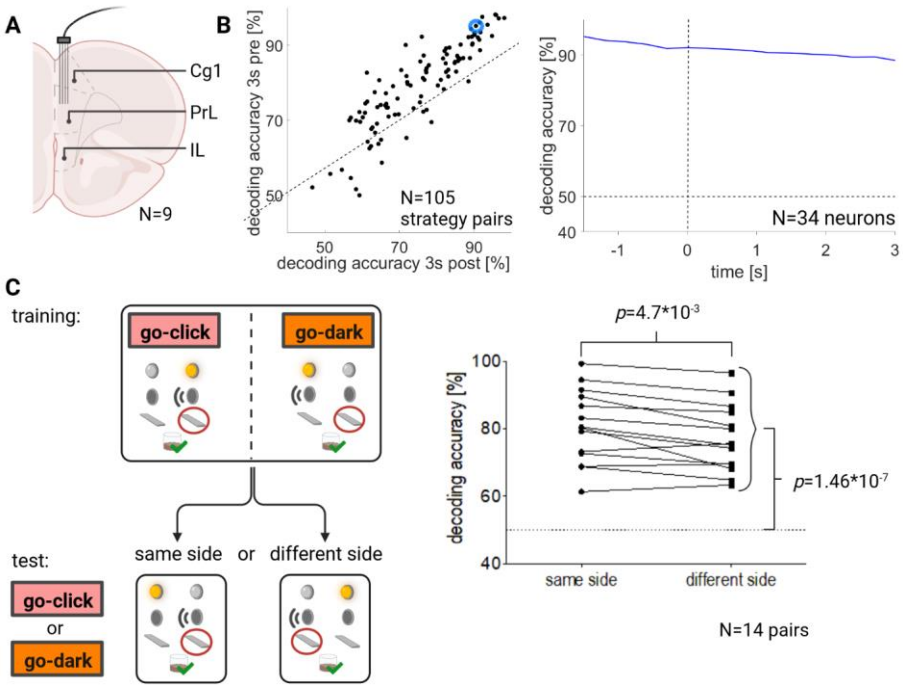

Fig. 4. The rat PFC encodes strategies as abstract rules.

(**A**) Population decoding of selected strategies based on prelimbic multiple single-unit activity recordings. The neural code has two features of an abstract rule. (**B**) Decoding accuracy peaks *prior* to cue onset, indicating that action selection is a top-down process and not driven by cues. *Inset:* decoding accuracy across time shown for the highlighted strategy pair. (**C**) Strategies are abstract actions and their prefrontal representation should thus be invariant with respect to motor response (i.e., left vs. right lever press) in a given trial. To test this prediction, we trained a classifier on trials with lever presses on one side and compared decoding accuracy when tested on trials with lever presses on either the same or opposite side (example shown *left*). Decoding accuracy in the 3s before cue onset was lower if sides were different in training and test conditions but still much higher than chance (*right*).

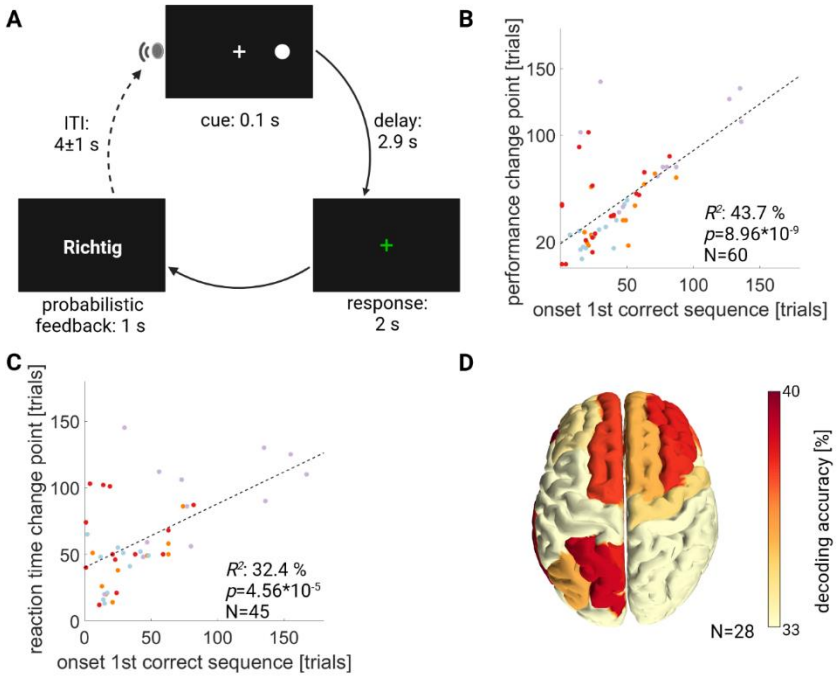

Fig. 5. Behavioral and MEG evidence for behavioral strategies in humans.

(**A**) Healthy adults solved an adapted version of the multidimensional rule-learning task during MEG recordings. (**B**) Significant correlation between the onset of the first correct strategy sequence and the performance change point mirrors rat findings. (**C**) Significant correlation between the onset of the first correct strategy sequence and a decrease in reaction time. (**D**) Decoding of the relevant task feature at the source level over the averaged time window of 0 to 1 s after stimulus onset using a searchlight approach. The spatial topography of decoding accuracy reveals a predominantly prefronto-parietal network, including superior and middle frontal gyrus, cingulate, superior parietal gyrus and precuneus (33 % corresponds to chance level).