

Evidence of Self – Referential Prioritization on the basis of Visual Features: Attributing Salience to Rule - Learning

Neelabja Roy (neelabja@iitk.ac.in)
Cognitive Science, IIT Kanpur

Harish Karnick (hk@cse.iitk.ac.in)
Cognitive Science, IIT Kanpur

Ark Verma (arkverma@iitk.ac.in)
Cognitive Science, IIT Kanpur

Abstract

Participants show faster and more accurate processing for arbitrary geometrical stimuli if they are paired with a self-relevant label (triangle = you). We ask whether participants only form self-associations with specific exemplars (triangle, circle, square), or whether they analyse the stimuli in terms of visual features, (for e.g. no. of vertices = 3), and can generalise the learned associations with the entire category of the stimuli (say, all triangles). In our experiments, participants showed the self – referential advantage not only to previously exposed exemplars of the same category, but also novel stimuli that could be categorised on the basis of similar visual features. Interestingly, they could generalise not only on the basis of a single rule, but also on the basis of a conjunction of more than one rule. The finding could be extended to explain social categorisation in the real world through group memberships

Keywords: self - reference, social categorisation, rule-learning, perception, social salience

Introduction

Decades of research has established that participants show processing benefits in terms of better memory, focused attention, and enhanced perceptual processing for self – referential stimuli, for e.g. self – name, self – face, self – referent adjectives, owned objects etc. (Cunningham & Turk, 2017). However, one could argue that the advantages in processing for these stimuli might be arising out of long-term familiarity and practice with these stimuli. More recently though, processing benefits have also been demonstrated when arbitrary geometrical stimuli (for e.g. triangle, square, circle) were linked to self (triangle = you) as compared to when these stimuli were linked with other socially relevant labels (circle = friend/mother/stranger) etc. (Sui et al., 2012). More specifically, Sui et al. (2012) asked participants to form and remember associations with arbitrary geometrical stimuli (for e.g. triangle = you, square = friend and circle = stranger), and later in a simple perceptual matching task, they asked participants to match the correct “social – labels” with the correct shape. Participants were found to be faster and more accurate, at matching shapes paired with the self – referent label (you), as compared to shapes linked with either the friend, stranger or even with mother labels. Sui et al. (2012) concluded that forming self – referent associations with even

arbitrary or neutral stimuli may modulate subsequent perceptual processing of the stimuli. These results were replicated and extended across a range of studies (Sui & Humphrey 2011; Sui, He & Humphrey 2012; Sui, Rotshtein & Humphrey, 2013).

Further, Sui & Humphreys (2015) suggested that self-reference may act as an integrating influence across perception, memory and decision making that may enhance or disrupt performance across task contexts. More specifically, based on results from a number of studies, Sui & Humphreys (2015) argue for the special role of self – reference in information processing, where it is asserted that self – reference may help bind memories with their source, increase perceptual integration, modulate the coupling between attention and decision making and also the interactions between various brain regions processing different aspects of task – relevant stimuli. It must be noted however, that these assumptions imply an almost global role of the ‘self – concept’ across various mental functions. Such results have however been challenged from quarters where the importance of the self – referential concept has been shown to be influenced by cultural variable. For e.g. Sparks, Cunningham, & Kritikos (2016) show that Asian participants displayed no ownership related biases in the accuracy of recognition memory for self – owned vs. other owned objects. The fact that the relative importance of the self – concept, and subsequently self – referential processing may vary across cultures calls into question any assumptions for its broad role in cognition as proposed by Sui & Humphreys (2015).

We were interested in a different question altogether, that required us to zoom in and try to analyze the mechanisms and processes underlying the self – referential advantage so robustly demonstrated by Sui et al. (2012, 2013, 2015, etc.). In the current set of experiments, we wondered whether participants in Sui et al. (2012)’s experiments merely formed associations with the specific shapes that they were exposed to (for e.g. triangle, circle, square) or whether they analyzed the target shapes for their properties for e.g. prominent visual features (say, number of vertices), and could use the analysis to develop some sort of categorizing rules that would allow them to distinguish these shapes from each other, eventually leading to the self – referential processing advantage. If the participants were learning to categorize stimuli on the basis

of these rules, it would follow that the self – referential advantage could in – principle be available for not just the specific triangle or circle, but to all triangles or circle.

Intuitively, the same mechanism of categorization may underlie the findings for prejudice not only against a specific member of a particular category, but rather against all of its members (say all males, or all females) (Akrami, Ekehammar & Berg, 2011). Moving on, we are aware that individuals in the real world may be using not one but several criteria for categorization in social settings (Liberman, 2017) and use the same to organize their social interactions, for e.g. stereotyping, prejudice etc. Similarly, it could be asked whether participants in a paradigm like the one used by Sui et al. (2012) may be sensitive to not just a single rule (say number of vertices) but a combination of rules (say, filled color and number of vertices) and may use the two in conjunction to actively distinguish self – relevant shapes from other – relevant shapes. Both of the above questions have not received much attention in recent literature which uses experimental paradigms to demonstrate the benefits of self – referential processing across various cognitive domains. Further insights from these questions may help in experimental demonstration of how individuals may pick up seemingly simple rules to categorize stimuli in social setting. If one were to go by the assumptions of Sui & Humphreys (2015), the categorization on the basis of relatively simple rules may underlie the manifestations of socially relevant phenomena like prejudice, stereotyping etc. To sum up we were interested in investigating (a) whether participants simply make associations with specific exemplars, or whether they learn to categorize on the basis of rules (such as on the basis of visual features), also (b) whether the self – referential advantage can then be extended beyond the specific exemplar to the entire category of the target stimulus (say all triangles) and finally, (c) whether participants can learn to differentiate stimuli on the basis of not a single but a combination of rules or features (say a filled color and number of vertices).

To investigate these questions, across five different experiments, we first replicate Sui et al. (2012)’s findings for self – referential processing with our stimuli (triangle, quadrilateral, and octagon). Then we exposed the participants to 20 exemplars of each kind of shape and demonstrated the self – referential advantage across the different shapes for each category. In the next set of experiments, we familiarized a different set of participants with specific shape stimuli in the association phase but exposed them to larger set of previously unseen exemplars from the same category in the test phase. We also check whether participants can learn to categorize not only on the basis of one, but two rules of categorization viz. color and number of vertices. The experiments equipped us to test the questions outlined above and make conclusions about whether participants were forming broad associations or were in-fact learning rules for categorization of stimuli. Findings indicate the latter to be true, which are discussed later in more detail.

Experiment 1a

Method

Participants: Participants were 18 college students of Indian Institute of Technology, Kanpur (six women; 22 to 32 years of age, $M=24.78\pm2.48$) who were awarded 100 INR as compensation. All except one participant were right-handed. All had normal or corrected-to-normal vision. Informed consent was obtained from all prior to experiment according to procedures approved by a local ethics committee.

Stimuli and Task: One of three geometric shapes (triangle, quadrilateral, and octagon, each $3.5^\circ \times 3.5^\circ$), presented above a fixation cross ($0.8^\circ \times 0.8^\circ$) displayed at the center of screen. Each shape had labels of ‘Self’, ‘Friend’, or ‘Stranger’ with them counterbalanced across subjects. One of those words was presented below the shape, below the place the fixation cross would appear otherwise. Distance between the center of the shape and the word was 3.5° . All stimuli were no-fill and black bordered and shown on a white background of a 15.6-inch monitor of Acer Predator Helios 300 ($1,920 \times 1080$ px at 60 Hz) where experiment was run. In testing block, participants recognized and judged consistency of presented pairings with the original associations. The experiment was coded and run on python 3.7 using PsychoPy library for GUI.

Procedure: In the learning stage, participants were presented with and asked to encode three geometric shapes (triangle, square, and octagon) with one of *Self*, *Friend*, *Stranger* labels randomly attached under them. In the matching-task stage, participants responded to fast presentations of various shape-label pairs as to whether the pairing actually matched the original associations or not. A fixation cross for 500ms would follow a blank screen at the start of each trial. A pair of a shape and label (*self*, *friend*, or *stranger*) is then presented for 100 ms. This pairing could either congruent to the initial association, or it could be a combination of an incongruent label with an incongruent shape. Shape-label pairings appeared at random. Immediately after, a blank frame is made visible for 1,200ms within which participants respond by pressing one of two response keys accurately and quickly (for immediate response). Feedback (‘correct’/‘incorrect’/‘no response’) shown for 500 ms after each presentation in trial block, but no feedback is shown in test block. Participants were informed of their accuracy after trial block, and again, not informed of any performance measure after test block.

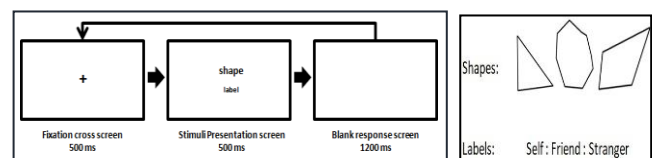


Figure 1.1: Expt-1 flow and the used Shape-Label stimuli

Results and Discussions: Experiment had shape category and matching judgement as two within subject variables.

Correct responses of each category which were two standard deviations away from the initial means were cleaned up.

Table 1: Mean RT and Mean Accuracy as function of Match Condition (Matched vs Unmatched) and Shape Category (Self, Friend, Stranger) in Experiment-1a

Condition	Category	MeanRT(ms)	Accuracy(%)
Matched	Self	621.91(105.08)	93.33 (6.5)
	Friend	675.15(118.07)	80.55 (16.5)
	Stranger	698.01(98.66)	76.94 (16.03)
Mismatch	Self	751.29 (98.31)	80.83 (16.26)
	Friend	733.59 (87.04)	82.91 (10.71)
	Stranger	749.823 (92.34)	84.30 (12.31)

RT=reaction time; Acc=proportions correct; SD in brackets

Through ANOVA for RT with a within-subjects factor shape category and match condition, significant effect of category shape, $F(2,34) = 4.616$, $p=0.017$, and also significant effect of category match $F(2,34)= 80.96$, $p=0.000$ was found. The interaction of the shape category and matching judgement was also significant $F(2,34)=11.68$, $p= 0.000$. For ANOVAs of accuracy, shape category was significant $F(2,34)=5.97$ and $p=0.006$. Even though Match Category wasn't significant ($p=.607$), interaction of shape category and matching judgement was significant, $F(2,34)=9.477$, $p=0.001$.

Bootstrapping method was applied to assess the overall effect to examine, by combining RTs and accuracy, the central tendencies and distribution of matching judgements in each condition. (Davison & Hinkley, 1997) The bootstrapped sample mean observations for self-matched responses fall in the lower right corner of the figure with higher accuracy and lower RT, the stranger-matched responses fall in the upper left locations with higher RT and lower accuracy, while the friend-matched responses are in the middle locations. They are separately discernible for the match cases, while the spread for responses to mismatching shape-label pairs seem overlapping.

The bootstrapped mean plots over 2000 iterations are as shown in Fig 1.2 for match and mismatch cases respectively.

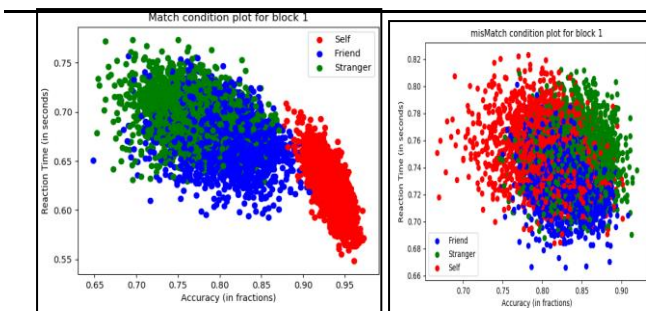


Figure 1.2: bootstrapped distribution for match (left) and mismatch (right) conditions of self, friend, other, Expt 1a

Experiment 1b

Methods

Participant, Stimuli and Task: These were the same as Expt-1a; but here we had a larger stimuli-sample with there being a group of similar shapes for each of the 3 different classes of triangle, quadrilateral, octagon exemplars where each class consisted of 20 exemplars generated randomly conforming to the same standard using MATLAB2016.

Procedure: In the first stage of this experiment, participants were made to undergo a familiarization task wherein they had to just observe and make associations of all shape-label pairings that were presented to them individually for 1.5 seconds each. The angle, size of stimuli remained similar to Expt 1a. All the true shape-label pairing samples are shown twice randomly one after another over a total of 120 runs and participant is asked to just attend to all the shapes as carefully as he can. After this, following an initial trial block, participants performed a match-identification task like in Expt-1, but this time over all possible combinations of shape-label pairing for each of 20 triangles, 20 quadrilaterals and 20 octagons associated to either of the three classes. The shape-label combinations, are presented for 100 ms followed by a blank response screen of 1200 ms. A 500 ms feedback screen is shown in the trial block. No feedback given in test block. Every response was followed by a 50ms flash of word 'next' marking end of that response window and the start of next.

Results and Discussions Expt-1b also had the same shape category and matching judgement as the two within subject variables. Again, RTs and accuracies of the correct responses of each category which were two standard deviations away from the initial means were cleaned up. Table 1b reflects the accuracy and reaction time data for Expt-1b.

Table 2: Mean RT and Accuracy of Expt 1b

Condition	Category	MeanRT(ms)	Accuracy
Matched	Self	690.46 (87.96)	89.74 (8.22)
	Friend	716.35 (75.91)	85.78 (10.13)
	Stranger	747.43 (94.10)	79.29 (11.94)
Mismatch	Self	792.13(85.89)	82.07 (13.89)
	Friend	789.63 (80.96)	84.46 (10.57)
	Stranger	793.06 (84.01)	83.99 (12.23)

RT=reaction time; Acc=proportions correct; SD in brackets

For ANOVA of RT, with within-subjects factor shape category (*self, friend, stranger*) and match conditions (match vs non match), significant effect found for category shape $F(2,34)=5.40$, $p=0.009$ and also significant effect of category match $F(2,34)= 184.15$, $p=0.000$. The interaction of the shape category and matching judgement was also significant $F(2,34)=8.34$, $p=0.001$. For ANOVAs of accuracy, shape category had significant effect $F(2,34) = 4.93$ and $p= 0.013$. Though the Match Category didn't show significance

($p=.625$), interaction of shape category matching judgement was significant, $F(2,34)=6.912$, $p=0.003$.

Bootstrap sample means analysis is reconducted for this new dataset and the distinct distributions and central tendencies for the different social salience groups of self, friend, stranger was observed for the match condition while there was again overwhelming overlap of distributions of the three different classes in the accuracy – RT plots.

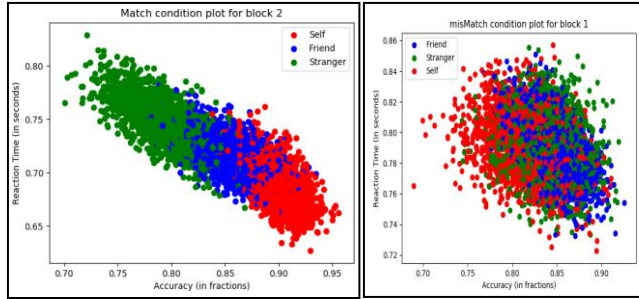


Figure 1.3: bootstrapped distribution for match (left) & mismatch (right) conditions of self, friend, other, Expt 1b

Experiment 2a, 2b, 2c

Here, we explore if learning of association and consequent variations in preferential processing, whether for the individual exemplar or generalized for the collective, is a consequence of associating salience with not only exemplars which are presented for learning but also attaching salience with one or more features/rules common to all the presented exemplars, like the shape (number of edges) in **Expt-2a** and colour in **Expt-2b** and coloured shape in **Expt-2c**. So here association is done on a limited subset of the stimuli-family, while testing is done on a wider set (whole family) consisting of seen and unseen exemplars. Order of presentation of **Expt 2a & 2b** is counterbalanced across subjects to negate any latent influence of practice effect on the data of single rules.

Method

Participants: Expt 2a, 2b, 2c had 20 participants, college students of IIT Kanpur, all but one right-handed, all corrected to normal vision (2 females; 19-34 years of age; $M=25.25 \pm 2.78$). Informed consent was obtained as approved by ethics committee. Rs.100 is rewarded as compensation to each.

Stimuli and Task: For **Expt-2a** colourless Triangles, Quadrilaterals and Pentagons are chosen as geometric shape stimuli. There were 10 variations of each of the 3 kinds, all differing intra-class conforming to the same standards. For **Expt-2b**, we used the triadic colours of cyan, magenta and yellow, again with 10 variations of each, controlled over a narrow range of luminance consistent across 3 colours. For **Expt-2c**, all colour-shape combinations were used with congruent-incongruent combinations as per social salience. All stimuli are generated using MATLAB and shown on a white background on a 24'' 120 Hz monitor. Like in Expt-1, each stimuli ($3.5^\circ \times 3.5^\circ$) had one of the labels of *Self*, *Friend*,

Stranger ($2.5^\circ/3^\circ \times 1.4^\circ$) attached with them at distance of 3.5° from centre of image and were presented after a fixation cross ($0.8^\circ \times 0.8^\circ$) and then responded-to accordingly

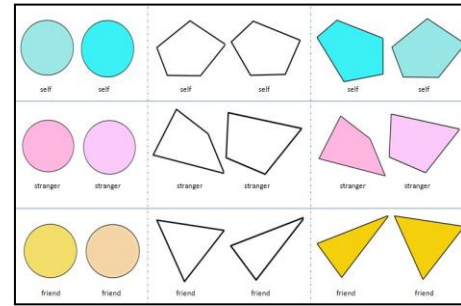


Figure 2.1: 2 exemplars of stimuli used in Expt 2b, 2a, 2c.

Procedure: In all 3 experiments, subjects were shown limited subset (half for Expt 2a, 2b; less than 25% for Expt 2c) and asked to learn the associations of all shape-label pairings that were shown on screen. After short trial blocks, participants did match-identification tasks like in Expt-1 on all combinations of shape-label pairing of 10 exemplars from each family of each of the geometric shapes (as in Expt-2a), each of colours (Expt-2b) and coloured shape (Expt-2c). Pairing was presented for 100 ms, followed by a blank response screen displayed randomly for 800-1200 ms within which one judges if the presented combination is a correct match or a non-match. There were 40 trials for each match-mismatch condition in Expt 2a & 2b; Expt 2c was additionally balanced with distracting coloured shapes that combined colours and shapes of differing salience, thus making that stimulus not belong originally to any class itself.

Results and Discussions: All 3 had type category (shapes in Expt 2a, colours in Expt 2b, coloured shapes in Expt 2c with labels of Self, Friend, Stranger) and matching judgement as the within-subject variables. Outlying measures were cleaned **Expt 2a** tested rule learning and preference of socially salient association with colourless shapes, and results are reflected.

Table 3: Mean RT and Accuracy, Expt-2a, as functions of Type Category and Match Conditions

Condition	Category	MeanRT(ms)	Accuracy(%)
Matched	Self	512.27 (111.88)	85.37 (12.28)
	Friend	574.73 (127.11)	68.87 (24.52)
	Stranger	585.95 (130.20)	62.75 (25.42)
Mismatch	Self	629.02 (136.70)	68.12 (17.35)
	Friend	633.06 (150.23)	72.62 (15.14)
	Stranger	608.72 (139.6)	73.00 (17.53)

RT=reaction time; Acc=proportions correct; SD in brackets

With a within-subjects factor type category (*self*, *friend*, *stranger*) and matching condition (matched vs Non-matched), ANOVA of RTs gave significant effect of category type, $F(2,38)=5.40$, $p=0.009$ and also significance in category

match $F(2,38)=184.15$, $p=0.000$. The interaction of the type category and the matching judgement was also significant $F(2,38)=8.34$, $p=0.001$. For ANOVAs of accuracy, type category had significant effect $F(2,38)=6.836$ and $p=0.003$. Even though the Match Category didn't show significant effect ($p=.765$), interaction of type category and matching judgement was significant, $F(2,38)=11.354$, $p=0.000$.

Bootstrap sample means plot showed discernable spread in match-case while overlapped distributions for mismatch case

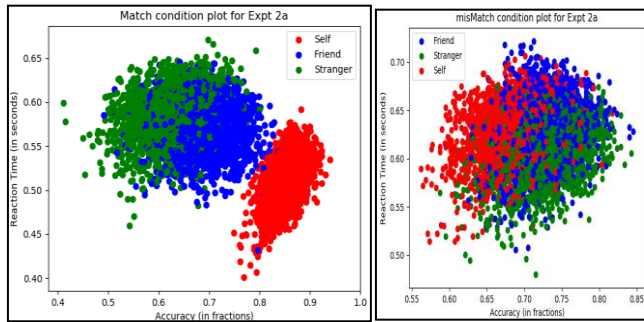


Figure 2.2: bootstrapped distribution for match (left) & mismatch (right) conditions of self, friend, other, Expt 2a

Expt 2b recorded responses to differing colours associated with the different labels and response measures are reflected.

Table 4: Mean RT and Accuracy for Experiment 2-b as functions of Type Category and Match Conditions.

Condition	Category	MeanRT(ms)	Accuracy(%)
Matched	Self	478.74 (115.36)	84.00 (19.72)
	Friend	527.61 (140.35)	74.00 (22.39)
	Stranger	550.77 (162.50)	66.25 (22.68)
Mismatch	Self	589.41 (154.89)	71.87 (15.15)
	Friend	586.47 (153.89)	70.25 (18.19)
	Stranger	586.97 (154.43)	74.25 (19.89)

RT=reaction time; Acc=proportions correct; SD in brackets

ANOVA with within-subjects factor *type-category* (*self*, *friend*, *stranger*) and match condition, showed significant effect of type category for RT's, $F(2,38)=5.596$, $p=0.007$ and also significance of match condition $F(2,38)=44.836$, $p=0.000$. The interaction of the type category and the matching judgement also had significant effect with $F(2,38)=6.021$, $p=0.001$. For ANOVAs of accuracy, type category had significant effect $F(2,38)=6.44$ and $p=0.004$. Even though Match Conditions didn't generate significance ($p=.517$), the interaction of the type category and matching judgement was again significant, $F(2,38)=9.52$, $p=0.001$. Bootstrapped sample means were again plotted over 2000 iterations and preferential distribution for Self is again seen at the bottom right corner for match case, while the spread for non-match cases remained largely overlapped.

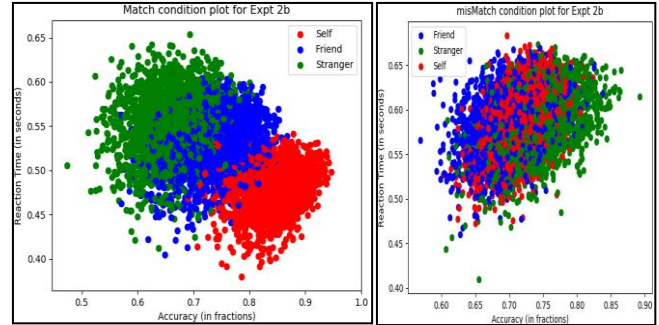


Figure 2.3: bootstrapped distribution for match (left) & mismatch (right) conditions of self, friend, other, Expt 1b

Expt 2c recorded responses to pre-associated coloured shape combinations and some distractions which are presented with the different labels and response measures are reflected.

Table 5: Mean RT and Accuracy of Experiment 2-c

Condition	Category	MeanRT (ms)	Accuracy (%)
Matched	Self	489.53 (110.17)	91.37 (7.79)
	Friend	530.81 (131.86)	80.12 (18.66)
	Stranger	568.78 (136.44)	74.00 (15.07)
Mismatch	Self	586.05 (132.48)	74.99 (18.38)
	Friend	584.93 (137.96)	76.22 (17.93)
	Stranger	591.59 (133.22)	75.31 (18.89)

RT=reaction time; Acc=proportions correct; SD in brackets

Using ANOVA, with within-subjects factor *type category* (*self*, *friend*, *stranger*), significant effect of category type for RTs, $F(2,38)=8.963$, $p=0.001$ was found as also significant effect of category match at $F(2,38)=25.137$, $p=0.000$. The interaction of the type category and matching judgement was also significant $F(2,38)=9.163$, $p=0.001$. For ANOVAs of accuracy, type category was significant with $F(2,38)=10.398$ and $p=0.000$. Matching judgement just had significance $F(2,38)=4.657$ and $p=.044$. Their interaction had significance- $F(2,38)=13.294$, $p=0.000$.

Again, discernible distributions and preferential spread for 'Self' relative to 'Strangers' in the matched data was found in the bootstrapped plot. Non-match case had overlaps, again.

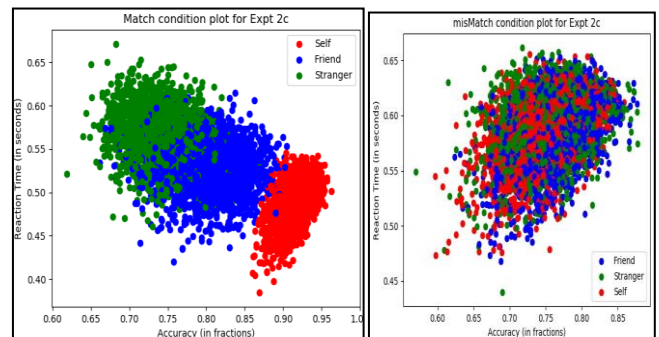


Figure 2.4: bootstrapped sample means distribution for match conditions (left) and mismatch conditions (right)

General Discussion

We replicated the findings of benefits in self – referential processing for shapes (Experiment 1A). We could demonstrate that the self – referential advantage in processing was not merely limited to specific exemplars of the associated stimuli but that it could be extended to various stimuli from the same category. Further, in the next set of experiments, we demonstrated that participants do indeed learn to categorize the stimuli into self – relevant or other – relevant, on the basis of learning rules based on visual features (say number of vertices (Experiment 2 A) and color (Experiment 2B)), and even a conjunction of both (number of vertices and color in conjunction (Experiment 2C)).

These findings suggest that participants may indeed be analyzing the target shapes in terms of the properties like visual features, and not merely forming associations with the specific shapes. Moreover, participants learn these properties and use them as rules for categorizing even novel (previously unseen) stimuli, using the same rules.

These findings may interpreted as evidence for the role of self – referential processing in social interactions. Similar findings have earlier been reported by Johnson and Gadon (2002), where they demonstrate that participants have better memory for adjectives encoded in terms of group – reference, rather than just on the basis of semantic features.

Also, following from the earlier discussion, the current findings may be taken as a demonstration of how people may use simple features or characteristics (visual features in these experiments) to organize and categorize their social world. Researchers have proposed that the ability to form abstract social categories may help people navigate the complex social world by structuring their thoughts, beliefs and actions according to group – memberships (Lieberman, Woodward & Kinzler, 2017). We propose that our findings demonstrate that individuals may form these abstract social categories, using self as a point of reference and may use these further to define the self vs. the other. Indeed, it has been proposed that negative consequences of this process such as prejudice may be explained as an outcome of self – definition in the context of the other and also in terms of overidentification of self-group as the in-group (Knight, 2015). More research is needed, however, to arrive at strong conclusions at this point. In our future experiments, we are testing whether this rule learning as demonstrated in experiments 2A – 2C, can lead to disadvantage (or prejudice) in processing stimuli identified as friend/other relevant in specific task contexts.

References

Akrami, Nazar & Ekehammar, Bo & Bergh, Robin. (2011). Generalized Prejudice: Common and Specific Components. *Psychological science*.

Cunningham, S. J., & Turk, D. J. (2017). Editorial: A Review of Self-Processing Biases in Cognition. *Quarterly Journal of Experimental Psychology*

Davidson, A.C., & Hinkey, D.V. (1997) *Bootstrap methods and their application*. Cambridge, UK: Cambridge University Press

Johnson, C. Gadon, O. (2002). Self Reference and Group Membership: evidence for Group Reference effect. *European Journal of Social Psychology* (pp. 261-274). John Wiley & Sons Ltd.

Knight, Zeldia G. (2015) The self in prejudice, *Journal of Psychology in Africa*

Lieberman, Z., Woodward, A. L., & Kinzler, K. D. (2017). The Origins of Social Categorization. *Trends in cognitive sciences*, 21(7)

Sui, J., & Humphreys, G.W. (2011). The boundaries of self-face perception: Evidence from response time distributions.

Sui, J. He, X. Humphreys, G.W (2012). Perceptual Effects on Social Salience: From Self Prioritization Effects. *Journal of Experimental Psychology Human Perception and Performance*.

Sui, Jie & Rotshtein, Pia & Humphreys, Glyn. (2013). Coupling social attention to the self forms a network for personal significance. *Proceedings of the National Academy of Sciences of the United States of America*.

Sui, J. He, X. Humphreys, G.W (2015). The Integrative Self: How Self-Reference Integrates Perception and Memory. *Trends in Cognitive Science*

Sparks, S., Cunningham, S. J., & Kritikos, A. (2016). Culture modulates implicit ownership-induced self-bias in memory. *Cognition*