UC Merced

Proceedings of the Annual Meeting of the Cognitive Science Society

Title

The lure of the self: How we misattribute our lesser likes to the “other” in perspective-taking and decision-making

Permalink

https://escholarship.org/uc/item/48p9t4gg

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 43(43)

ISSN

1069-7977

Authors

Ongchoco, Joan Danielle K.
Inchingolo, Marco
Paul, L.A.

Publication Date

2021

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

The lure of the self: How we misattribute our lesser likes to the "other" in perspective-taking and decision-making

Joan Danielle K. Ongchoco (joan.ongchoco@yale.edu)
Department of Psychology, Yale University

Marco Inchingolo (m.inchingolo@gmail.com)
Institut Jean-Nicod, ENS, EHSS, CNRS

L. A. Paul (la.paul@yale.edu)
Department of Philosophy, Yale University

Abstract

How do we represent other people? Our representations are prone to a wide range of biases. We project our mental states onto others (especially when we assume they are similar to us), or rely on existing stereotypes (when we think they are different). But sometimes, it can be unclear how similar or different a person actually is from us. How does this affect how we represent their preferences? Here, subjects declared their favorite and least favorite colors and were introduced to another person whose preferences were neither completely similar or dissimilar. Across experiments, people successfully remembered the other person's preferences, but they also tended to falsely ascribe their own lower ranked preferences to the other player in multiple memory and decision-making tasks. These results suggest a tendency to distance ourselves from preferences we identify with the least in our hierarchy of preferences, and to associate them instead with the "other".

Keywords: self, memory, preferences, perspective-taking

Introduction

We make decisions for ourselves and for others. Decisions for others can range from simple (e.g. what to give our partner for their birthday), or weightier ones (e.g. whether to continue treatment for a loved one who can no longer decide for themselves). Making these choices requires "perspective-taking": we need to remember and keep track of another person's preferences, while performing a calculation that is consistent with these preferences. The nature of this calculation can depend on how similar or different another person is from us. But people are complicated: often times, a person's preferences can be similar to ours in some respects, and different in others. How, then, do we adjust to this complexity? Sometimes, it can be good to represent the preferences of others as similar to our own, as when we need to reach a common ground. Other times, it can be important to represent another person's preferences as different, as when we need to make decisions for them that we would not otherwise make for ourselves. Given the complexity of people's preferences in relation to our own, do we default to the ways they are similar to us? Or the ways they are different?

Representing others as similar

The egocentric bias is the tendency to overestimate the extent that other people (and by extension their preferences) are similar to us. In other words, it is the tendency to infer other people's thoughts, beliefs, and affective states from our own point of view. People often fail to discriminate between their own and other people's beliefs (e.g. Keysar, Lin, & Barr, 2003), are prone to project their own affective states onto others (e.g. Van Boven & Loewenstein, 2003), and tend to overestimate the salience of their actions in the eyes of others simply because it is salient from their own egocentric point of view (e.g. Gilovich, Kruger, & Medvec, 2002). People's preferences can also influence the very way that they learn about another's preferences (e.g. Tarantola, Kumaran, Dayan, & De Martino, 2017).

This egocentric bias can be helpful, especially when we are interacting with people who are similar to us. But there are many other instances when this bias can lead to errors, primarily because others can think, feel, and judge differently from the way we do. A great deal of work has explored how we can become more adept at correcting this bias, perhaps by adjusting our inferences to move closer to what the other person may actually be feeling (e.g. Epley, Keysar, Van Boven, & Gilovich, 2004; Tamir & Mitchell, 2013). This process of 'correction' can be moderated by a person's own disposition or motivations to connect or distance themselves from another person. For instance, empathic individuals tend to mimic other people's postures and mannerisms (e.g. Chartrand & Bargh, 1999), and conversely, psychopaths fail to automatically take the perspective of others (e.g. Drayton, Santos, & Baskin-Sommers, 2018).

Representing others as different

Another mechanism by which we represent other people and their mental states is through stereotypes. Stereotypes are beliefs about categories of people that give us information about what they might be likely to think or prefer. For example, if a person is a member of a conservative political party, this might inform the way we represent their preferences, where we consider what other people in this party might typically prefer. How much these stereotypes matter to our representations seems to be a function of similarity between self and other. For instance, students project their own attitudes onto other students of the same university, but rely instead on stereotypes to infer attitudes of students of a different university (e.g. Ames, 2004).

When making decisions for other people, some work also suggests that we observe different neural patterns when choosing for ourselves than for another person (e.g. Nicolle

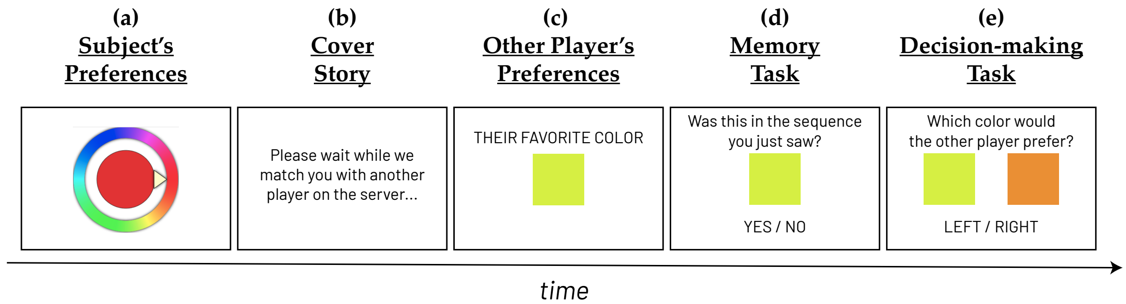

Figure 1: Experiment flow. (a) Subjects chose their favorite and least favorite colors from an interactive color wheel. (b) Subjects were matched with an 'other player' on the server. (c) Subjects were shown the favorite and least favorite colors of the other player. (d) Subjects performed a recognition memory task. (e) Subjects were asked to choose which of two colors the other player would prefer.

et al., 2012). Consider an inter-temporal discounting task. Subjects would be asked to choose between a large value delivered later (e.g. $30 in 3 months) and a smaller value delivered sooner (e.g. $10 today). Critically, sometimes, subjects would be asked to choose for another person. A particular activity pattern across brain regions was observed in trials where subjects chose for themselves—which was qualitatively different when subjects instead chose for their partner. This suggests a 'switching' of representations that occurs when we make decisions for others.

The current study

So far, we've discussed accounts of how we represent others when they are similar to or different from us in a particular dimension. But sometimes, setting biases and stereotypes aside, people can be both similar and dissimilar to us across multiple dimensions, and across a hierarchy of preferences. For instance, your favorite cuisine might be Italian, and another person's favorite might be Chinese, but you both also like Mexican. Or your favorite color might be red, and another person's favorite color might be yellow—but for each of you, your second favorite colors may be some shade of blue. One's preferences can be more or less similar to another person's preferences in complex ways.

This is where the interaction between representation and decision-making can matter. Even when we can remember another person's preferences well, we might rely on heuristics when deciding for them. If so, do we presume similarity, or do we presume difference? Thus, we tested not only people's ability to make decisions for others, but also how they remembered another's preferences in the first place. Subjects first declared their own preferences (Figure 1a)—their favorite and least favorite colors—and were then presented with another player's preferences (Figure 1b-c). This player was in fact just an 'avatar', introduced with minimal information to avoid stereotype-reasoning from being deployed.

The color preferences of this other player were manipulated to be neither completely similar nor dissimilar from the subject's own preferences. These manipulations were designed to avoid swaying people one way (e.g. triggering the egocentric bias because the other player is obviously similar) or another (e.g. triggering the stereotype bias because the other player is obviously dissimilar). To blur obvious similarities or dissimilarities, we leveraged the structure of color space and color preferences. First, because of the nature of color space, some colors will be obviously different (as when two colors are on opposite sides of the space, e.g. red and blue), but for most of the time, colors are sampled on the continuum (such that it is unclear what the boundary between light-orange or yellow, or green-blue or blue-green actually is). Second, it was important that we never presented subjects with only one color as a presentation of the other player's preferences. Rather, we let them build a hierarchy of the other player's preferences across multiple colors. Thus, even if some colors in the other player's palette were obviously different from the subject's, other colors could be similar. In Figure 3, for example, the player's favorite color could be red, while the other player's favorite color could be yellow – but their next favorite colors are may be a shade of blue-indigo. This allowed us to create complexity in the way the other player's hierarchy of preferences related to the subject's.

To tap into how they were representing other people's preferences, subjects were first tested for their memory of the other player's preferences (Figure 1d). To look at how they make decisions for the other people, we then tested their ability to correctly identify the other player's preferences in a two-forced choice task (Figure 1e; Experiment 1), or to correctly choose for them (Experiments 2 and 3). To ensure that results were due to subjects' own preferences, and not just prior stimuli interfering with reasoning about another person's preferences (i.e. maybe the most recent colors we see or interact with interfere with memory more), results were compared to a Memory-Control condition where subjects were shown a random set of colors prior to the introduction of the other player's preferences.

Experiment 1: Representing the 'Other'

How do we represent other people's preferences, when these are not entirely different from our own? Subjects encountered another person's preferences and were subsequently tested in recognition and two-forced choice tasks. Across tasks, we wanted to see whether they could accurately identify the other person's preferences.

Method

All methods and analyses were pre-registered at: https://aspredicted.org/blind.php?x=wb73id.

Participants. 200 non-excluded subjects were recruited online through Prolific (www.prolific.co). This sample size was chosen based on pilot experiments, was pre-registered, and is the same across the experiments reported here. We excluded subjects based on the following criteria: We asked subjects in a post-experimental debriefing phase how well they paid attention (on a continuous scale, with 1 being very distracted and 100 being very focused), and we excluded subjects who reported an attention level of 80 or below. We also excluded subjects whose total experiment time was more than 2 standard deviations from the population mean. These subjects were replaced until we obtained our target sample size.

Stimuli. The entire experiment was programmed using the jsPsych library (De Leeuw, 2015) and custom javascript plug-ins. Due to the nature of online experiments, we cannot specify here the exact size, color, or brightness (etc.) of the stimuli because we cannot know each subject's viewing conditions. However, any distortions or differences would have been equated across all stimuli and conditions.

The main stimuli in the experiment involved an interactive color wheel, and color tiles that subjects had to respond to. The color wheel was adapted for our purposes from an existing code library, https://github.com/techslides/huewheel, and its colors were defined in HSL values (see Figure 2). Saturation of the colors was always fixed at 80%, and brightness, at 50%. Square colored tiles (such as those in Figure 1c-e) were defined in similar HSL values.

Procedure and Design. In the first section, subjects first saw a color wheel at the center of their screen, with a row of four rectangles that formed an empty color 'palette' that was always visible at the bottom of the screen. Subjects could either be assigned to one of two conditions. In the 'Self' condition, subjects could interact with the color wheel using the yellow-white triangle, and click a button to add colors to their palette. They were told to choose their favorite, second best, third best, and least favorite colors, and the palette

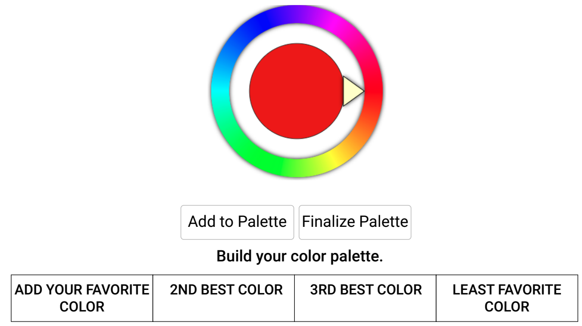

Figure 2: Sample color wheel display with color palette in the bottom. In the Self condition (Experiments 1 and 2), subjects filled in their favorite and least favorite colors, as depicted here. In the Memory-Control condition (Experiments 1 and 2), subjects were presented four randomly chosen colors from the wheel. In the Interactive-Control condition (Experiment 3), subjects chose four colors corresponding to blue, purple, orange, and red (in a different random order for each subject).

rectangles were labelled accordingly (see Figure 2). In the Memory-Control condition, subjects saw the exact same display, except the yellow-white triangle was removed, and subjects were told to wait as four colors were randomly selected and added to the palette. Palette rectangles were instead simply labelled with 1, 2, 3, and 4. The rates at which these colors were sampled and added matched the average time it took for subjects to choose colors in the Self condition. Once four colors were selected, subjects proceeded to the next section.

In the next section (the Memory task), subjects were told that other players chose their own favorite and least favorite colors, and that they would now be matched with one of these players. These 'players' were ultimately computer-generated avatars. To construct this other player's color palette, we took the subject's reported color preferences, shuffled them around, and varied the hues by differing degrees (see Figure 3). For the other player's favorite color, we took the subject's least favorite color—i.e. their lowest ranked color—and added/subtracted 30 °(i.e. a red hue at H=0°would then be orange at H=30°). For the other player's lowest ranked color, we took the subject's favorite color and added/subtracted 30°. For the other player's second best and third best colors, we used the subject's second best and third best colors, and added/subtracted 60°(i.e. a red hue at H=0 would be yellow at H=60°). Thus, the colors from the other player's color palette were not completely similar or dissimilar from the subject's preferences. We further ensured that no colors were ever the same in raw hue value, randomly jittering values within the range of 5-10°. Subjects were then shown the other player's colors (that appeared in square tiles for 1000ms) one by one (see Figure 1d), and were simply asked to remember each color as best as they could. Each color was accompanied with the corresponding label of 'favorite color', '2nd best color', '3rd best color', and 'least

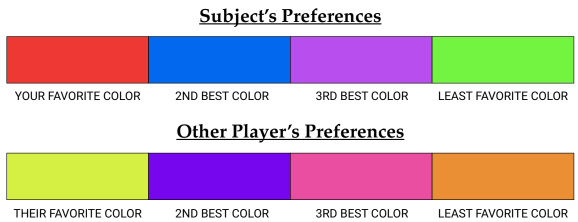

Figure 3: Sample hierarchies of color preferences of the subject and the 'other player'. Here, the player's favorite color is of a red hue, while the other player's favorite color is of a yellow hue—but their second favorite colors are both of a blue-indigo hue, and their third favorite colors are both of a purple-pink hue.

favorite color'. After subjects saw all four colors, they were tested for their memory of the other person's colors. They were shown a square tile and determined if this was in the other player's palette. Half of the time, the tile's color would be from the other player's palette. The other half of the time, the color would be from the first section (from the set of their own selected colors, or of the randomly selected ones).

In the final section (the Decision-making task), subjects saw two colored square tiles, and decided which the other player would prefer (see Figure 1e). Subjects were presented with three choice pairs in the same sequence: (1) the other player's favorite (highest ranked) color vs. the other player's least favorite (lowest ranked) color; (2) the other player's second ranked color vs. the subject's favorite color; and (3) the other player's favorite color vs. the subject's lowest ranked color. In all these three choice pairs, the 'correct' choice was always either the other player's favorite or second ranked color. These three choice pairs were designed to test whether subjects could accurately identify the other person's preferences (pair #1), and whether subjects would confuse their highest ranked and lowest ranked colors with the other player's highest ranked and lowest ranked colors (pairs #2 and #3).

Results and Discussion

First, we looked at whether subjects successfully encoded and identified the other player's color preferences. Memory performance for the other player's color preferences (measured in terms of a *hit rate*, or whether the subject said 'yes' when the color was indeed from the other player's palette) was above chance for both the Self (M=75.12%; $t(99)=12.57$, $p<.001$, $d=1.26$) and Memory-Control conditions (M=81.12%; $t(99)=21.99$, $p<.001$, $d=2.20$). Thus, regardless of the condition, when shown the preferences of another player, subjects accurately kept track of which colors belonged to the other player.

But would the subject's own preferences interfere with how they represented the other player's color preferences? To assess this, we looked at whether subjects false alarmed when they were presented with their own colors. Results

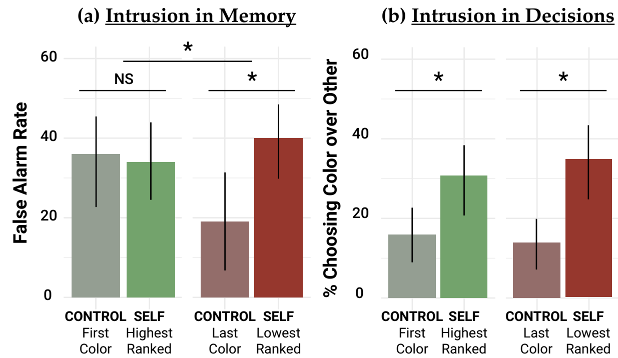

Figure 4: Main results from Experiment 1. (a) False alarm rates between Self and Memory-Control conditions for favorite (first) and lowest ranked (last) colors. False alarms were greater for the subjects' lowest-ranked colors compared to the the last colors they saw in the Memory-Control condition. (b) Proportions of choosing one's personal preferences over the other player's preferences. Subjects were more likely to choose their favorite and least favorite colors compared to critical Memory-Control conditions. Error bars reflect 95% corrected and accelerated bootstrapped confidence intervals.

revealed significantly higher false alarm rates in the Self condition than the Memory-Control condition ($t(198)=2.02$, $p=.045$, $d=0.29$), demonstrating a trace of subjects' preferences on their representations of the 'other'. Further probing revealed that this 'self-intrusion' was driven by false alarms for the last color that subjects chose compared to the last color they were shown (Self vs. Memory-Control, Fisher's test, $p=.002$), but not the first color they chose compared to the first color they were shown (Self vs. Memory-Control, Fisher's test, $p=.882$)—with a reliable interaction across conditions ($F(1, 198)=6.17$, $p=.013$, $n_p{}^2=.015$); see Figure 4a. In other words, when shown their lowest ranked color, subjects tended to false alarm and report that they saw this color in the other player's palette—and this was not just because this color was the most recent color that subjects interacted with or saw in the first section. That we only find an effect for the lowest-ranked color and not the highest-ranked color suggests that this is not just a mere "preference" bias. Instead, the misattribution errors here were sensitive to the ranks of the subjects' preferences.

Does this self-intrusion carry over to more explicit evaluations of the other player's preferences? We then looked at the likelihood of subjects' choosing the right colors for the other player in a forced-choice task. When choosing between the other player's favorite vs. their lowest ranked color, subjects correctly chose the other player's favorite colors above chance (binomial test, 82 of 100, $p<.001$)—and there was no difference between the Self and Memory-Control conditions (Fisher's test, $p=.484$). This suggests that subjects could accurately assess the other player's preferences when choosing between two of the other player's colors.

The next two choice pairs involved choosing between the subject's chosen colors versus the other player's colors. Note, though, that the task was still the same: select the color that the other player would prefer. So, subjects simply had to ignore their chosen colors, and instead choose the other player's colors. However, results revealed that this is not what they did. When choosing between the other player's second ranked color vs. the subjects' favorite color, more subjects chose their favorite color instead of the other player's second ranked color in the Self condition than in the Memory-Control condition, when subjects were shown the first color they saw in the earlier section (Fisher's test, p=.019). When choosing between the other player's favorite vs. the subjects' lowest ranked color, more subjects chose their lowest ranked colors instead of the other player's favorite color in the Self condition than in the Memory-Control condition (Fisher's test, $p<.001$); Figure 4b. Subject's choices in these last two choice pairs were striking because these decisions were ultimately about the other player and their preferences. These results suggest that traces of our preferences can intrude in our evaluations of another person's preferences, perhaps confusing us, even when the decision is supposed to be for the other and not for ourselves.

Experiment 2: A More Ecological Replication

Does this self-intrusion carry over to actual decisions that we have to make for another person? Instead of merely asking subjects to about the other player's preferences, we now asked them to make a selection for the other player.

Method

This experiment was identical to Experiment 1, except where noted. 200 unique subjects were again recruited through Prolific. To test whether color preferences translated to more ecological choices, we generated 36 mugs in 36 different hues (the first mug started with a hue of 0°, and we changed the mug's hue in 10°increments all the way to 350°). To determine which mug corresponded to a hue, we rounded hues up/down to the closest tens (e.g. we used a mug of a 160°hue when the color hue was 162°). In the Decision-making task, instead of choosing between two colors, subjects decided which of two mugs they would select for the other player (see Figure 5).

Figure 5: Sample decision-making display in Experiment 2.

Results and Discussion

Memory performance for the other player's colors was again above chance for both the Self (M=75.50%; t(99)=13.33, $p<.001$, d=1.33) and Memory-Control conditions (M=80.38%; t(99)=18.30, $p<.001$, d=1.83)—with significantly higher false alarm rates in the Self condition than the Memory-Control condition (t(198)=2.68, p=.008, d=0.38). Further probing revealed that this self-intrusion was again driven by false alarms for the subjects' lowest ranked colors (Self vs. Memory-Control, Fisher's test, $p<.001$), but not their favorite colors (Self vs. Memory-Control, Fisher's test, p=.771)—with a reliable interaction across conditions (F(1, 198)=10.70, p=.001, $n_p{}^2$=.026); Figure 6a.

When choosing between mugs in the other player's favorite vs. lowest ranked color, subjects again correctly chose the correctly colored mug (i.e. the mug in the other player's favorite color) above chance (binomial test, 63 of 100, p=.012)—and there was no difference between the Self and Memory-Control conditions (Fisher's test, p=.768). For the next two choice pairs, subjects again just had to ignore the mugs in their chosen colors and choose the mugs in the other player's colors. In the first of these two, when choosing between the other player's second ranked color vs. the subjects' favorite color, subjects' preferences did not seem to intrude and they did not choose differently across the Self and Memory-Control conditions (Fisher's test, p=1). (This is different from the previous experiment, but we suspect that this may partly be because of the different nature of the task.) Critically, however, when choosing between the other player's favorite color vs. the subjects' lowest ranked color, more subjects again chose mugs in their lowest ranked colors instead of mugs in the other player's favorite color in the Self condition than in the Memory-Control condition (Fisher's test, $p<.001$)—with a reliable interaction across conditions (F(1, 198)=7.21, p=.008, $n_p{}^2$=.015); see Figure 6b. These results serve as a more ecological replication of the results from the previous experiment. They further confirm an intrusion of the self in both our representations of the other's preferences, and decisions that we might then make for them. But they also point to a very specific, counter-intuitive type of intrusion—such that consistently, across memory and decision-making tasks, this intrusion (surprisingly) came in the form of our ascription of our lowest ranked preferences to the other person.

Experiment 3: A Mere Interaction Bias?

In the previous experiments, the Memory-Control condition ensured that whatever memory/decision effects involving the subjects' *least* preferred colors were not simply due to these colors also being the most recent colors that they chose. But in the Memory-Control condition, another critical difference was that the colors were randomly presented to the subjects.

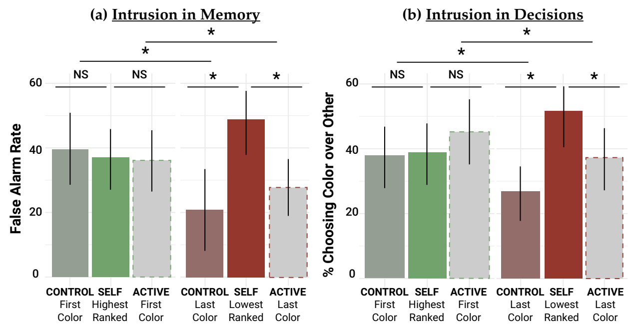

Figure 6: Main results from Experiments 2 and 3. (a) False alarm rates between Self, Memory-Control, and Interactive-Control conditions for favorite and lowest ranked colors. False alarms were greater for the subjects' lowest-ranked colors compared to the the last colors they saw (or interacted with) in the Memory-Control and Interactive-Control conditions. (b) Proportions of choosing one's personal preferences over the other player's preferences. Subjects were more likely to choose their least favorite colors compared to critical Memory-Control and Interactive-Control conditions. Error bars reflect 95% corrected and accelerated bootstrapped confidence intervals.

Perhaps the intrusion of the self in the previous experiments was not about preferences per se, but rather about active engagement with the color wheel? Here, to isolate the effect of preference versus engagement, we ran a separate Interactive-Control, where subjects interacted with the color wheel, but were not asked to choose their favorites.

Method

This experiment was identical to the Memory-Control condition in Experiment 2, except where noted. 100 unique subjects were recruited through Prolific. In the first section, instead of subjects simply being presented random colors, subjects could interact with the color wheel (as in the Self condition), but this time, were asked to choose a blue, purple, orange, and red color (in a different random order). This preserved the 'interactive' aspect of the Self condition, without explicitly asking subjects for their color preferences.

Results and Discussion

Memory performance for the other player's colors was again above chance (M=85.50%; $t(99)=14.26$, $p<.001$, $d=1.42$), and when choosing between mugs in the other player's favorite vs. lowest ranked color, subjects again correctly chose the correctly colored mug (i.e. the mug in the other player's favorite color) above chance (binomial test, 72 of 100, $p<.001$). The key result, however, was how false alarm rates and choice biases especially involving subjects' lowest-ranked preferences would compare to the Self and Memory-Control conditions from Experiment 2. False alarm rates in this Interactive-Control condition was not significantly different from the Memory-Control condition ($t(198)=0.34$, $p=.737$ $d=0.05$), but was significantly different from the Self condition ($t(198)=2.37$, $p=.019$, $d=0.34$)—with this difference again present only for subjects' lowest ranked colors (Self vs. Interactive-Control, Fisher's test, $p=.004$), but not their favorite colors (Self vs. Interactive-Control, Fisher's test, $p=.958$)—with a reliable interaction across conditions ($F(1, 198)=4.33$, $p=.038$, $n_p{}^2=.011$); Figure 6a. The proportion of choosing the other player's favorites as opposed to one's own was again no different between Self and Interactive-Control (Fisher's test, $p=.474$)—and critically, when choosing between the other player's favorite color vs. the subjects' lowest ranked color, more subjects again chose mugs in their lowest ranked colors instead of mugs in the other player's favorite color in the Self condition than in the Interactive-Control condition (Fisher's test, $p=.046$)—with a reliable interaction across conditions ($F(1, 198)=5.28$, $p=.023$, $n_p{}^2=.011$); see Figure 6b. (There were no differences across either choice pair between the Memory-Control and the Interactive-Control, Fisher's test, $p=.389$ and $p=.172$ respectively).

These results confirm the role of one's own preferences, and not just mere interaction or engagement, in memory and decision-making—an effect that is particularly driven by the misattribution of our lowest ranked preferences to the other person.

General Discussion

One of the most complex features of our representations of others involves the set of preferences they hold. When we get to know another person, whether acquaintance, colleague, or partner, we can discover ways that we are similar to and different from each other, encoded in ways our preferences are similar or different. Previous work has mostly treated the relationship between self and other as all-or-none, with the "other" as simply similar or dissimilar to a subject. Here, we blur this simple distinction, through making the other person's preferences and their relation to ours more complex.

Our results can be summarized in three main findings. First, people successfully identified the other person's preferences in memory and decision-making tasks. But second, and more interestingly, people falsely remembered their colors (especially their lowest ranked colors) as being in the other player's palette—and this was not simply because the colors were the most recent colors people saw (i.e. the Memory-Control condition), or because these colors were interacted with (i.e. the Interactive-Control condition). Third, when making color-based choices for the other, they were more likely to choose their own personal colors (especially their lowest ranked colors) for the other. These results show an intrusion of our preferences in how we represent others: we dissociate from the "other" and falsely ascribe to them our "lesser likes", or our lower-ranked preferences.

This ascription of our lesser likes to the other was surprising to us. Given the work on the egocentric bias (e.g. Van Boven & Loewenstein, 2003), we expected people to assume that their preferences would be shared by others, which should have led them to falsely ascribe their *favorite* colors. Why then are we seeing a different result—where beyond just a mere preference bias, people are instead ascribing their least valued preferences? Because of confusion stemming from a self-other dissociation.

We think our results speak to how we represent ourselves in relation to others. Intuitively, our concept of self involves our preferences—and existing work has shown this to be true across people's own self-descriptions (e.g. Kanagawa, Cross, & Markus, 2001). Moreover, we can understand preferences ordinally, taking some options to be better than others, and this ranking matters for how we understand ourselves. Options that rank higher for a person might seem closer or 'more essential' to who they are, and those that rank lower, further away. Previous work has explored where the self might be 'located' in the physical body (e.g. more associated with one's eyes than with one's feet; Starmans & Bloom, 2012), and here we instead focus on where a self might be located in the representational space of one's preferences. Our results further suggest where the "other" might be located in relation to ourselves. To the extent that we distance ourselves from our lesser likes, this might affect our representation of the preferences of the "other". People might be take preferences that are less essential to who they are as a proxy for this separate person or the "other". If people representationally distance themselves from both their lesser likes and from the "other", this might lead to the confusion that we observed with false memories and faulty decisions. We're curious about how such 'distancing' may be amplified or reduced as we get to know an "other" better.

Here, we focused on color preferences because we think these seemingly inconsequential preferences still form part of people's concepts of self, affecting how they behave and think of themselves. Beyond this, such a test case in a minimal context can also suggest new ways of thinking about how our preferences might interfere with how we represent others, in a way that might be obscured in more complex decision-making tasks. But of course, deciding for others can often also involve morally/politically-colored preferences (e.g. Harman, 2009; Van Boven, Loewenstein, Dunning, & Nordgren, 2013)—sometimes even involving life-and-death decisions, as whether to terminate treatment for someone who cannot decide for themselves. It can become crucial to be able to dissociate ourselves from the other person, especially when their preferences differ from ours. Further work can explore whether we might find simultaneous self-projection and dissociation in these weightier contexts.

When we preserve the complexity of the "other", it seems that we project onto them, but also simultaneously represent them as distinct from us. These together demonstrate one aspect through which we construct a representation of the preferences of the "other" in relation to our own—which can have further consequences for how we ultimately make decisions for them.

References

Ames, D. R. (2004). Strategies for social inference: A similarity contingency model of projection and stereotyping in attribute prevalence estimates. *Journal of Personality and Social Psychology*, *87*, 573–585.

Chartrand, T. L., & Bargh, J. A. (1999). The chameleon effect: The perception–behavior link and social interaction. *Journal of Personality and Social Psychology*, *76*, 893–910.

De Leeuw, J. R. (2015). JsPsych: A javascript library for creating behavioral experiments in a web browser. *Behavior Research Methods*, *47*, 1–12.

Drayton, L. A., Santos, L. R., & Baskin-Sommers, A. (2018). Psychopaths fail to automatically take the perspective of others. *Proceedings of the National Academy of Sciences*, *115*, 3302–3307.

Epley, N., Keysar, B., Van Boven, L., & Gilovich, T. (2004). Perspective taking as egocentric anchoring and adjustment. *Journal of Personality and Social Psychology*, *87*, 327–339.

Gilovich, T., Kruger, J., & Medvec, V. H. (2002). The spotlight effect revisited: Overestimating the manifest variability of our actions and appearance. *Journal of Experimental Social Psychology*, *38*, 93–99.

Harman, E. (2009). "I'll be glad I did it" reasoning and the significance of future desires. *Philosophical Perspectives*, *23*, 177–199.

Kanagawa, C., Cross, S. E., & Markus, H. R. (2001). Who am I? The cultural psychology of the conceptual self. *Personality and Social Psychology Bulletin*, *27*, 90–103.

Keysar, B., Lin, S., & Barr, D. J. (2003). Limits on theory of mind use in adults. *Cognition*, *89*, 25–41.

Nicolle, A., Klein-Flügge, M. C., Hunt, L. T., Vlaev, I., Dolan, R. J., & Behrens, T. E. (2012). An agent independent axis for executed and modeled choice in medial prefrontal cortex. *Neuron*, *75*, 1114–1121.

Starmans, C., & Bloom, P. (2012). Windows to the soul: Children and adults see the eyes as the location of the self. *Cognition*, *123*, 313–318.

Tamir, D. I., & Mitchell, J. P. (2013). Anchoring and adjustment during social inferences. *Journal of Experimental Psychology: General*, *142*, 151–162.

Tarantola, T., Kumaran, D., Dayan, P., & De Martino, B. (2017). Prior preferences beneficially influence social and non-social learning. *Nature Communications*, *8*, 1–14.

Van Boven, L., & Loewenstein, G. (2003). Social projection of transient drive states. *Personality and Social Psychology Bulletin*, *29*, 1159–1168.

Van Boven, L., Loewenstein, G., Dunning, D., & Nordgren, L. F. (2013). Changing places: A dual judgment model of empathy gaps in emotional perspective taking. In *Advances in Experimental Social Psychology* (Vol. 48, pp. 117–171). Elsevier.