Conjunctive Representations that Integrate Stimuli, Responses, and Rules are Critical for Action Selection

Atsushi Kikumoto and Ulrich Mayr

University of Oregon

Acknowledgements: This research was supported by NIA grant R01 AG037564-01A1, and by NSF grant NSF grant 1734264.

Abstract

People can use abstract rules to flexibly configure and select actions for specific situations. Yet how exactly rules shape actions towards specific sensory and/or motor requirements remains unclear. Both research from animal models and human-level theories of action control point to the role of highly integrated, conjunctive representations, sometimes referred to as event files. These representations are thought to combine rules with other, goal-relevant sensory and motor features in a nonlinear manner and are a necessary condition for action selection. However, so far, no methods exist to track such representations in humans *during* action selection with adequate temporal resolution. Here, we applied time-resolved representational similarity analysis to the spectral-temporal profiles of the EEG signal, while participants performed a cued, rule-based action selection task. In two experiments, we found that conjunctive representations were active throughout the entire selection period and were functionally dissociable from the representation of constituent features. Specifically, the strength of conjunctions was a highly robust predictor of trial-by-trial variability in response times, and was selectively related to an important behavioral indicator of conjunctive representations, the so-called partial-overlap priming pattern. These results provide direct evidence for conjunctive representations as critical precursors of action selection in humans.

Significance Statement

A tennis player planning the next stroke has to consider disparate pieces of information: the type of stroke, the trajectory of the incoming ball, or where to hit the ball. By decoding EEG signals while participants executed simple, rule-based actions, we demonstrate here that such different, action-relevant aspects are integrated within a unified, *conjunctive* representation—rather than being processed in a piecemeal manner. Furthermore, the strength of conjunctive representations is a highly robust predictor of how quickly actions are executed in each trial. Human-level theories and recent single-cell evidence from animal models suggests that conjunctive representations are a necessary condition for successful action selection. Our results provide the first direct evidence in humans that is consistent with this hypothesis.

Conjunctive Representations that Integrate Stimuli, Responses, and Rules are Critical for Action Selection

Flexible, goal-directed action requires the use of abstract rules that can be applied to a range of specific situations. However, we know little about how such rules connect with lower-level sensory or response representations, as a specific action is planned and executed. In traditional stage-based processing models, rules or task sets regulate the flow of information from stimulus to response in form of a cascade of relatively independent processing steps (1-5). In contrast, recent results from research in non-human primates suggest a critical role of neurons with nonlinear, mixed selectivity response properties that integrate various aspects (rules, stimuli, and responses) in a conjunctive manner (6, 7). Similarly, some cognitive psychologists have proposed—mostly on the basis of behavioral results—that as a necessary step for action selection, relevant features, including rules, need to be combined into highly integrated, conjunctive representations, referred to as *event files* (8-11). However, no direct, neural-level indicator of event files exists, making it difficult to bridge the gap between theories about integrated representations in human action selection and the literature on mixed selectivity neurons in animal models.

Currently, the main signature of event files is an indirect, behavioral aftereffect, the *partial-overlap priming cost* (Fig.1A): When either all or none of the action-relevant features repeat across consecutive trials (e.g., both rule and response either repeat or change), performance is relatively fast. In contrast, when only some, but not all features overlap across trials (e.g., response repeats, but rule changes), response times (RT) and/or errors increase. According to event-file theory, entire event files can be easily repeated or replaced. However, when an overlapping feature needs to be extracted

from a recently activated event file, RT or error costs arise, thus leading to the partial-overlap priming pattern.

There is also neuroimaging evidence about how the partial-overlap cost pattern is expressed neuroanatomically (12, 13) or in evoked EEG components (14). However, given that partial-overlap costs are an *aftereffect* of event file formation, this pattern provides no information about how conjunctive representations versus their constituent features behave during response selection, and whether or not they are indeed a critical precursor of successful action. Moreover, partial-overlap costs can also be explained by alternative models, such as in terms of interactions between distinct hierarchical levels of control (15, 16), or as an indirect consequence of response inhibition (17).

In order to evaluate the hypothesized role of conjunctive representations, we used the EEG signal to decode information about action-relevant representations in a time-resolved manner (18-20), while participants selected responses on the basis of randomly cued, action rules (11) (Fig.1AB). Conjunctive representations are per definition correlated with the representations of constituent features. To tease apart these correlated representations, we used representational similarity analysis (RSA) (21, 22). Standard RSAs require information about the similarity of the multivariate neural signals across conditions (e.g., based on correlations), which however, cannot be computed on the level of individual trials. Therefore, we performed RSAs using confusion profiles that resulted from an initial step of decoding each of the possible action-relevant constellations on the level of individual trials and time-points (Fig.1CD, see *Representational Similarity Analysis* in *Methods*). This two-step procedure allowed

us to examine the role of conjunctive and constituent representations in predicting trial-by-trial variability in performance.

In two experiments our results showed—for the first time in humans—temporally precise evidence for the activation of conjunctive representations *during* action selection. Consistent with both findings from mixed selectivity neurons and event-file theory, conjunctions were robust and unique predictors of variability in performance and were specifically related to the pattern of partial-overlap priming effects.

Results

We conducted two experiments that we report together. In Experiment 1, we used the spatial rules task with three different rules (see Fig.1AB), which did not allow us to differentiate between different types of conjunctions (i.e., S-R conjunctions versus rule-S-R conjunctions). In Experiment 2, we used an expanded task space with four different rules (Fig.1EF), which included conjunctions that shared the same S-R pairs, but different abstract rules (11). This allowed us to dissociate conjunctions that integrate rules (i.e., rule-S-R conjunctions) from rule-independent conjunctions (i.e., S-R conjunctions).

Behavior

For all analyses, error-trials, post-error trials, and trials in which RTs were larger than 99.5 percent of each individuals' RT distributions were excluded. In both experiments, and consistent with previous work (11), we observed partial-overlap costs in RTs and errors as a function of the different trial-to-trial transitions (Fig.2): In Experiment 1, when either rules and responses repeated or when both changed, responses were fast and accurate, whereas costs emerged in the case of partial

updates of either rules or stimuli/responses. In Experiment 2, the repetition of rule-S-R settings produced RT and error benefits, whereas any partial updates (including S-R repetitions with rule changes) generated costs (for statistical analyses, see SI Appendix, Tables S1 and S2).

Tracking Representational Dynamics

To directly assess the role of conjunctive representations during action selection, we used time-resolved RSAs on the level of single trials (Fig.3A and 4A). Consistent with previous results, the cascade of decoded representations unfolded consistent with the expected flow of information: The rule was activated during the pre-stimulus phase, followed by a strong expression of the stimulus, and finally by the response (18, 20). Critically, over and above these constituent features, the conjunctive representations were active during the entire post-stimulus period (Fig.3A and 4A) in both Experiments. In Experiment 2, rule-S-R conjunctions were more strongly expressed than rule-independent, S-R conjunctions. Also, in both experiments, conjunctions emerged in tandem with (in Experiment 1), or clearly before response activation (in Experiment 2, see also SI Appendix, Fig.S8). Consistent with event-file theory, this temporal pattern suggests that conjunctions arise during response selection and not just as a response-selection aftereffect.

Note, that expression of conjunctions was statistically robust even though we accounted for subject-specific differences in RTs between action constellations. Thus, decoding results cannot be explained in terms of unspecific difficulty differences between action constellations (SI Appendix, Fig.S1 and S2). In Experiment 2, we observed that the rule representation diminished after stimulus onset (Fig.4A).

Excluding conjunction models restored the post-stimulus rule representation, suggesting that the rule-S-R conjunction model captures the same variance as the rule model explains in this phase of action selection (Fig.4A inset).

Do Conjunctions Uniquely Predict Trial-to-trial Performance Fluctuations?

To test the prediction from event-file theory that conjunctive representations are critical for action selection, we regressed trial-to-trial variation in RTs onto the strength of each expressed representation. Using multilevel modeling, we performed these analyses for each time-point and with all predictors entered simultaneously. The resulting "impact-trajectories" are shown in Fig.3B and 4B (statistical results for a-priori selected time intervals are summarized in SI Appendix, Table S3 and S4; see also SI Appendix, Fig.S3 and S4 for corresponding results from standard decoding analyses). Note that negative *t*-values indicate that stronger representations lead to faster responses. Consistent with the prediction from event-file theory, conjunctive representations were the dominant predictors of performance in both experiments. In Experiment 2, both rule-S-R conjunctions and S-R conjunctions explained substantial, independent variability in trial-to-trial RTs (Fig.4B), with a slight edge for the rule-specific conjunctions. Combined, these results indicate that conjunctive representations emerge during response-selection, and predict upcoming behavior over and above the influence of the constituent features.

Are Conjunctions Related to Partial-Overlap Costs?

In order to directly connect the EEG-decoded conjunctive representations with the theoretical, event-file construct, we examined how these representations relate to the partial-overlap priming pattern. As Fig.3C and 4C show, the strength of decoded

conjunctions expresses the partial-overlap pattern in both Experiments 1 and 2. Conjunctive representations were particularly strong in those transitions in which RTs were fast (i.e., when either everything repeated or everything changed, see Fig.2). In Experiment 1, conjunctive representations showed the partial-overlap pattern in the expected direction during the early, b=.024, SE=.010, $t(20)$=2.58, but not in the late post-stimulus phase, b=.004, SE=.010, $t(20)$=.39, and none of the constituent features showed the critical interaction pattern, all $t(20)$<.21. In Experiment 2, only the strength of rule-S-R conjunctions showed the partial-overlap costs, b=.021, SE=.009, $t(21)$=2.22, for the early selection phase; b=.021, SE=.009, $t(21)$=2.24, and for the late selection phase (Fig.4C). None of the constituent features, all ts(21)<.72), or S-R conjunctions showed such an effect, b=.012, SE=.009, $t(21)$=1.27 for the early selection phase; b=.007, SE=.010, $t(21)$=.72, for the late selection phase.

Another important prediction that can be derived from the event-file model is that strong conjunctions should be particularly difficult to “unbind” on the following trial. Thus, the stronger the conjunction on trial n-1, the larger should be the partial-overlap costs on trial n. Our results, shown in Fig.3D and 4D, confirm this prediction for both experiments: A stronger conjunctive representation in trial n-1, late in the selection period, led to a greater RT partial-overlap costs on trial n, b=.025, SE=.011, $t(20)$=2.25. Importantly, this pattern was unique for conjunctive representations, and was not found for any of the constituent representations, all ts(20)<.05. In Experiment 2, only rule-S-R conjunctions significantly modulated RT partial-overlap costs on the next trial (Fig.4D), b=.031, SE=.011, $t(20)$=2.81. Again, this pattern was absent for S-R conjunctions or any other of the constituent representations, all ts(21)<.38. Thus, it is specifically the

highest-order, rule-S-R conjunctions that relate to the partial-overlap cost pattern. Taken together, the behavior of decoded conjunctive representations was highly consistent with predictions from the event-file model.

Discussion

We tested whether integrated, conjunctive representations between task-relevant features emerge during action selection, as predicted from results with mixed selectivity neurons (6), and by event-file theory (8, 9). In our paradigm, action settings had to be updated flexibly for each trial, creating unique constellations between rules, stimuli, and responses. We combined a standard, linear decoding approach with a subsequent, time-resolved RSA in order to track the emergence of conjunctive representations and their constituent features over time, and for each individual trial.

The time course of decoded information showed a highly plausible cascade of action representations (rule, stimulus, and then response). Most critically, we found robust evidence for conjunctive representations—emerging shortly after stimulus onset and persisting until response execution. Analyses with response-locked EEG data fully confirmed this pattern of results (SI Appendix, Fig.S8). The fact that conjunctive representations were continuously present from stimulus processing to response execution is consistent with their role in translating sensory codes into response codes based on the current task rules.

Even though conjunctive representations were on average less strongly expressed than those of constituent features, they were statistically, highly robust (Fig.3B and Fig.4B and SI Appendix, Fig.S10). Moreover, conjunctive representations were strong and unique predictors of trial-by-trial variability in RTs, over and above

other constituent features. These results are difficult to reconcile with traditional stage theories (1-5) and hierarchical control models (16, 23), where information flows in a strict feed-forward manner and thus allows no integrated representations to emerge. This is the more remarkable as our task design, with explicitly cued rules that appeared prior to each stimulus, should have been clearly compatible with a hierarchical selection architecture (e.g., first selection of rule, then of rule-specific S-R link). Instead, our results indicate that action selection is established by tying together the disparate, task-relevant features from the entire selection event into a common representation.

In Experiment 1, conjunctions could entail any pairwise, or complete combination of rule, stimulus, or response features; in Experiment 2, we were further able to dissociate between rule-specific, rule-S-R conjunctions and rule-independent, S-R conjunctions. The fact that in Experiment 2, both rule-S-R and S-R conjunctions emerged is an important result in its own right. It suggests that integrated representations that match the contingencies in the environment develop in parallel on different levels of specificity. This combination of both rule-specific and rule-independent representations can account for past findings showing that S-R associations learned within one rule can transfer to another rule, albeit in a limited manner (11, 24).

A key behavioral indicator of event files is the partial-overlap priming pattern (see Fig.1A) (11, 25). In both experiments, we found this pattern not only in RTs and errors (Fig.2), but also in the strength of conjunctions (Fig.3C and 4C). More importantly, the strength of conjunctions in trial *n*-1, predicted the size of partial-overlap costs in trial *n* (Fig.3D and 4D), suggesting the tighter the integration between action features, the harder it is to “unbind” the features in order to integrate them into a new conjunction.

Importantly, only conjunctions, and not the basic features, showed such a relationship with the partial-overlap pattern, thus functionally dissociating conjunctions from their constituent codes. The results in Experiment 2 also indicated that specifically rule-S-R conjunctions were related to the partial-overlap cost, not, however S-R conjunctions. It is noteworthy that the conjunctions in Experiment 1 (where we were not able to distinguish between S-R and rule-S-R conjunctions) showed a similar priming pattern as the rule-S-R conjunctions in Experiment 2, suggesting integration of not just stimuli and responses, but also of rules in both experiments.

While our results provide temporal and functional information about specific representations, they are relatively silent about the underlying neural mechanisms or their neuroanatomical location. We had no strong a-priori predictions about frequency bands that might contain conjunction-specific information, we used a broad spectrum of frequencies for decoding. In post-hoc analyses we found that the pattern of EEG responses that underlies conjunctive representations is idiosyncratic (SI Appendix, Fig.S5), but most strongly expressed in the delta-band frequency signal (SI Appendix, Fig.S6 and S7). This latter result is generally consistent with previous evidence that decision-relevant representations can be decoded from oscillations in the delta band (26).

Regarding the question of neuroanatomical location, research with animal models points to the hippocampus as particularly critical for representing highly contextualized, conjunctive information (6, 27-29). There is also some evidence from human neuroimaging work that implicates the hippocampus in retrieving incidentally learned associations between actions and their consequences, albeit using paradigms

that involve learning across longer time frames (30, 31). In addition, single-neuron electrophysiological work with non-human animals indicates that neurons coding task-relevant features are distributed across frontal and parietal cortex (32, 33). A large proportion of recorded neurons in these areas integrate multiple features in a nonlinear manner (6, 34). Such heterogeneous, neural responses allow both efficient, linear read-out of information to downstream neurons and can also code conjunctive information in a high-dimensional format (35). In human neuroimaging work, attempts to decode high-level, task-relevant representations in frontal areas have proven more challenging (36).

An important finding from the research on mixed selectivity neurons is that the degree of nonlinear information coded in these neurons is functionally distinct from the representation of linear features. For example, nonlinear responses were found to be highly robust in correct trials, but was largely missing in error trials, whereas simple, linear information was equally present on correct and error trials (6, 7, 37). This pattern is consistent with our finding that the strength of conjunctive representations uniquely predicts trial-by-trial performance, beyond the predictive strength of constituent, simple features (Fig.3B and 4B). As mentioned earlier, further evidence for a functional dissociation comes from our finding that only conjunctive representations, but not the representations of constituent, simple feature, express the partial-overlap priming pattern (see Fig.3CD and 4CD).

Our results about the relevance of conjunctions for both efficient action selection and for the partial-overlap priming pattern, directly confirm predictions derived from event-file theory. Therefore, they provide an important missing link between two distinct lines of research: The relatively abstract, event-file conceptualization, designed to

explain the architecture of human action selection, and the recent progress in characterizing the format of representations arising from mixed selectivity neurons. For example, from the here established relationship between the partial-overlap pattern and the neural-signature of event files, we can derive the testable prediction that the partial-overlap pattern should also be selectively expressed in the activity of mixed-selectivity neurons in animal models. In addition, the current results raise a range of new questions about the functional properties of conjunctive representations: We do not know how these representations are constrained by capacity limitations (38), how they respond to distracting information (7), to what degree they allow integration of action outcomes or goals (39), or how they change through experience (SI Appendix, Fig.S9) (11). The EEG decoding approach used here, provides the tools to address these and related questions in human participants.

Method

See SI Appendix for additional method details.

Participants

Forty-four people participated after signing informed consent following the protocol approved by the University of Oregon's Human Subjects Committee in exchange for $10 per hour and additional performance-based incentives. Participants with a pre-defined criterion of more than 35% of trials with EEG artifacts were removed from further analysis, leaving 20 out of 22 participants for Experiment 1 and 21 out of 22 for Experiment 2.

Stimuli, Tasks and Procedure

Participants performed a cued rule-selection task, in which one of the pre-instructed action rules was randomly selected to determine possible S-R mappings, on a trial-by-trial basis (Fig.1B) (11). Based on the cued rule, participants responded to the location of a circle (1.32° in radius) that randomly appeared in the corner of a white frame (6.6° off-center) by selecting one of the four response keys that were arranged in a 2 x 2 matrix. Each action rule specified four S-R links using a simple spatial transformation rule. For instance, the “vertical” rule mapped the top-left circle to the bottom-left response. In order to ensure that decoding of rule information was not driven by superficial, perceptual aspects, we used two cues for each rule (a pair of verbal cues in Experiment 1 and symbol/word pair in Experiment 2, see Fig.1C and 1E) that appeared in either even or odd trials to prevent immediate cue repetitions. In Experiment 1,”vertical”, “horizontal” and “diagonal” rules were used (i.e., 66.6 % switch rate). In Experiment 2, for different rules, ”vertical”, ”horizontal”, “clockwise” and “counterclockwise” rules were used (i.e., 75% switch rate). This specific set of rules ensured that each S-R link occurred in two different rules (e.g., a top-left circle leads to a bottom-left response in both the vertical and the clockwise rule), allowing us try to decode both rule-S-R conjunctions and rule-unspecific, S-R conjunctions in Experiment 2 (Fig.1EF).

We presented two practice blocks and 200 experimental blocks per experiment. Participants were instructed to complete as many correct trials as possible within each 16-second block. Trials that began within the 16 seconds were allowed to complete.

EEG Recording and Processing

Electroencephalographic (EEG) activity was recorded from 20 tin electrodes using the International 10/20 system and preprocessed to remove artifacts (see *EEG Recording and Processing* in the SI). Furthermore, temporal-spectral profiles of single-trial EEG data were obtained via complex wavelet analysis(40) by applying time-frequency analysis (1-35 *Hz*) to preprocessed EEG data (see SI Appendix, *Time-frequency Analysis*). This analysis resulted in a frequency-band specific power estimate at each sample point. As in our previous work (18), we averaged five different frequency bands:1-3 *Hz* for the delta-band, 4-7 *Hz* for the theta-band, 8-12 *Hz* for the alpha-band, 13-30 *Hz* for the beta-band, 31-35 *Hz* for the gamma-band to prepare training data for the decoding analyses. Within individuals, frequency-specific power values were z-transformed across electrodes at every sample to remove the effects that uniformly influenced all electrodes. While we had no a-priori predictions about the role of specific frequency bands in representing different action-relevant representations, we present post-hoc analyses probing the relevance of each frequency band in the SI (see SI Appendix, Fig.S6 and S7).

Representational Similarity Analysis

To obtain information about the strength of each feature and conjunction on the level of individual trials and time-points, we used a two-step procedure. First, we performed a linear decoding analysis to discriminate between all 12 different action constellations in Experiment 1, or all 16 constellations in Experiment 2. This analysis was conducted for each time-point and used the average power of rhythmic EEG activity within the predefined frequency bands (delta, theta, alpha, beta, and gamma),

generating 100 features (5 frequency-bands x 20 electrodes) to train decoders. Following cross-validation, this decoding step yielded a vector of “confusion profiles” of classification probabilities for both the correct and all possible incorrect classifications, and for each time-point and trial (Fig.1D). As a second step, we applied RSA (21) to each profile of classification probabilities in order to determine their underlying similarity structure for each time-point and trial. Specifically, we regressed the classification-probability vector onto model vectors as simultaneously entered predictors, which were derived from a set of RSA model matrices (Fig. 1D).

Each model matrix represented a potential, underlying representation. In Experiment 1, we constructed RSA models for the rules, stimuli, responses, and conjunctions (Fig.1D). In Experiment 2, we used separate matrices for the rule-specific S-R conjunction model (rule-S-R conjunction) and the rule-independent S-R conjunction model (S-R conjunction; Fig.1F). Complete orthogonalization of basic features could be established within each of two equal-sized subspaces, but not across the entire space of action constellations. Specifically, one subspace (G1 in Fig.1E) contained constellations with stimuli at the top-left or bottom-right corner (leading to a bottom-left or bottom-right response for all rules), whereas the second subspace (G2 in Fig.1E) contained trials with stimuli at the left-bottom or top-right corner (leading to a top-left or bottom-right response). Within each subspace, conjunctions were defined by the combination of four rules (vertical, horizontal, clockwise, and counterclockwise), two stimulus positions, and two responses, ensuring that each S-R link could occur in the context of two different action rules.

Multilevel Modeling

For the results shown in Fig.3C and 4C, and well as in Fig.3D and 4D, we used multilevel linear modeling to analyze within-subject variability in RSA scores as a function of trial-to-trial transition variables (Fig.3C and 4C), or in RTs as a function of trial *n*-1 RSA scores and trial-to-trial transition variables (Fig.3D and 4D). In each case, subject-specific intercepts and slopes were included as random effects. Log-transformed RTs as dependent variable were pre-whitened by linear and quadratic trends of experimental trials and blocks. We performed statistical tests for a-priori selected time intervals: cue-to-stimulus period from the onset of cue to the onset of stimulus (-300 to 0 ms for Experiment 1 and -500 to 0 ms for Experiment 2), early post-stimulus period (0 to 300 ms of the post-stimulus segment for both experiments), and late post-stimulus period (300 to 600 ms of the post-stimulus segment for both experiments). We predicted trial-to-trial RTs/RSA scores in the current trials with EEG signals from pre-stimulus and early post-stimulus periods in order to capture processing prior to response execution (see also SI Appendix, Fig.S8 for results using signals aligned to the response onsets). The late post-stimulus interval was used to assess how partial-overlap costs are modulated by the strength of action representations developed during selection in *n*-1 trials (i.e., Fig.3D and 4D). In addition, to visualize the impact of different decoded features on RTs across time, we also ran fixed-effect models plus random intercepts at each sample point, but without random slopes (Fig.3B and 4B).

References

1. F. C. Donders, On the speed of mental processes. *Acta Psychol.* **30**, 412-431 (1969).
2. S. Sternberg, The discovery of processing stages: Extensions of Donders' method. *Acta Psychol.* **30**, 276-315 (1969).
3. S. Kornblum, T. Hasbroucq, A. Osman, Dimensional overlap: cognitive basis for stimulus-response compatibility--a model and taxonomy. *Psychol. Rev.* **97**, 253 (1990).
4. A. F. Sanders, A. Sanders, *Elements of human performance: Reaction processes and attention in human skill* (Psychology Press, 2013).
5. M. I. Posner, R. F. Mitchell, Chronometric analysis of classification. *Psychol. Rev.* **74**, 392 (1967).
6. M. Rigotti *et al.*, The importance of mixed selectivity in complex cognitive tasks. *Nature* **497**, 585 (2013).
7. A. Parthasarathy *et al.*, Mixed selectivity morphs population codes in prefrontal cortex. *Nat. Neurosci.* **20**, 1770-1779 (2017).
8. B. Hommel, Theory of Event Coding (TEC) V2. 0: Representing and controlling perception and action. *Atten. Percep. Psychophys.*, 1-16 (2019).
9. B. Hommel, J. Müsseler, G. Aschersleben, W. Prinz, The theory of event coding (TEC): A framework for perception and action planning. *Behav. Brain Sci.* **24**, 849-878 (2001).
10. E. H. Schumacher, E. Hazeltine, Hierarchical task representation: Task files and response selection. *Curr. Dir. Psychol. Sci.* **25**, 449-454 (2016).
11. U. Mayr, R. L. Bryck, Sticky rules: integration between abstract rules and specific actions. *J. Exp. Psychol. Learn. Mem. Cogn.* **31**, 337-350 (2005).
12. A. W. Keizer *et al.*, When moving faces activate the house area: an fMRI study of object-file retrieval. *Behav. Brain Funct.* **4**, 50 (2008).
13. S. Kühn, A. W. Keizer, L. S. Colzato, S. A. Rombouts, B. Hommel, The neural underpinnings of event-file management: Evidence for stimulus-induced activation of and competition among stimulus–response bindings. *J. Cognit. Neurosci.* **23**, 896-904 (2011).
14. A. Opitz, C. Beste, A.-K. Stock, Using temporal EEG signal decomposition to identify specific neurophysiological correlates of distractor-response bindings proposed by the theory of event coding. *NeuroImage*, 116524 (2020).
15. T. Kleinsorge, H. Heuer, Hierarchical switching in a multi-dimensional task space. *Psychol. Res.* **62**, 300-312 (1999).
16. T. Kleinsorge, H. Heuer, Hierarchical switching in a multi-dimensional task space. *Psychol. Forsch.* **62**, 300-312 (1999).
17. K. R. Grzyb, R. Hübner, Excessive response-repetition costs under task switching: How response inhibition amplifies response conflict. *J. Exp. Psychol. Learn. Mem. Cogn.* **39**, 126 (2013).
18. J. Hubbard, A. Kikumoto, U. Mayr, EEG Decoding Reveals the Strength and Temporal Dynamics of Goal-Relevant Representations. *Sci. Rep.* **9**, 9051 (2019).
19. J. O. Garcia, R. Srinivasan, J. T. Serences, Near-real-time feature-selective modulations in human cortex. *Curr. Biol.* **23**, 515-522 (2013).

20. S. Hall-McMaster, P. S. Muhle-Karbe, N. E. Myers, M. G. Stokes, Reward Boosts Neural Coding of Task Rules to Optimize Cognitive Flexibility. *J. Neurosci.* **39**, 8549-8561 (2019).
21. N. Kriegeskorte, M. Mur, P. Bandettini, Representational similarity analysis - connecting the branches of systems neuroscience. *Front. Syst. Neurosci.* **2**, 4 (2008).
22. T. Grootswagers, S. G. Wardle, T. A. Carlson, Decoding dynamic brain patterns from evoked responses: A tutorial on multivariate pattern analysis applied to time series neuroimaging data. *J. Cognit. Neurosci.* (2017).
23. F. M. Korb, J. Jiang, J. A. King, T. Egner, Hierarchically organized medial frontal cortex-basal ganglia loops selectively control task-and response-selection. *J. Neurosci.* **37**, 7893-7905 (2017).
24. A. G. Collins, M. J. Frank, Cognitive control over learning: Creating, clustering, and generalizing task-set structure. *Psychol. Rev.* **120**, 190 (2013).
25. B. Hommel, Event files: Evidence for automatic integration of stimulus-response episodes. *Vis. Cogn.* **5**, 183-216 (1998).
26. V. Wyart, V. de Gardelle, J. Scholl, C. Summerfield, Rhythmic fluctuations in evidence accumulation during decision making in the human brain. *Neuron* **76**, 847-858 (2012).
27. E. Marcos, S. Tsujimoto, M. Mattia, A. Genovesio, A Network Activity Reconfiguration Underlies the Transition from Goal to Action. *Cell Rep.* **27**, 2909-2920.e2904 (2019).
28. K. Duncan, B. B. Doll, N. D. Daw, D. Shohamy, More Than the Sum of Its Parts: A Role for the Hippocampus in Configural Reinforcement Learning. *Neuron* **98**, 645-657.e646 (2018).
29. I. C. Ballard, A. D. Wagner, S. M. McClure, Hippocampal pattern separation supports reinforcement learning. *Nat. Commun.* **10**, 1073 (2019).
30. B. Elsner *et al.*, Linking actions and their perceivable consequences in the human brain. *Neuroimage* **17**, 364-372 (2002).
31. T. Melcher, M. Weidema, R. M. Eenshuistra, B. Hommel, O. Gruber, The neural substrate of the ideomotor principle: An event-related fMRI analysis. *NeuroImage* **39**, 1274-1288 (2008).
32. M. Siegel, T. J. Buschman, E. K. Miller, Cortical information flow during flexible sensorimotor decisions. *Science* **348**, 1352-1355 (2015).
33. A. Woolgar, J. Jackson, J. Duncan, Coding of visual, auditory, rule, and response information in the brain: 10 years of multivoxel pattern analysis. *J. Cognit. Neurosci.* **28**, 1433-1454 (2016).
34. M. Rigotti, D. D. Ben Dayan Rubin, X.-J. Wang, S. Fusi, Internal representation of task rules by recurrent dynamics: the importance of the diversity of neural responses. *Front. Comput. Neurosci.* **4**, 24 (2010).
35. O. Barak, M. Rigotti, S. Fusi, The sparseness of mixed selectivity neurons controls the generalization–discrimination trade-off. *J. Neurosci.* **33**, 3844-3856 (2013).
36. A. Bhandari, C. Gagne, D. Badre, Just above Chance: Is It Harder to Decode Information from Prefrontal Cortex Hemodynamic Activity Patterns? *J. Cogn. Neurosci.* **30**, 1473-1498 (2018).
37. S. Fusi, E. K. Miller, M. Rigotti, Why neurons mix: high dimensionality for higher cognition. *Current opinion in neurobiology* **37**, 66-74 (2016).

38. I. R. Olson, K. Page, K. S. Moore, A. Chatterjee, M. Verfaellie, Working memory for conjunctions relies on the medial temporal lobe. *J. Neurosci.* **26**, 4596-4601 (2006).
39. B. Elsner, B. Hommel, Effect anticipation and action control. *J. Exp. Psychol. Hum. Percept. Perform.* **27**, 229 (2001).
40. M. X. Cohen, *Analyzing Neural Time Series Data: Theory and Practice* (MIT Press, 2014), pp. 600.

Figures Captions

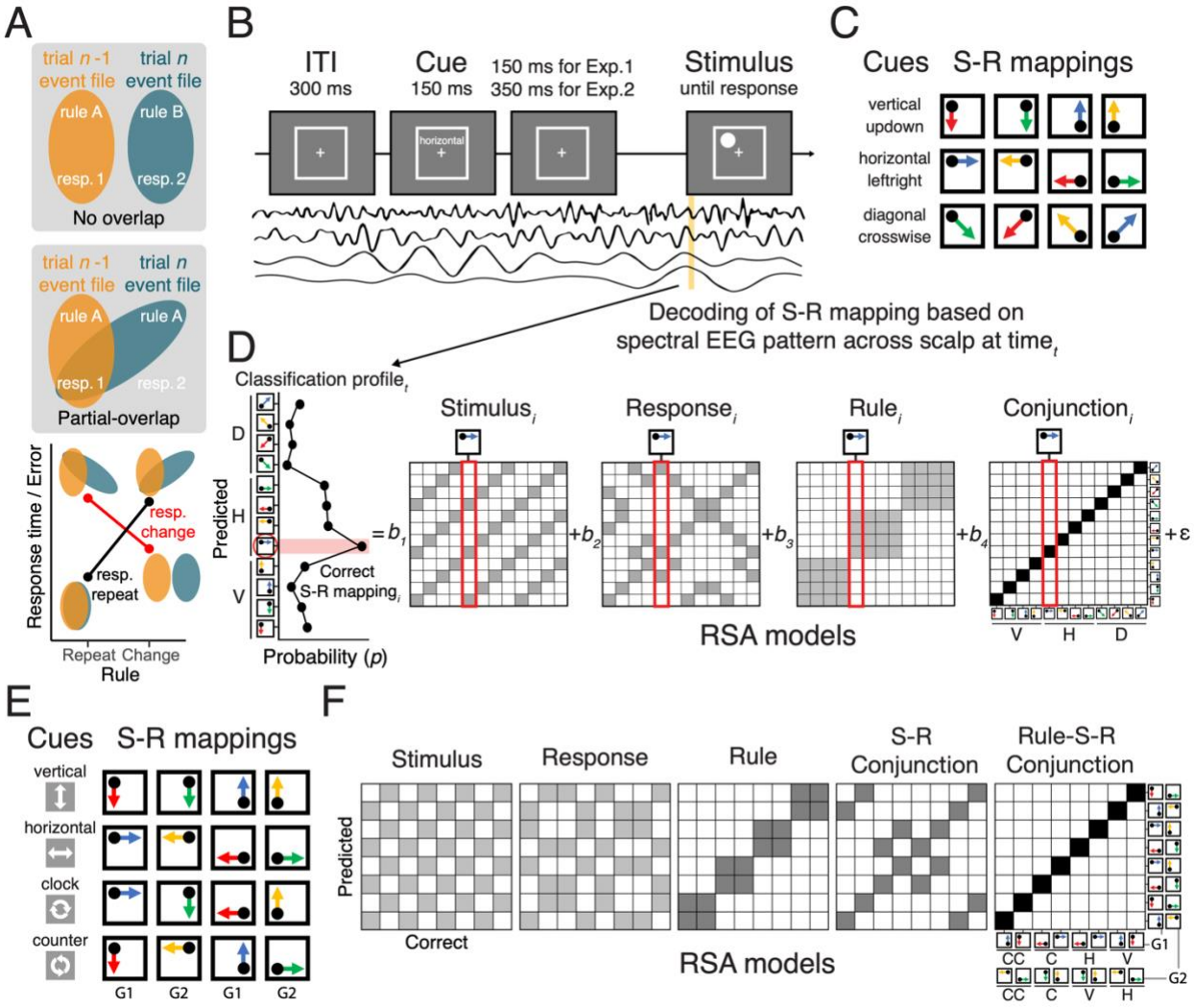

Fig.1. (*A*) Event files and partial-overlap priming pattern. Shown are two examples from simplified situations with two possible action rules and response options, along with the idealized data pattern. The trial *n*-1 event file is shown as the yellow oval, the trial *n* event file as the green oval. In the first example, consecutive event files do not overlap, which, just as when there is complete overlap (not shown), allows efficient performance. In the second example, the response representation overlaps, but the rule does not, leading to partial-overlap costs—just as would be the case for rule overlap, but non-overlapping responses (not shown). Note, that examples of all possible trial-to-trial transition types in the current paradigm are embedded in Figure 2. (*B*) Sequence of trial events in the rule-selection task. (*C*) Spatial translation rules mapping specific stimuli to responses in Experiment 1 (horizontal, vertical, and diagonal rule). Two different cue words were used for each rule. Responses were made on four keys, each spatially compatible with one of the four possible stimulus locations. (*D*) Schematic steps of the representational similarity analysis. For each sample time (*t*), a scalp-distributed pattern of EEG power (see *Representational Similarity Analysis* in *Methods*) was used in a first step to decode trial-specific rule/stimulus/response configurations, producing classification probabilities for each of the possible configurations. In the second step, these classification profiles for each trial and time-point were simultaneously regressed onto model vectors for the potentially relevant representation. The figure shows all

possible vectors as model matrices (the x-axis represents the correct constellations for the decoder to pick, and the y-axis the “confidence” with which each constellation is predicted). For each specific trial, the “vertical” vectors corresponding to the relevant action constellation are picked as predictors (i.e., the red boxes show all four vectors for one specific action constellation; the shading of matrix elements indicates the predicted classification probabilities with darker shadings representing higher probabilities). Note, that the conjunction matrix arises from multiplying corresponding elements of the constituent feature matrices with each other. The idealized classification profile represents an example where a unique conjunction and rule information are expressed as a peak at the correct label of the S-R mapping$_i$ and some confusion to other instances with the same rule. The coefficients associated with each predictor (i.e., in terms of *t*-values) reflect the unique variance explained by each of the constituent features and their conjunction. (*E*) Spatial translation rules mapping specific stimuli to responses in Experiment 2. A word or a symbol were used as a cue for each rule. As described in *Methods*, to achieve complete orthogonalization, rule-S-R constellations were divided into two groups of eight constellations each (G1 and G2) that were separately analyzed and then averaged. (*F*) Models for the representational similarity analysis (RSA) used within each of the two 8-constellation groups in Experiment 2.

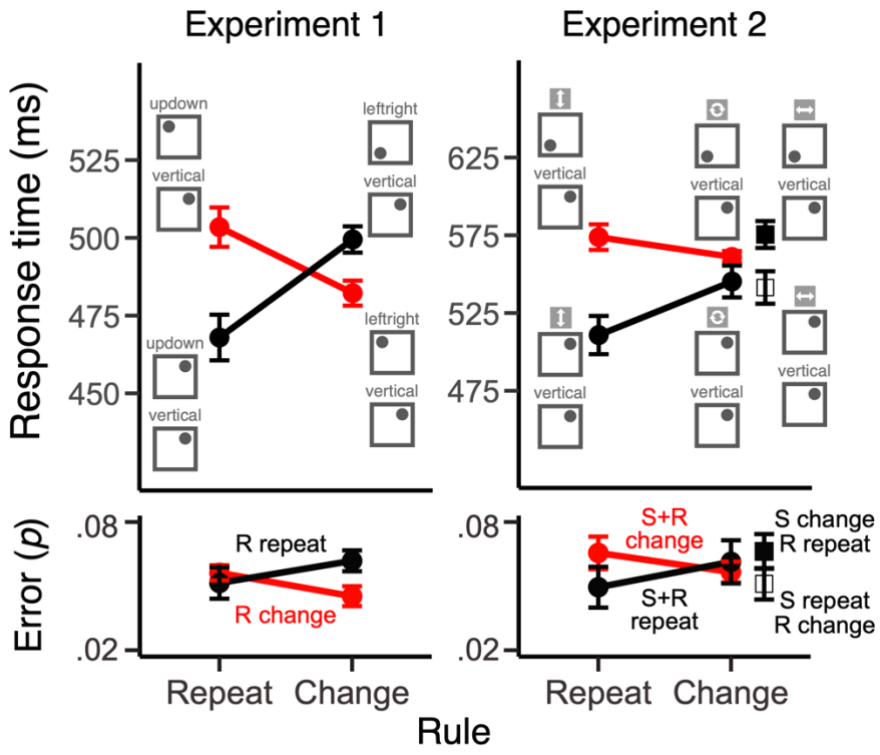

Fig.2. Mean response times (RTs) and errors as a function of all possible transition types in Experiments 1 and 2. Alongside with each data point, examples for the corresponding trial-to-trial transition type are shown (i.e., trial *n*-1=top, trial *n*=bottom). Experiment1 allows transitions resulting from crossing the rule repetition/change factor and the response repetition/change factor. Experiment 2 also allowed differentiating pure stimulus changes and response changes in case of rule-change trials. However, for subsequent analyses on the relationship between conjunctive representations and overlap costs (see Fig.4D), we focused on the transitions with complete S-R changes or repetitions. Error bars specify 95% within-subject confidence intervals.

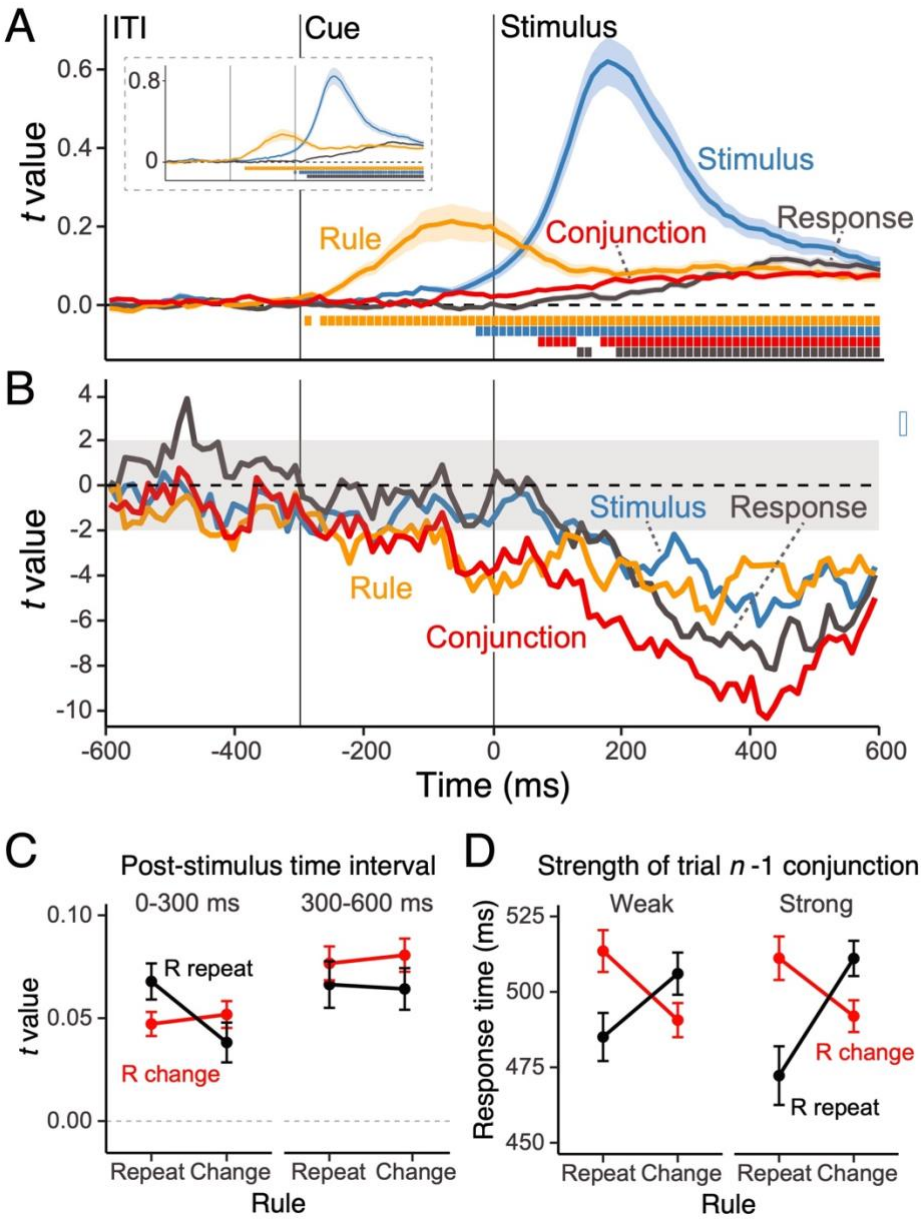

Fig.3. EEG-related results for Experiment 1. (*A*) Average, single-trial *t*-values associated with each of the basic features and their conjunction derived from the RSA analysis (see Fig.1C). Shaded regions specify the standard error around the mean. The colored squares at the bottom of the figure denote the significant time-points using a non-parametric permutation test. The insert shows the same RSA fit scores when the conjunction was not included as predictor in the RSA analysis. (*B*) Time-course of *t*-values from multilevel, linear models predicting the variability in trial-to-trial RTs (i.e., the impact of representations on behavior), using RSA scores of all features as the simultaneous predictors. (*C*) Average RSA scores of the conjunction model as a function of rule repetition/change and the response repetition/change factors for the early (0-300 ms) and the late (300-600 ms) periods in the post-stimulus interval. (*D*) Modulation of RT partial-overlap costs in trial *n* as a function of the strength of conjunction codes (median split) in trial *n*-1. Error bars specify 95% within-subject confidence intervals.

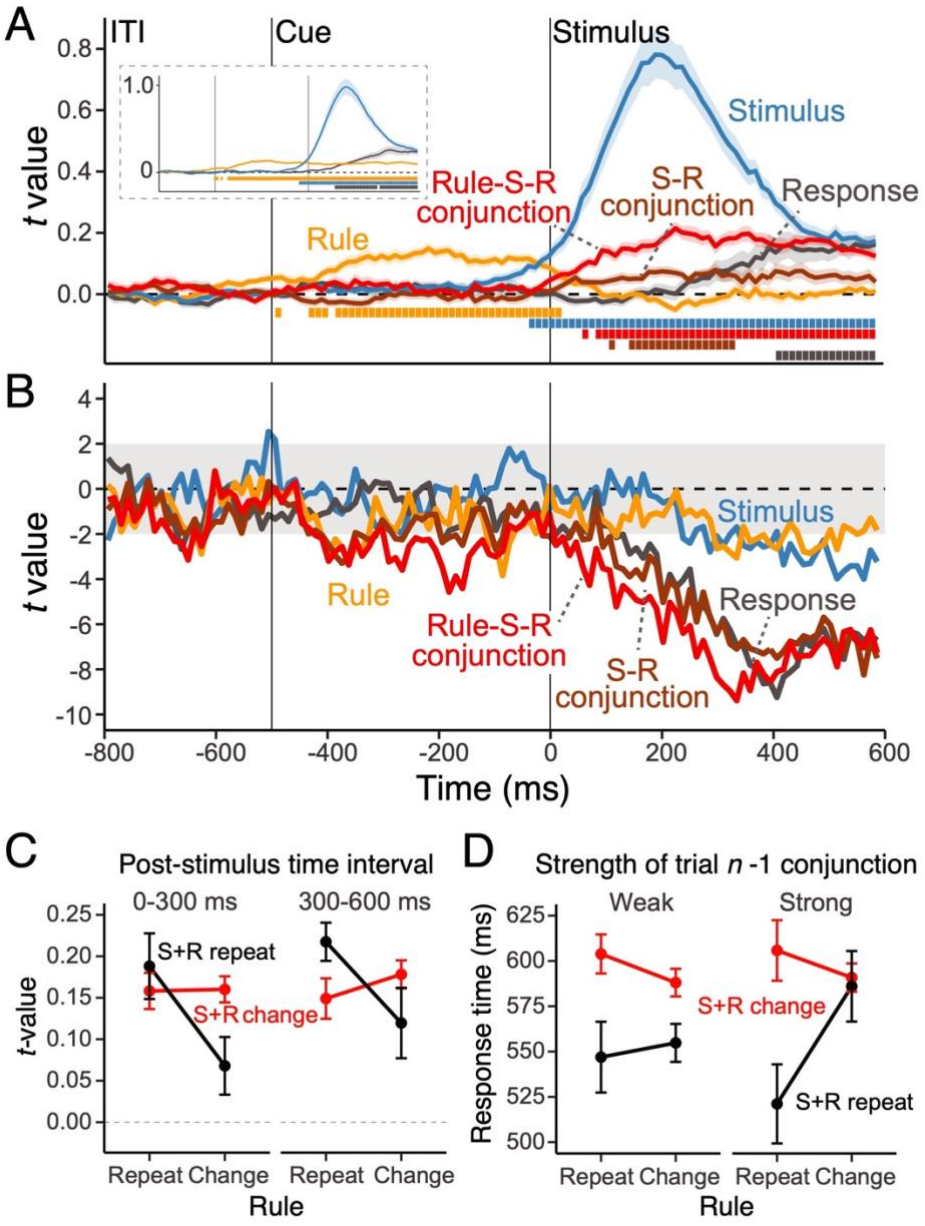

Fig.4. EEG-related results for Experiment 2. See caption to Fig.3 for a description of panels A-D. Different than Experiment 1, this experiment allowed differentiation between S-R and rule-S-R conjunctions.