Title
Common and distinct neural bases for rule- and similarity-based category learning

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 45(45)

Authors
LI, Christy
Sun, Yinghao
Deng, Sophia W

Publication Date
2023

Peer reviewed

Common and distinct neural bases for rule- and similarity-based category learning

Jianhua Li (yb97301@um.edu.mo)
Department of Psychology, Faculty of Social Sciences, University of Macau
Center for Cognitive and Brain Sciences, University of Macau
Taipa, Macau S.A.R. 999078 China

Yinghao Sun (bc00511@um.edu.mo)
Faculty of Business Administration, University of Macau
Taipa, Macau S.A.R. 999078 China

Sophia W. Deng* (wdeng@um.edu.mo)
Department of Psychology, Faculty of Social Sciences, University of Macau
Center for Cognitive and Brain Sciences, University of Macau
Taipa, Macau S.A.R. 999078 China

Abstract

Category learning is a core competence for minimizing cognitive load and optimizing decision-making. An identical problem can be solved by employing a rule-based or a similarity-based strategy. This work examined whether the use of the two strategies was supported by common or distinct neural substrates. We conducted a category learning experiment with rule-plus-similarity stimuli using EEG-fNIRS fusion methodology. Participants learned two artificial categories using either a rule-based or similarity-based strategy. The results showed a common visual-perceptual-analysis process and distinct decision-making processes between the uses of the two strategies. Larger P300 and N400 amplitudes and Wernicke's area activation indicated that hypotheses testing and verbal rule abstraction processes were critical for rule-based categorization. In contrast, increased frontopolar cortex activity indicated that integration of multiple dimensions was critical for similarity-based categorization. These results were consistent with COVIS theory, implying an explicit system in rule-based category learning whereas an implicit system in similarity-based learning.

Keywords: EEG-fNIRS fusion; rule-based category learning; similarity-based category learning; COVIS

Introduction

Categorization is a core component in human cognition that organizes things into equivalent classes. It reduces cognitive load and promotes decision making. For example, when giving a diagnosis report, a physician may classify a patient into a known disease category based on a diagnostic rule or based on multiple specific symptoms similar to those previously seen in other patients. While applying a rule is a verbal process that involves semantic analysis and relies on working memory, assessing the similarity of symptoms to previous patients is a nonverbal process that requires access to the connections between multiple exemplars. It seems a categorization problem can be solved with two critical types of strategies: the rule-based strategy and the similarity-based strategy. Using different strategies implies that there may be variety in neural substrates for different category boundaries in human categorization (Ashby et al., 1998), but the same kind of decision output could also indicate that different strategies may be supported by a common neural system.

Correspondingly, while some evidence points to the possibility of multiple learning systems (Ashby et al., 1998; Ashby & Maddox, 2011), others claim that it can be explain by a single system (Edmunds et al., 2015). As a result, there is currently no clear consensus on this issue: whether dual-learning systems support the rule-based and similarity-based learning strategies with separate neural substrates or if there is any overlap. Since semantic processing, working memory, and cognitive control processes can be characterized by cortex activation and brain potentials (Morrison et al., 2015), in the current work, we investigated the common and distinct neural substrates between rule-based and similarity-based category learning strategies by assessing electrophysiological and hemodynamic responses.

Single versus Multiple System

Several models of human category learning have been developed, suggesting that single or multiple learning systems are possible. Specifically, rule-based models (e.g., RULEX—rule-plus-exception model, Nosofsky et al., 1994) posit that people classify items by developing logical rules and occasional exceptions to those rules, while exemplar-based models (e.g., GCM—generalized context model, Nosofsky, 1986) suggest that categorization decisions are dependent on the similarity of stimuli to the stored exemplars. There is also evidence consistent with multi-process models. For example, the Competition between Verbal and Implicit Systems (COVIS; Ashby et al., 1998; Ashby & Maddox, 2004, 2011) indicates that the explicit system dominates the learning of verbalizable, rule-based category structures and the implicit system dominates the learning of non-verbalizable, information-integration category structures. There is no consensus on this issue. A complementary approach is to use the brain dynamic method to investigate whether there are neurally dissociable processes, since single-system accounts fail to explain the evidence that separable neural systems are engaged during different types of category learning (Carpenter et al., 2016).

In M. Goldwater, F. K. Anggoro, B. K. Hayes, & D. C. Ong (Eds.), *Proceedings of the 45th Annual Conference of the Cognitive Science Society.*

COVIS claims that human category learning requires at least two systems (Ashby et al., 1998; Ashby & Maddox, 2004, 2011). The explicit/verbal system posits that rule-based strategy depended on a hypothesis-testing approach, in which executive attention, working memory, and semantic processing were crucial for choosing, testing, and switching different hypotheses. The implicit/procedural-learning system is mediated by an implicit system that is difficult to describe verbally and requires integration across multiple input dimensions (Ashby & Maddox, 2004, 2011). Such non-declarative system was employed implicitly in similarity-based category learning, especially for category structure defined by the overall similarity to category examples or prototypes (Koenig et al., 2005; Milton et al., 2017).

Neural Distinction between Rule-Based and Similarity-Based Category Learning

To the best of our knowledge, no research has directly compared the neural responses to rule-based and similarity-based category learning using the same stimuli sets with identical category structure. Looking into the research using different stimuli, there is still debate over whether dual-learning systems supported the rule-based and similarity-based learning strategies with different neural substrates (Ashby & Maddox, 2004, 2011; Carpenter et al., 2016). Some research suggested that the rule-based and similarity-based category learning may share brain substrates (Carpenter et al., 2016; Milton et al., 2017). However, some studies revealed different neural correlates (Koenig et al., 2005). These disagreements could be attributed to the different sets of stimuli to encode, hence in the present study, we applied the rule-plus-similarity structure (Deng & Sloutsky, 2015, 2016) to directly compare the neural responses to rule-based and similarity-based category learning.

Event-related potentials (ERPs) are directly associated to neuronal-electrical-activity and provide excellent temporal resolution for identifying cognitive processes including attention, working memory, and semantic processing (Kayser & Tenke, 2015). As such, based on the cognitive processes revealed by the theoretical frameworks of category learning (e.g., COVIS), a few ERP components could be revealed: (1) An P1 component, a positive deflection peaked in posterior electrodes around 100ms post-stimulus, is associated with early visual processing or encoding (Liang et al., 2007) as well as attention modulation (Fu et al., 2005; Luo et al., 2001), with higher P1 amplitude indicating increased attentional effort. (2) A P300 component, a positive deflection peaked around 300ms, indexing the working memory, especially the rule information updating in memory (Morrison et al., 2015). (3) An N400 component, a negative deflection in the 200-600ms time window, related to semantic processing (Kutas & Federmeier, 2011), or acquisition of abstract rules (Sun et al., 2012).

Brain region activity involved in category learning could be measured by functional near-infrared spectroscopy (fNIRS) simultaneously to investigate the frontal and temporal activity, which are shown related to working memory and semantic processing, respectively. fNIRS measures blood's intrinsic optical absorption utilizing near-infrared light from a source to scalp detector probes (Lloyd-Fox et al., 2010). Based on the amount of light absorbed, concentration changes of oxygenated and deoxygenated hemoglobin can be deduced and are thought to be comparable to fMRI (Cui et al., 2011). By concurrently recording the EEG and fNIRS, both the temporal signatures and spatial activations in the brain would be depicted and the relatively limited spatial resolution of EEG can be compensated.

Present Experiment

To investigate whether rule- and similarity-based category learning was supported by common and distinct cognitive processes, an experimental study with EEG and fNIRS devices recorded concurrently was conducted using the same stimuli set. Specifically, we investigated (1) if rule-based and similarity-based category learning had the same or different P1, P300, and N400 components; (2) whether the left frontal and temporal lobes were active differently. Based on the dual-system processing theory, we hypothesized that distinctive cognitive processes could be observed in working memory, and semantic processing related neural components or brain activation. Compared to similarity-based category learning, rule-based category learning required working memory and verbal system to evaluate hypotheses, eliciting a larger P300 and N400 and activating functionally supported brain regions.

Method

Participants

We recruited 73 healthy, right-handed, normal, or corrected-to-normal vision volunteers (40 females; Mean: 20.81 years, SD: 2.37 years). All of them were naive about the hypotheses of the study and were tested in a sound-attenuated and electrostatically shielded room in the psychology laboratory.

Behavioral data of all 73 participants were analyzed (39 rule-based and 34 similarity-based). Eight neural data were excluded from the analyses because of excessive ocular artifacts, or bad data quality, leaving 65 participants in the final analyses (34 rule-based, 31 similarity-based).

Materials

Materials were similar to those used previously by Deng and Sloutsky (2015) and consisted of colorful drawings of artificial creatures. These creatures were accompanied by the novel labels, *lulu* and *momo*. These categories had two prototypes (Lulu0 and Momo0, respectively) that were distinct in the color and shape of seven of their features: body, hands, feet, head, tail, antennae, and a neck button (see Figure 1). With the variations in these seven features, two sets of stimuli were formed.

The High-match items and Switch items were stimuli that included five of seven features from one prototype and two features from the other (i.e., five features of Lulu0 plus two features of Momo0, or five features of Momo0 plus two features of Lulu0, forming the seven features of each item). For the rule-based category learning group, the classification of stimuli depended on the deterministic feature (neck button)—the stimulus with a rain-drop button was *lulu* and the stimulus with a cross button was *momo*. For the similarity-based group, they made decisions based on the overall similarity among the exemplars, five out of seven features from the Lulu0 would be *lulu* and five features from the Momo0 would be *momo*. Since each stimulus had seven features, switching the two features results in 42 variants, including 30 High-match items and 12 Switch items, half of which were *lulu*, and the other half were *momo*.

The New-feature items and Filler items were stimuli whose five of seven features were from one prototype and the other two features were new, which were not from Lulu0 or Momo0 (i.e., five features of Lulu0 plus two new features, or five features of Momo0 plus two new features, forming the seven features of each item). For the similarity-based category learning group, the classification of stimuli depended on the similarity with Lulu0 or Momo0. *Lulus* were items with five Lulu0 features and two new features, whereas *momos* were items with five Momo0 features and two new features. For the rule-based group, the classification of stimuli would be determined if their deterministic feature were from Lulu0 (rain-drop button) or Momo0 (cross button), whereas there would be no correct response for those stimuli whose deterministic feature is a new feature (Filler items). These filler items were designed (1) to avoid participants inferring the classification of stimuli during testing (if all other features were changing but the deterministic feature stayed static, participants may give the feature a salient status during testing); (2) to ensure that both learning groups were presented the identical stimuli (this manipulation is critical for the rule-based group to avoid learning by inference in testing); and (3) to verify the effectiveness of strategy manipulation (the filler trials can assistant us to identify whether the deterministic feature differs from the other features in the rule-based group, but it has no significant status for the similarity-based group). Each stimulus had seven features, five features of Lulu0 or Momo0 and two new features resulted in 42 variants, including 30 New-feature items and 12 Filler items, of which half were *lulu* and the other half were *momo*.

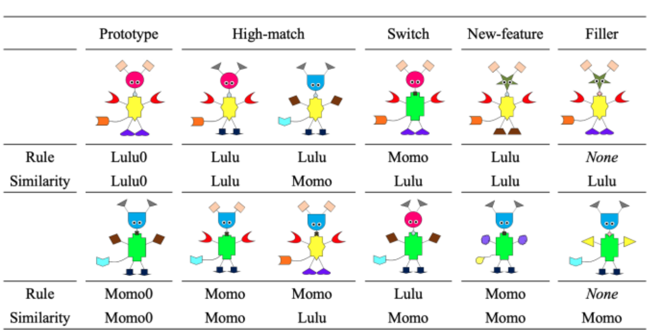

Figure 1: Examples of stimuli.

Procedure

The experiment consisted of instructions, training, and testing. The training was a between-subject factor, with participants being presented with either rule- or similarity-based categorization training. Instructions and testing were identical for both training groups. The experiment was presented on the computer and controlled by E-prime 2.0.

Instructions and Training For both rule-based and similarity-based category learning groups, all features of Lulu0 and Momo0 were given to participants explicitly one by one before training. In the rule-based categorization group, they were told that all *lulu* (or *momo*) had a raindrop-shaped (or a cross-shaped) button and most of the *lulu* (or *momo*) had the other features (body, hands, feet, head, tail, and antennae, one at a time). This information was repeated in the corrective feedback on each trial during training using the following script: *This one is lulu (or momo), and it has the lulu's (or the momo's) button*. In the similarity-based categorization group, all the features (body, hands, feet, head, tail, antennae, neck button) were presented one at a time, and participants were told that most of the *lulu* (or *momo*) had this feature. The corrective feedback was provided by: *This one is lulu (or momo), and it looks like lulu (or momo)*. Testing was not mentioned during the training phase. Participants were randomly assigned to one of the two training conditions. Participants in both groups were asked to classify the stimuli into the *lulu* and *momo* categories (one stimulus was provided in each trial). They were given 30 training trials (randomly chosen from the High-match items) and each trial was accompanied by corrective feedback.

Testing The testing phase was administered immediately after training, and it was identical for both groups. During the testing phase, 30 High-match items, 12 Switch items, 30 New-feature items, and 12 Filler items were presented three times randomly in three blocks. No feedback was provided.

Behavioral Data Analysis

The 30 training trials were divided into three learning blocks, and only the data whose accuracy in the final block above the chance threshold of 0.5 were retained to ensure that the subject had understood the task. To verify the effectiveness of strategy manipulation, linear mixed effect models were performed on the RT and ACC of testing trials, with the subject as a random factor, learning strategy (rule-based, similarity-based), stimuli conditions (High-match, Switch, New-feature, Filler), and their interaction were entered into as fixed factors, using the *lmer*, *glmer*, and *ANOVA* functions in the *lme4* and *lmerTest* package in R (Brown, 2021). It should be mentioned that even though rule-based learners were instructed to categorize based on the deterministic features, they might not completely disregard other probabilistic features (Deng & Sloutsky, 2016), and thus they would perceive that if the similarity of stimuli was significantly altered. As a result, since the deterministic feature (neck button) should differ from other features for the rule-based learners but had no special status for similarity strategy learners, we predicted that rule-based learners should perform differently on High-match and Switch items and on New-feature and Filler items, whereas the similarity-based learners should not.

EEG Recording and Analysis

EEG and fNIRS data were collected simultaneously using an EasyCap. For EEG acquisition, 32 electrodes were arranged based on the international 10/20 system (Figure 2, green) and digitized at 1000 Hz. During the online acquisition, the left mastoid electrode was used as a reference, and the impedance of all electrodes was kept below 20 kΩ.

For offline analysis in EEGLAB v2021.0 (Delorme & Makeig, 2004) in MATLAB. Continuous EEG data were first down-sampled to 250 Hz, re-referenced to the grand average of all channels, and filtered with a band pass of 0.1-30 Hz. Bad channels were interpolated by averaging the spherical electrodes, which took up less than 2% of all the channels. Then epochs lasting from 200 ms before the target onset to 1000 ms afterwards were extracted. Eye movement-induced muscle activity was removed from the segmented data by the Independent Component Analysis (ICA) algorithm and *ADJUST* automatic classification algorithm (Mognon et al., 2011), and manual screening on topographical distribution were employed. The EEG signal exceeded ±75μV in amplitude during the epochs was then deleted.

In light of the COVIS theoretical framework, previous studies (Chen et al., 2015) and the wave of the current results, the mean P1, P300, and N400 amplitudes of selected electrodes were measured during an 80-120 ms, 200-440 ms, and 200-500 ms time window after the stimuli onset, respectively. To improve the statistical power and reduce false effects (Luck & Gaspelin, 2017), the average values of FP1, AFF5h, FC1, FP2, F4, FCz, and FC2 electrodes were collapsed as an anterior activity indication, and the averages of P3, Pz, P4, O1, Oz, and O2 electrodes were collapsed as a posterior activity indication.

It should be noted that Switch items and Filler items were presented to ensure that participants of both groups precepted the same stimuli. There weren't enough Switch or Filler trials to study their brain signals and comparing responses to them did not provide further information on our primary issue. As a result, P1, P300, and N400 amplitudes were then exported in linear mixed effects models, with random intercepts for the subject, learning strategy, condition (High-match, New-feature), and their interaction as fixed factors.

fNIRS Data Acquisition and Data Analysis

Eight sources and 8 detectors were placed in the left frontal and temporal cortex, thus forming 22 fNIRS channels (Figure 2; 8 sources, orange; 8 detectors, blue; 22 channels, purple line). Each source transmitted LED lights at the wavelength of 760 nm and 850 nm. The distance between the source and detector was maintained at 3 cm. Optical signals were sampled at 7.81 Hz. The MNI coordinates of optodes and channels were obtained from their spatial information at the international 10/20 system, which was then imported to NIRS_SPM to generate anatomical labels and percentage of overlap (Ye et al., 2009). Channels 2, 5, 7, 8, 9, 10, 12, and 13 reflected the contributions of the dorsolateral prefrontal cortex (dlPFC); channels 1, 2, 3, 4, 6, and 7 were located over the frontopolar cortex (FPC); channels 12 was located over the Broca's area; channel 16 and 17 were located over the middle temporal gyrus (MTG); channel 14, 18, and 19 were located over the pre-motor and supplementary motor cortex (MC); channels 15 and 16 were located over the superior temporal gyrus (STG); and channels 21 and 22 were located over the supramarginal gyrus part of Wernicke's area. In sum, seven regions of interest: dlPFC, FPC, Broca's area, MTG, MC, STG, and Wernicke's area, were selected for analysis.

The pre-processing of fNIRS data was performed with nirsLAB (Xu et al., 2014). Motion artifacts were removed from the raw data by the built-in algorithm, which was subsequently filtered with a band pass of 0.01-0.2 Hz. Modified Beer-Lambert Law was then used to calculate hemodynamic states. Changes in oxygenated hemoglobin (HbO), deoxygenated hemoglobin (HbR), and total hemoglobin (HbT) were modeled with the canonical HRF function in the Level 1 module of statistical parametric mapping. As a result, general linear model (GLM) coefficients (beta values) were calculated for High-match and New-feature conditions for each participant.

Although both HbO and HbR beta values were obtained from the GLM estimations, only HbO signals were compared, since it is a more sensitive indicator of changes in regional cerebral blood flow (Fu et al., 2014). To illustrate the activation for using the rule-based strategy and the similarity-based strategy, an independent-sample *t*-test on HbO beta values was used to compare across two groups by the activation averaged across channels for each region of interest (seven regions of interest, dlPFC, FPC, Broca's area, MTG, MC, STG, and Wernicke's area). Since the purpose of Switch and New-feature trials was to verify the effectiveness of learning strategy manipulation, we compared the brain activation between two groups of High-match condition and New-feature conditions.

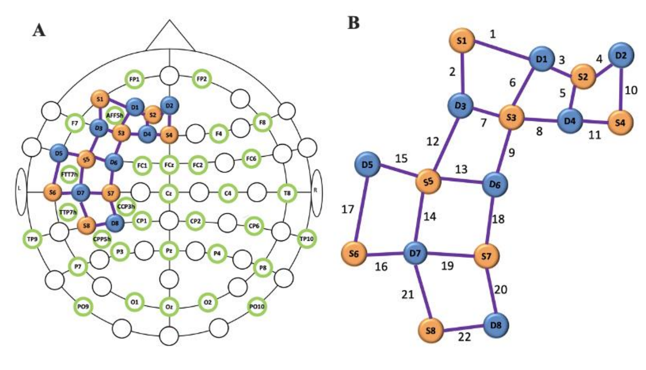

Figure 2: A. The layout of fused EEG-fNIRS setup. B. The layout of fNIRS channel positions.

Results

Behavioral Results

For the training data, for RT, the significant interaction between the learning strategy and block ($p < .001$) revealed no difference between the three blocks for the similarity-based group ($p > .382$), whereas RT of the rule-based group of block1 was significantly longer than the other blocks ($ps < .001$) and no difference was found between block2 and block3 ($p = .435$). For ACC, no difference was found between the three blocks in the rule-based group ($ps > .050$) and the similarity-based group ($ps > .409$). ACC of the last training block of all the participants was above 0.5 (rule = 0.913 ± 0.161; similarity = 0.725 ± 0.209).

To verify the effectiveness of strategy manipulation, linear mixed-effect models were used on testing RT and ACC, showed significant interactions ($ps < .001$). We primarily reported the difference between the two groups on High-match and Switch items, and on New-feature and Filler items. Consistent with our prediction, participants in the rule-based group (1) showed a lower accuracy ($p < .001$) to Switch trials (0.840 ± 0.271) than to High-match trials (0.926 ± 0.114), whereas no difference was found in the similarity-based group ($p = .277$, Switch: 0.770 ± 0.269; High-match: 0.792 ± 0.18); (2) responded slower and with a lower accuracy (ps < .001) to Filler trials (RT: 1601 ± 660 ms; ACC: 0.865 ± 0.222) than to the New-feature trials (RT: 960 ± 444 ms; ACC: 0.968 ± 0.074), whereas no difference was found in the similarity-based group ($ps \geq .340$, Filler RT: 2195 ± 1527 ms, ACC: 0.865 ± 0.217; New-feature RT: 2113 ± 1278 ms, ACC: 0.874 ± 0.180).

In sum, our manipulation for learning strategies was successful as our results revealed that the deterministic feature was learned differently from other features in the rule-based learning group, while for the similarity-based group, these features are not specific.

ERP Results

Figure 3 depicts the grand-averaged ERP waveforms of selected brain areas and scalp topographies of P1, P300, and N400 in the High-match and New-feature conditions.

For P1, Figures 3B and 3D showed that the P1 component was observed in the posterior brain region, without main effects or interaction effect ($ps > .229$). For P300, Figures 3B and 3D showed a large positive wave with maximum amplitude at 200 to 440 ms after stimulus onset in the posterior brain region. The main effect of the learning strategy ($p = .030$) showed that the amplitude of the rule-based (1.80 ± 2.26 μV) group was higher than the similarity-based (0.62 ± 2.33 μV) group. The N400 effects were observed in the anterior brain region, in Figures 3A and 3C. A marginal main effect of learning strategy ($p = .085$) showed that the N400 amplitude of the rule-based (-2.16 ± 2.15 μV) was marginally larger than the similarity-based (-1.27 ± 2.24 μV) group. No interaction was observed.

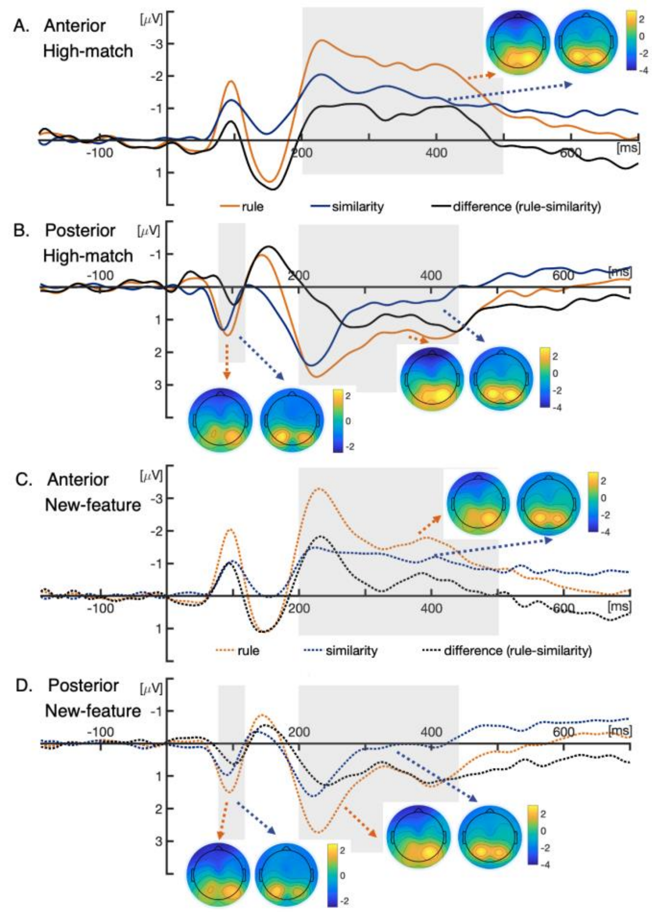

Figure 3. The grand-averaged waveform of the rule-based group (orange), similarity-based group (blue), and difference between them (black, rule minus similarity). EEG component windows were indicated by shaded grey vertical bars and their topographies were plotted.

Our analysis showed that (1) the ERP response to High-match and New-feature stimuli showed a similar pattern, for which the P1 (80-120 ms), P300 (200-440 ms), and N400 (200-500 ms) components were observed; (2) an identical component, P1, was revealed in two learning groups; (3) distinct components, P300 and N400, were observed, with the average P300 and N400 components of the rule-based group were higher than the similarity-based learning group.

fNIRS Results

As shown in Table 1, in both conditions, the similarity-based group was likely to activate the FPC more whereas the rule-based group was likely to activate the Wernicke's area more.

Similar to the ERP results, our fNIRS data revealed the same pattern in High-match and New-feature conditions, and the difference between rule-based and similarity-based learning groups showed in the FPC and Wernicke's area.

Table 1: Independent sample *t*-tests for HbO beta values across rule- and similarity-based learning groups.

Condition	ROI	Comparisons		
		t	*p*	Cohen's *d*
High-match	dlPFC	-0.423	0.675	0.105
	FPC	**2.477***	0.018	0.639
	Broca's area	0.524	0.603	0.133
	MTG	0.529	0.600	0.135
	MC	0.281	0.780	0.070
	STG	0.320	0.750	0.079
	Wernicke's area	**1.747#**	0.091	0.454
New-feature	dlPFC	0.509	0.613	0.123
	FPC	**-1.88#**	0.066	0.477
	Broca's area	0.926	0.358	0.229
	MTG	0.632	0.530	0.159
	MC	1.155	0.253	0.281
	STG	0.465	0.644	0.113
	Wernicke's area	**2.025#**	0.051	0.524

***** $p < .05$, **#** $p < .1$

Discussion

In our study, we evaluated ERP responses and hemodynamic activations to the same stimuli sets of individuals who were manipulated to use rule-based and similarity-based category learning strategies. Behavioral results demonstrated the success of strategy manipulation. ERP responses showed that a similar P1 component, but the rule-based group showed a larger P300 and N400 amplitudes than similarity-based category learning group. The distinct activations between two learning groups showed in the frontopolar cortex and Wernicke' area, with similarity-based group exhibiting higher frontopolar cortex activation and lower Wernicke' area activation. These findings indicated the common and distinct cognitive processes between rule-based and similarity-based category learning strategies.

Object categorization has been regarded as a two-stage process: stimulus representations and processes for decision making. The first stage, visual feature processing could be reflected by early ERP components, such as P1, reflecting the early visual processing (Liang et al., 2007) and higher P1 amplitude indicating distributed visuospatial attention focusing (Fu et al., 2005; Luo et al., 2001). It seems more features were needed to be represented in the similarity-based category learning, compared to the rule-based learning. However, no difference was observed in the P1 component in the current research, which could be due to it was the supervised learning strategy and no difference in visuospatial attentional effort. Future research should examine the differences between unsupervised, internal, unintentional rule-based and similarity-based strategies (Love, 2002). Nevertheless, the current results showed shared visual feature processing or visuospatial attention allocation between rule-based and similarity-based category learning.

The second step, forming a membership criterion and making a choice, evidenced by late ERP components, the P300 and N400. Rule-based and similarity-based classification diverge in the anterior P300, reflects working memory changes (Polich, 2007). Rule-based categorization requires concentrating attention on a single dimension, whereas paying attention to several dimensions are required in similarity-based categorization. Besides, rule-based categorization had higher N400 amplitude, an index of the degree of the semantic processing (Kutas & Federmeier, 2011), and thus, is associated with rule learning or abstraction (Kutas & Federmeier, 2011; Sun et al., 2012).

The role of semantic processing in rule-based strategy use was also confirmed by our fNIRS results, showing that Wernicke's area was activated more. Perry and Lupyan (2014) found that Wernicke's area plays a key role linking stimuli and linguistic representations in categorization.

Moreover, in contrast to the rule-based, the similarity-based strategy integrates information into a category. Similarity-based strategies activated the FPC more. The level of FPC blood oxygen has been correlated with integrating the outcomes of several cognitive activities for optimizing behavioral goal (Ramnani & Owen, 2004). Converging with the implicit learning system of COVIS, similarity-based strategy elicited enhanced activation in the FPC, showing the integration across several input dimensions was required.

Conclusion

Since an identical problem can be solved by employing different strategies, we investigated the common and distinct cognitive processes between rule-based and similarity-based category learning by evaluating electrophysiological and hemodynamic responses. Participants were instructed to employ a rule- or similarity-based strategy. We compared the ERP responses and fNIRS activation to the same stimuli sets with a rule-plus-similarity structure. The common cognitive process was unpacked. A similar visual perceptual processing or visuospatial attention allocation process was involved, as evidenced by a similar P1 component. The distinct cognitive processes were extracted, which were consistent with COVIS theory. Compared with similarity-based category learning, larger P300 and N400 were elicited, and Wernicke's area was activated more, indicating that the hypotheses testing, and rule verbal abstraction processes were critical in rule-based categorical representation. In contrast, the frontopolar cortex (FPC) was engaged more in similarity-based category learning to integrate several dimensions into a categorical representation. In sum, the shared stimulus representation process and distinct decision-making processes were involved in rule- and similarity-based category learning, implying an explicit system and an implicit system in rule-based and similarity-based category learning, respectively.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (31900769) to S. W. Deng. Trip to the conference is partially supported by the Conference Grant for the Faculty of Social Sciences at University of Macau to S. W. Deng. We thank members of the Attention Brain and Cognitive Development Lab for their assistance in data collection and helpful comments.

References

Ashby, F. G., Alfonso-Reese, L. A., Turken, U, & Waldron, E. M. (1998). A neuropsychological theory of multiple systems in category learning. *Psychological Review*, *105*(3), 442–481.

Ashby, F. G., & Maddox, W. T. (2004). Human category learning. *Annual Review of Psychology*, *56*(1), 149–178.

Ashby, F. G., & Maddox, W. T. (2011). Human category learning 2.0. *Annals of the New York Academy of Sciences*, *1224*(1), 147–161.

Brown, V. A. (2021). An introduction to linear mixed-effects modeling in R. *Advances in Methods and Practices in Psychological Science*, *4*(1), 2515245920960351.

Carpenter, K. L., Wills, A. J., Benattayallah, A., & Milton, F. (2016). A Comparison of the neural correlates that underlie rule-based and information-integration category learning. *Human Brain Mapping*, *37*(10), 3557–3574.

Chen, Q., Liang, X., Li, P., Ye, C., Li, F., Lei, Y., & Li, H. (2015). The processing of perceptual similarity with different features or spatial relations as revealed by P2/P300 amplitude. *International Journal of Psychophysiology*, *95*(3), 379–387.

Cui, X., Bray, S., Bryant, D. M., Glover, G. H., & Reiss, A. L. (2011). A quantitative comparison of NIRS and fMRI across multiple cognitive tasks. *NeuroImage*, *54*, 2808–21.

Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*(1), 9–21.

Deng, W., & Sloutsky, V. M. (2015). The development of categorization: Effects of classification and inference training on category representation. *Developmental Psychology*, *51*(3), 392–405.

Deng, W., & Sloutsky, V. M. (2016). Selective attention, diffused attention, and the development of categorization. *Cognitive Psychology*, *91*, 24–62.

Edmunds, C. E., Milton, F., & Wills, A. J. (2015). Feedback can be superior to observational training for both rule-based and information-integration category structures. *Quarterly Journal of Experimental Psychology*, *68*(6), 1203–1222.

Fu, G., Mondloch, C. J., Ding, X. P., Short, L. A., Sun, L., & Lee, K. (2014). The neural correlates of the face attractiveness aftereffect: a functional near-infrared spectroscopy (fNIRS) study. *NeuroImage*, *85*, 363–371.

Fu, S., Caggiano, D. M., Greenwood, P. M., & Parasuraman, R. (2005). Event-related potentials reveal dissociable mechanisms for orienting and focusing visuospatial attention. *Brain Res Cogn Brain Res*, *23*(2-3), 341–353.

Kayser, J., & Tenke, C. E. (2015). Issues and considerations for using the scalp surface Laplacian in EEG/ERP research: A tutorial review. *International Journal of Psychophysiology*, *97*(3), 189–209.

Koenig, P., Smith, E. E., Glosser, G., DeVita, C., Moore, P., McMillan, C.,& Grossman, M. (2005). The neural basis for novel semantic categorization. *NeuroImage*, *24*, 369–383.

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology*, *62*, 621–647.

Liang, P., Zhong, N., Lu, S., Liu, J., Yao, Y., Li, K., & Yang, Y. (2007). The neural mechanism of human numerical inductive reasoning process: A combined ERP and fMRI study. *Web Intelligence Meets Brain Informatics*, 223–243.

Lloyd-Fox, S., Blasi, A., & Elwell, C. E. (2010). Illuminating the developing brain: the past, present and future of functional near infrared spectroscopy. *Neuroscience & Biobehavioral Reviews*, *34*(3), 269–284.

Love, B. C. (2002). Comparing supervised and unsupervised category learning. *Psychonomic Bulletin & Review*, *9*(4), 829–835.

Luck, S. J., & Gaspelin, N. (2017). How to get statistically significant effects in any ERP experiment (and why you shouldn't). *Psychophysiology*, *54*(1), 146–157.

Luo, Y. J., Greenwood, P. M., & Parasuraman, R. (2001). Dynamics of the spatial scale of visual attention revealed by brain event-related potentials. *Brain Res Cogn Brain Res*, *12*(3), 371–381.

Milton, F., Bealing, P., Carpenter, K. L., Bennattayallah, A., & Wills, A. J. (2017). The neural correlates of similarity- and rule-based generalization. *Journal of Cognitive Neuroscience*, *29*(1), 150–166.

Mognon, A., Jovicich, J., Bruzzone, L., & Buiatti, M. (2011). ADJUST: An automatic EEG artifact detector based on the joint use of spatial and temporal features. *Psychophysiology*, *48*(2), 229–240.

Morrison, R. G., Reber, P. J., Bharani, K. L., & Paller, K. A. (2015). Dissociation of category-learning systems via brain potentials. *Frontiers in Human Neuroscience*, *9*, 389.

Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General*, *115*(1), 39–61.

Nosofsky, R. M., Palmeri, T. J., & McKinley, S. C. (1994). Rule-plus-exception model of classification learning. *Psychological Review, 101*(1), 53–79.

Perry, L. K., & Lupyan, G. (2014). The role of language in multi-dimensional categorization: evidence from transcranial direct current stimulation and exposure to verbal labels. *Brain and Language*, *135*, 66–72.

Polich, J. (2007). Updating P300: an integrative theory of P3a and P3b. *Clinical Neurophysiology*, *118*(10), 2128–2148.

Ramnani, N., & Owen, A. M. (2004). Anterior prefrontal cortex: insights into function from anatomy and neuroimaging. *Nature Reviews Neuroscience*, *5*, 184–194.

Sun, F., Hoshi-Shiba, R., Abla, D., & Okanoya, K. (2012). Neural correlates of abstract rule learning: an event-related potential study. *Neuropsychologia*, *50*(11), 2617–2624.

Xu, Y., Graber, H. L., & Barbour, R. L. (2014, 2014/04/26). nirsLAB: A computing environment for fNIRS neuroimaging data analysis.*OSA Technical Digest (online)* Biomedical Optics 2014, Miami, Florida.

Ye, J. C., Tak, S., Jang, K. E., Jung, J., & Jang, J. (2009). NIRS-SPM: statistical parametric mapping for near-infrared spectroscopy. *NeuroImage*, *44*(2), 428–447.