THE UNIVERSITY OF CHICAGO

The Idiosyncrasy of Preferences Across Domains

By

Xiaying Deng

July 2022

A paper submitted in partial fulfillment of the requirements for the Master of Arts degree in the
Master of Arts Program in the Social Sciences

Faculty Advisor: Alexander Todorov, Reid Hastie

Preceptor: Rubén Rodriguez

Abstract

There has been much research in the realm of the social sciences in investigating the construction of preferences. However, there is a gap in the literature that examines the stability of our preferences and the idiosyncratic components of preferences within different stimuli domains. Within the current study, the reliability-sharedness of preferences is examined through judgments across different stimulus domains: money gambles, abstract artwork, and faces. Raters in Study 1 were asked to rate 42 money gambles and 42 face stimuli twice on a scale of "Not at all attractive" to "Very attractive". Using Pearson correlations and variance component analysis, the study found that within both domains, preferences tended to be stable. However, the gamble domain was more likely to have shared preferences among the raters than the face domain, where raters had both individual and shared components. Raters in Study 2 were asked to rate 42 money gambles and 42 abstract artworks twice on a scale of "Not at all attractive" to "Very attractive." Using the same methods as done in Study 1, the study found that preferences tended to be stable, though artwork preference reliability was lower than the other two domains. The gamble stimuli results remained consistent to Study 1, but artwork VCA found that preferences tended to be much more individualized than the other domains.

Keywords: Interrater agreement; Judgment; Preferences; Artwork; Gambles; Face perception; Variance component analysis

Introduction

For decades now, social science and the behavioral sciences have been enraptured by understanding the construct of values and their impact on human behavior (Woodruff, 1951): doing so would allow insight on why humans act the way they do across contexts, their motivations, and ultimately allows for the prediction and understanding of future behaviors. Much research in the realm of motivation theorizes towards goal-directed behavior, in which people base their behaviors on what will lead to their desired outcome (Deci & Ryan, 2000). Goals and values then determine individual preferences and decisions. In order to further our understanding of values, it then becomes imperative to understand the output of these values, the preferences and decisions. Research within the realm of decision-making assumes preferences are based on values and goals that lead to purposeful behavior. Decision-making agents in natural environments may have *a priori* preferences that impact the comparison process when faced with making a choice and thus produce default actions (Lopez-Persim, et al., 2016). The present research specifically addresses two fundamental questions about preferences: (A) how stable and reliable are our preferences? (B) How unique, individualized, or shared are our preferences?

Values and preferences have been measured a multitude of ways. One such way is the usage of rating scales, as done in many studies on attraction and attractiveness (Cunningham et al., 1995; Vartanian et al., 2013; Langlois & Roggman, 1990; Dongwon et al., 2020). Another frequently used method is through choice, in which participants are able to specifically pick based on their own individual inclination, as seen in studies regarding sentencing disparity and racial biases (Austin & Williams III, 1977; Sarnikar et al., 2007; Rehavi & Starr, 2014). Within other fields such as marketing, values are frequently measured through a willingness-to-pay method, where respondents are asked to assign a monetary value on an object, service, or experience

(Brookshire & Norum, 2011; Homburg, 2005). In economics, gambles are often used within decision research to examine heuristics such as sunk costs, risk aversion, and rationality (Bateman et al., 2007; Kim et al., 2012; Stetzka & Winter, 2021; Gainsbury et al., 2018). Regardless of the type of measurement used, there are two characteristics necessary that are pertinent to understanding values: reliability and uniqueness. Reliability tests values for identical objects at different points in time. Uniqueness, or idiosyncrasy, is measured by assessing the similarity of different individuals' values across the group. In order to establish the basis of rationality, having predictable changes in preferences is necessary; idiosyncrasy allows us to understand why we differ from one another, in what areas our values are shared, biological, or based on cultural elements. Together, predictions can then be made to predict the behavior of masses like economies.

Studies on social and aesthetic perceptions find idiosyncrasy for similar judgments made by different raters varies across stimulus domains. Kurosu and Todorov (2017) examined the reliability of face and object judgments. Within the study, participants were asked to evaluate two types of stimuli (novel objects, faces) and rated their impressions of the stimuli (how approachable, dangerous, beautiful, likable). The study found that participants often shared impressions for faces, but not always for novel objects. Martinez et al. (2020) were able to replicate these findings using variance component analysis (VCA). VCA allows for the quantification of shared judgments in specific clusters through the estimation of their variance components (Hönekopp, 2006). Observed variance can be attributed to the objects of measurement (participants) or to the facets (stimuli). The analysis calculates variance components of Rater, Stimulus, and Rater x Stimulus. With more Rater agreement, more variance is attributed to the Stimuli, indicating shared taste. If more variance is found within the Rater cluster, it can be concluded that there is a difference between the raters. The Rater x Stimulus cluster factor measures the differences in how participants rated

the stimuli, thus representing the idiosyncratic component of judgments. For the stimulus cluster, some stimuli may just be more attractive and preferable than other stimuli (i.e., most people should prefer to take a gamble where you have a 75% chance of winning $5 and 25% of losing $2 rather than the gamble where you have a 25% chance of winning $5 and 75% chance of losing $7. The interpretation of rater variance can be ambiguous; differences can reflect true behavioral differences or differential use of the scale. For the latter case, when asked to make a judgement on how attractive a face is, Participant 1 may always rate faces higher than Participant 2. Real behavioral differences would exist in the instance of gambling, where if having to play a gamble for a cost, the participant will only do so if they find the gamble to be an attractive one to play upon. The rater by stimulus is the cleanest measure of idiosyncratic variance: for example, Participant 1 likes gamble 1 more than gamble 2, Participant 2 likes gamble 2 more than gamble 1.

Within the current study, the reliability-sharedness of preferences is examined through judgments across different stimulus domains: money gambles, abstract artwork, and faces. When examining idiosyncrasy of preferences and variance, much research has been done to include stimuli such as art and faces; however, such research has rarely used gambles, a type of stimuli often used in behavioral science and economics. If judgments are reliable, then it is then also necessary to examine how shared they are across people and the potential difference in how these preferences are shared. There is empirical evidence in which aesthetic experiences, specifically the experience of beauty, involves also a more objective process. This relates especially to studies on attraction and facial preferences. Langlois and Roggman (1990) find that there is a correlation in finding faces attractive when the facial composition is close to the average attraction of the population, arguing that “attractive faces are only average.” Other studies find that there is a

correlation between face attraction preference to smooth skin texture and skin homogeneity, a correlation in self-rated body rating with preferences to masculine or feminine men, and other standards such as facial symmetry (Fink et al., 2001; Penton-Voak et al., 2003; Little & Perrett, 2002). For many of these studies, judgements are also aggregated across participants, making it difficult to identify idiosyncratic contributions to preferences.

These studies suggest a biological component to how objects get their value. For faces, there is a heavy emphasis on evolutionary theory which then creates beauty standards and norms (Fink et al., 2001; Penton-Voak et al., 2003; Little & Perrett, 2002; Little et al., 2011; Rhodes, 2006; Swami & Furnham, 2008). People are more likely to base their preference to faces off of mate selection criteria and reproductive condition. Other studies point at individual experience affecting preferences (Palmer et al., 2013; Leder et al., 2004; Leder and Nedal, 2014). Elements such as proficiency, culture specific values, and personal experiences may lead to individual taste. Other objects like money and gambles, may have value based more on practicality. People may be more likely to look at expected gain and loss in a more rational manner when confronted with these objects than others. When confronted with a money puzzle, individuals are more likely to think in their own self-interest in order to promote more gain (Vohs et al., 2006; DeScioli et al., 2014). Many monetary gamble questions call upon rational choice theory as an explanation as to why a decision-maker chooses to take or not take a gamble (Reid & Dawes, 2016). There is a more specific focus on logic and an inherent intentionality to be rational in one's choices so that more utility can be gained. Based off of these sources of values, we can hypothesize towards more stable and shared preferences to gambles and faces as compared to artworks.

Research on aesthetic experiences and preferences have found both shared and idiosyncratic components (Leder & Nadal, 2014; Leder et al., 2004). The appreciation of art, for examples, is held to be a very subjective experience in which individuals tend to differ in the type of art that they prefer (Leder et al., 2012). There is support to show that for the degree of art experience, familiarity and novelty matters in how beholders experience the art (Song et al., 2021); people tend to prefer novel abstract pieces rather than representational painting, though this effect may be moderated by the artistic properties as well as differences in the observer. Gender and age differences also modulate preference for works of art, where school boys tend to prefer abstract art and girls tend to prefer realistic art, leading to more individualized experiences of aesthetic appreciation (Salkind., L & Salkind., N, 1997). It is also difficult for beholders of an artwork to all reach an agreement due to varying artistic evaluations and interpretations; this is especially true for abstract works (Palmer et al., 2013; Leder et al., 2004; Vessel & Rubin, 2010; Belke et al., 2006; Leder et al., 2016). While some artists and historians see abstract art as a "universal language", empirical research finds that the perception of abstract paintings creates more intersubjective disagreement than for representational artworks (Brinkmann et al., 2014; Schepman et al., 2015; Leder et al., 2016). This may be due to the lack of semantic content and interpretations, where it is more difficult for a viewer to place a meaning behind an abstract artwork as compared to the Mona Lisa. Rather, for abstract artworks, the beholder primarily hold aesthetic judgment based on the style and composition (Vessel et al., 2010).

Gambles have been used in decision making research as "a vehicle for examining fundamental processes with presumably important implications outside the laboratory" (Bateman et al., 2007; Loewenstein, Weber, Hsee, & Welch, 2001; Peters, Vastfjall, Garling, & Slovic, 2006; Slovic, Finucane, Peters, & MacGregor, 2002). Much research has been done with gamble stimuli

to examine decision making behaviors and more recently on preference construction (Bateman et al. 2007; Lichtenstein & Slovic, 2006). These studies primarily examine individual differences in decision-making and preference construction, i.e., examining whether individuals fall within certain categories of risk adversity, to counter utility theory, to examine emotionality, cognitive ability, personality, etc. Monetary gambles have been found useful to help make predictions for stock market participation and related financial puzzles through narrow framing (Barberis et al., 2006). However, there are few studies of the reliability of preferences on gamble stimuli (Lin et al. 2013; Kim et al., 2012). Using variance analysis, we can further examine this domain. Due to the monetary and rational expectation one may expect with gamble decision-making, we predict for there to be more shared variance in the stimulus cluster as compared to the other clusters. Participants should make a rational choice based on the perceived advantages and disadvantages of the gamble. This can mathematically be expressed as the expected value of the gamble. For example, when given a 25% chance to win $3 (0.25 x 3= 0.75) and a 75% chance to lose $4 (0.75 x 4= 3), the expected value of such a gamble is $-2.25 (0.75 - 3). Within a situation in which an individual is told this gamble but also that they must rate this gamble by how much they like it with no financial consequence to themselves, the individual should rational say they do not like the gamble due to the low expected value.

These three different types of stimuli conceptually differ from each other and would provide good point of comparison relevant to behavior and judgment research. They hold interesting points of comparison when examining values and preferences. Though not tested within this research, the stimuli could have interesting differences in how some objects may induce intuitive judgments while others may involve a more deliberate and analytic judgment.

Aims of the Current Study

We are interested in (A) how stable and reliable are our preferences? (B) How unique, individualized, or shared are our preferences? That is, are people more likely to share similar ratings towards gambles compared to their ratings of art or faces?

Variance component analysis will be used to compare idiosyncratic and shared components of preference ratings between stimulus domains. For instance, we expect when asked to make judgments about gambles, preferences will be more shared than when asked to make judgments about art. Participants are expected to have more shared preferences for gambles due to the mathematical hierarchy of gambles. That is, some gambles are better than most and gambles have an expected value that can be calculated. In contrast, when judging works of art, judgments will be more individualized and less shared among participants.

Overview and Methods

Correlation indices were used to provide descriptive analyses of preferences within the three domains. Inter-rater and intra-rater correlations were calculated in order to describe reliability and rater agreement. Inter-rater correlation was computed between individuals and the group's average ratings across stimuli within domain. With higher inter-rater correlation, we can infer greater consensus and agreement. In order to rule out measurement error and unreliability of individual raters, we computed intra-rater correlations by measuring the consistency of each rater across two time points (Martinez et al., 2020).

To estimate idiosyncratic or shared contributions to judgments, we conducted variance partitioning analysis. Through this method, variance components from clusters (e.g., raters, stimuli, block, etc) within the dataset can be examined. By quantifying the variance components of certain clusters, variance component analysis allows us to ascribe systematic portions of

variance in preference judgments to each component. The size of the variance components can be compared to get a sense of how important different clusters are to the ratings. Higher variance indicates more systematic differences between judgments within the cluster. (Martinez et al., 2020). By examining these cluster variances, we are able to systematically assess the shared and idiosyncratic contributions to preferences. If there is a greater proportion of variance for the stimulus cluster, then there is more consensus across participants on the value of each stimulus; everyone was more likely to rate a certain stimulus highly and a certain stimulus lowly. If more variance is found within the Rater cluster, it can be concluded that there is a difference between the raters, though it is unknown whether that difference lies in different usage of the rating scale or in behavioral differences. The Rater x Stimulus cluster can evaluate idiosyncratic contributions, where a high proportion of variance shows that participants differ in how they personally ranked their preferences for different stimuli.

VCA was estimated using methods in Martinez et al. (2020), in which the Tests in Linear Mixed Effects Models (lmerTest) library and Linear Mixed Effect 4 (lme4) package was used within the R statistical language. Random effects included Rater, Stimulus, and Rater × Stimulus, Block, Block × Rater, and Block × Stimulus clusters. Within the study, participants saw each stimulus type twice in order to generate these analyses.

Once the VCA analysis was conducted, we examined variance partitioning coefficients (VPCs), which represent the ratio of a cluster's variance to the total variance (Martinez et al., 2020; Goldstein et al., 2002). A higher VPC expresses high between-cluster variance and low within-cluster variance. Conversely, as the VPC gets lower and closer to zero, there is more within-cluster variance and less between-cluster variance.

Study 1

In Study 1, we examined two sets of stimuli: money gambles and faces. Participants (n = 100) from the online research company Prolific were recruited for the study. Participants were drawn from a pool of USA participants, with an even distribution of male to female aged 18+ (50 male, 50 female). 67 participants were White, 16 were Black or African American, 7 were Asian, and 10 were of other ethnicity. The average age of participants was 34 years old, with the youngest being 18 and oldest being 60 years old. Before the beginning of the study, participants indicated their consent and upon completion received monetary compensation.

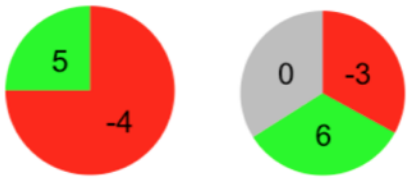
Participants were asked to rate 42 gambles twice and 42 faces twice (84 ratings in each domain). They were asked to rate each stimulus on a 10-point slider scale labeled: "How attractive is this face (gamble)?" with endpoints labeled "not at all attractive" and "very attractive." Each set of stimuli was separated into 2 blocks of 42 gambles or faces that the participants saw in a random order:

- 1st block of faces, 1st block of gambles, 2nd block of faces, 2nd block of gambles
- 1st block of gambles, 1st block of faces, 2nd block of gambles, 2nd block of faces
- 1st block of faces, 1st block of gambles, 2nd block of gambles, 2nd block of faces
- 1st block of gambles, 1st block of faces, 2nd block of faces, 2nd block of gambles

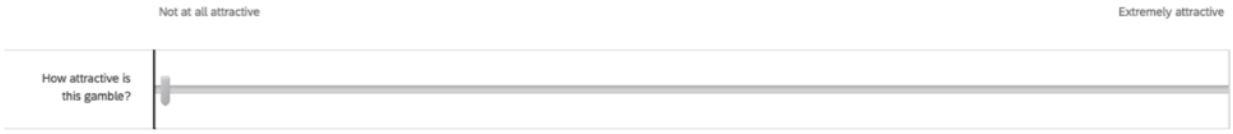
The 42 stimuli within each of the blocks were randomized in order to reduce the chance of participants remembering what they saw as well as reduce possible order confounds. At the end of the survey, the participants were able to play one of the gamble stimuli and had the chance to win a bonus $7 on top of the $5 base pay. They were then asked about the strategies they used while making their ratings as well as for demographic information.

The gamble stimuli (Figure 1) were presented as pie charts, in which each gamble was defined by six characteristics: amount to win (green), probability of winning (green), amount to lose (red), probability of losing (red), a null outcome (gray), and the probability of the null outcome (gray). (see Appendix 1 for a list of the values for the 6 components of each of the 42 stimulus gambles.)

Figure 1. Example gamble stimuli and the rating scale used in Experiments 1 and 2.



(Figure 1. Examples of gamble stimuli)



Within this first study, 21 images of hyper realistic male and 21 hyper realistic female adult faces (Peterson et al., 2022) were used. Like with the gamble stimuli, participants were asked "How attractive is this face? and asked to indicate their rating on the same slider scale (Figure 2).

Figure 2. Example face stimulus and the rating scale used in Experiment 1.

Study 2

In Study 2, we examined two sets of stimuli: money gambles and artworks. Participants (n = 100) from the online research company Prolific were recruited for the study. Participants were drawn from a pool of USA participants, with an even distribution of male to female aged 18+ (50 male, 50 female). 71 participants were White, 10 were Black or African American, 12 were Asian, 1 were American Indian or Alaska Native, and 6 were of other ethnicity. The average age of participants was 34 years old, with the youngest being 18 and oldest being 60 years old. Before the beginning of the study, participants indicated their consent and upon completion received monetary compensation.

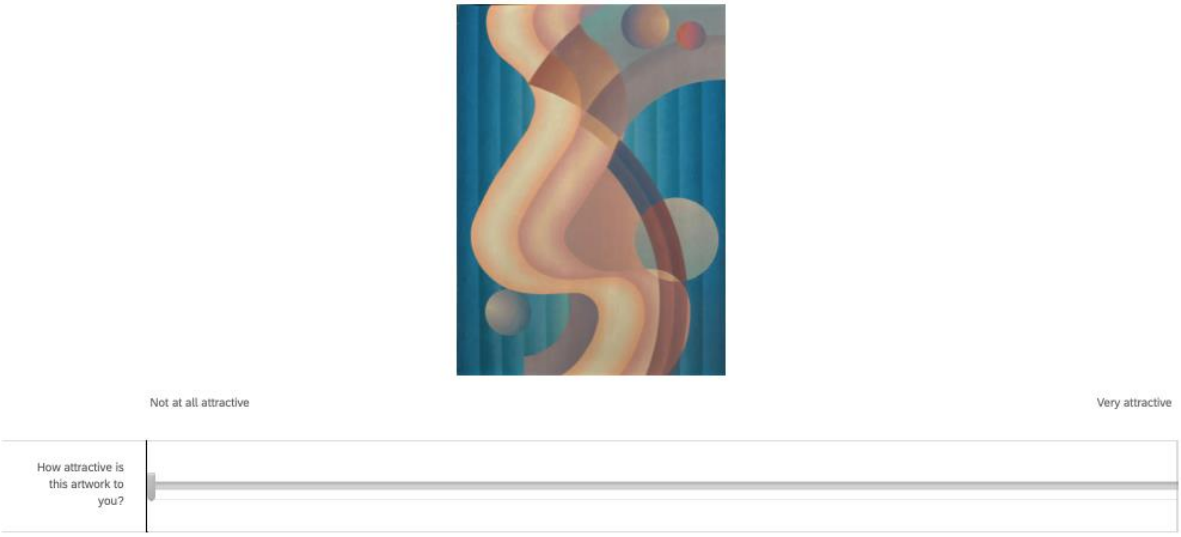
Participants were asked to rate 42 gambles twice and 42 artworks twice (84 ratings in each domain). They were asked to rate each stimulus based on how attractive they found it. Each set of stimuli was separated into 2 blocks of 42 gambles or artworks that the participants saw in a one of the following orders:

- 1st block of artworks, 1st block of gambles, 2nd block of artworks, 2nd block of gambles
- 1st block of gambles, 1st block of artworks, 2nd block of gambles, 2nd block of artworks
- 1st block of artworks, 1st block of gambles, 2nd block of gambles, 2nd block of artwork
- 1st block of gambles, 1st block of artworks, 2nd block of artworks, 2nd block of gambles

The 42 stimuli within each of the blocks were randomized in order to reduce the chance of participants remembering what they saw as well as reduce order confounds. At the end of the survey, the participants were able to play one gamble and had the chance to win a bonus $7 on top of the $5 base pay. They were then asked about the strategies they used while making their ratings as well as for demographic information.

As in Study 1, the gamble stimuli were presented as pie charts, in which participants saw how likely it was to win, lose, or gain nothing (Figure 1). The study used abstract artworks sourced from online museum databases, as using non-abstract artworks may increase the chance of the participant recognizing the work from personal experience prior to the experiment (Figure 3).

Figure 3. Example artwork stimulus from Experiment 2



Analyses

Descriptive statistics and variance partitioning analysis methods were based on Martinez et al. (2020) and Hönekopp (2006). Intra-rater correlations were calculated to measure correlations in ratings for the same individual while the inter-rater correlations measured the consensus between raters.

Table 1. Test-retest correlations, measuring reliability, across artwork, faces, and gamble stimuli.

	Intra-rater Correlation			
	Gambles Study 1	Faces Study 1	Gambles Study 2	Artwork Study 2
Mean	0.80	0.80	0.80	0.63
Standard Error	0.02	0.02	0.02	0.03
Minimum	0.10	0.08	-0.01	-0.33
Maximum	0.99	0.98	0.99	0.96

To calculate the intra-rater correlation, or test-retest correlation, each response from each respondent was examined based on how consistent the respondent was across the two blocks. Pearson's correlations were calculated with an average intra-rater correlation of 0.63 for the artwork and higher values for the gambles (0.80 in the first study and 0.80 for the second study) and face stimuli (0.80). Respondents were consistent in their ratings when asked to rate the same stimuli a second time. In both studies, the test-retest correlations are positively correlated across domains (0.80 for gambles and faces in Study 1, 0.80 for gambles in Study 2, and 0.63 for artworks in Study 2). In other words, participants who were reliable for one domain of stimuli tended to be reliable for the other domain of stimuli.

Table 2. Inter-rater correlations across artwork, faces, and gamble stimuli.

	Inter-rater Correlation			
	Gambles Study 1	Faces Study 1	Gambles Study 2	Artwork Study 2
Mean	0.82	0.67	0.82	0.37
Standard Error	0.01	0.02	0.02	0.02
Minimum	0.09	-0.13	-0.12	-0.27
Maximum	0.95	0.90	0.94	0.79

Pairwise Pearson's correlations between raters as well as correlations between a rater and the group's average ratings were calculated. These calculations yielded high inter-rater correlations for the gamble stimuli in both studies, of about 0.82, suggesting high agreement between the raters. Conversely, inter-rater correlations begin to fall for face stimuli (0.67) and were low for artwork stimuli (0.37), suggesting that a participant tended to rate these stimuli differently when compared to the other participants. This is also observed when examining the variance components.

Study 1 Variance Components

Gamble judgments (left) are largely explained by stimulus VPC (i.e., shared variance). That is, for certain gambles, raters were more likely to reach a consensus. Preferences within the gambles domain are largely shared. Face judgments (right panel) are subject to greater between-rater differences (i.e., potentially idiosyncratic variance) and stimuli VPC (i.e., shared variance). There is also a sizeable Rater x Stimulus effect (i.e., idiosyncratic variance). That is, face judgments can be explained both by individual behavioral differences and shared judgments of certain faces. When comparing the two domains, it seems that the gamble domain's shared preference lies in the nature of the stimuli itself and is not so much influenced by the subject perceiving it. Contrarily, differences and similarities in judgment within the face domain lies in both the subject and the stimuli.

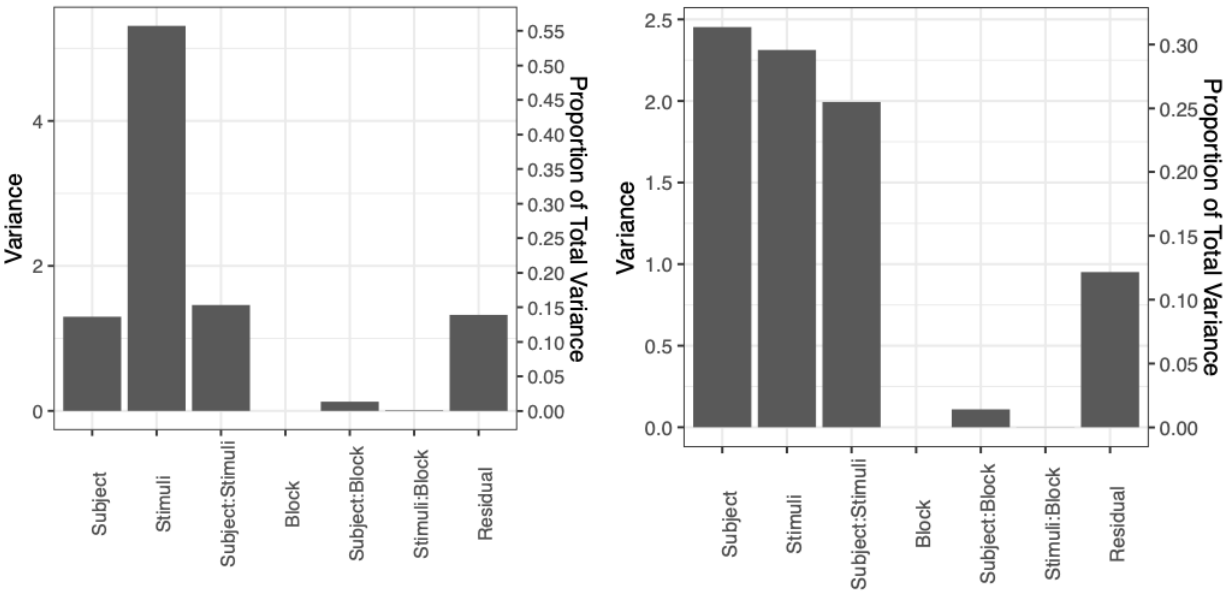


Figure 4. Study 1 gamble variance components (left) and face variance components (right).

Study 2 Variance Components

Gamble judgments (left panel) are again largely explained by stimulus VPC (i.e., shared variance). For gambles, raters were more likely to reach a consensus. Artwork judgments (right panel) are largely explained by between-rater differences (potentially idiosyncratic variance) and

Rater x Stimulus VPC (idiosyncratic variance). That is, preferences within this domain are not shared, the preferences that result in the judgments of liking for artwork are private to the person making the judgment. These results remain supportive of previous studies on aesthetic judgments within the art domain (Leder & Nadal, 2014; Leder et al., 2004).

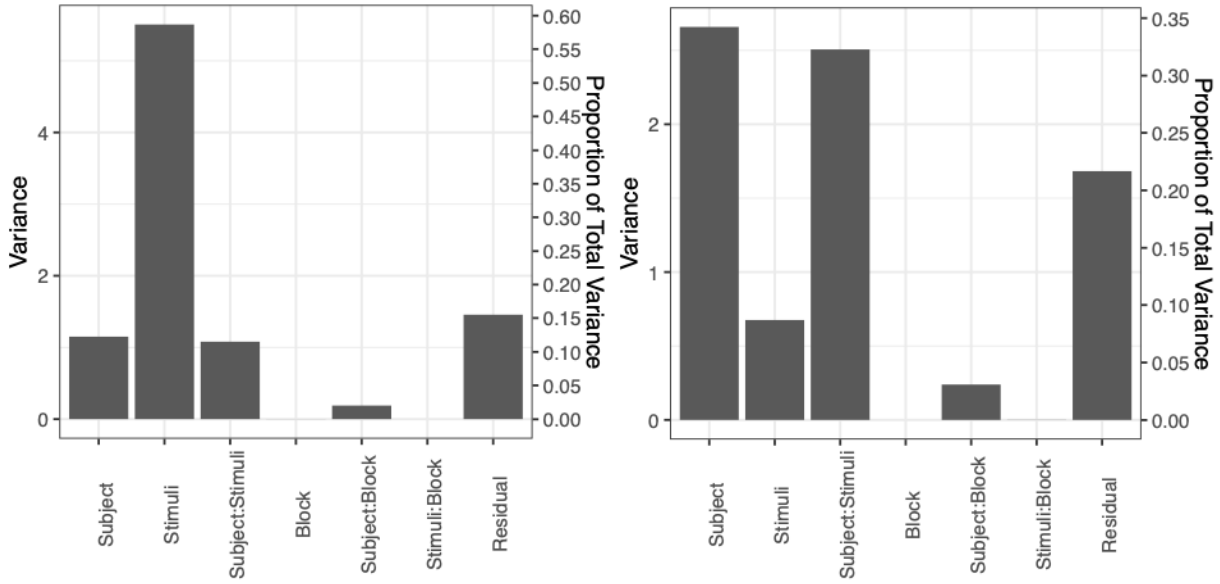


Figure 4. Study 2 gamble variance components (left) and artwork variance components (right).

Results

Based on the analyses of these two studies, we can conclude: (A) our preferences within these domains are quite reliable and that (B) the uniqueness of these preferences depend on the domain. Participants had high intra-rater correlations when asked to rate stimuli again. Interestingly, when examining inter-rater correlations, we found that participants were most likely to rate similarly to the others for gambles (0.82), somewhat likely for faces (0.67), and least likely for artwork (0.37). When examining variance components, we see this result reflected. Preferences within the gamble domain tended to be shared, with a large proportion of variance being explained by the stimuli component; participants were more likely to share preferences on certain gambles. Preferences within the face domain could be explained by individual behavioral differences and

shared judgments of certain faces; there was greater between-rater differences, shared variance (seen in the stimuli component), and idiosyncratic variance (seen in the rater by stimulus component). Artwork judgments are largely explained by between-rater differences (potentially idiosyncratic variance) and Rater x Stimulus VPC (idiosyncratic variance), showing that preferences within this domain are not as collective as compared to the other two.

Future Research

Future research can examine objectivity and subjectivity as well as the ways in which analytic and intuitive thinking may impact values. Objective domains can be defined as those in which the majority of people believe that only one claim or perspective is correct or should be most preferred. Cheek et al. (2019) hold that there is some inherent logic to viewing one's own judgments as objective. Implicitly, in an objective domain analytical reasoning should converge to a single "true," objective conclusion. Subjective domains, on the other hand, are those where people think there is more than one correct way of looking at things or in which there is no one correct way of looking at things at all (Cheek, et al., 2019). Such domains would include works of art. An evaluation of an artist's work is believed to be an expression of the observer's own personal experiences, knowledge, preferences, and feelings (Sandelands & Buckner, 1989). The link between an art piece and the viewer is then part of a unique aesthetic experience. Strong emotional connection between the artist and her work directs viewers' attention to the subject matter. However, individual preferences towards the artwork are not held to a general consensus; viewers generally form their own opinions about various aspects of the art. Comparing different aesthetic domains, Vessel, et al. (2017) found that there tended to be stronger shared tastes for naturally varying stimuli, such as facts, compared to domains relating to culturally generated aesthetic principles (i.e., artwork).

Studies on preference also examine intuitive modes of decision-making in which intuition, "can sometimes produce attitudes and judgments that are superior to those generated from conscious, analytic thinking" (Halberstadt, 2010) and where analytical thinking can "degrade decision quality" (McMackin & Slovic, 2000). Further research done by Wilson and Schooler (1991) shows similar findings, in which they found that preferences for jam differed from expert opinions when participants were asked to analyze the jam, but were close to the experts' opinions when they chose and evaluated intuitively. These results suggest that deliberate analysis causes changes in how people weigh criteria: they focus on criteria that would not have originally been weighted heavily. Through conscious introspection, people may also come to be less satisfied with their choices (Wilson et al., 1993).

When using these heuristics on the stimuli within the present research, gamble stimuli may be considered more objective and analytic while face and artworks would be the subjective and intuitive. Global values of individual gambles can be defined by the expected value of each gamble. And experts agree that expected value is a rational standard for values and preferences. Natural comprehension of the gambles is systematic and deliberate, making the evaluation process relatively analytic. In contrast, judgments of artwork are more subjective. Natural perceptions of artwork are relatively intuitive, relying heavily on only partly conscious and often emotionally rich evaluation processes.

Conclusion

Within the studies, it is evident that there is an idiosyncratic component in terms of what participants tended to prefer. For the gamble stimuli, which may be considered more objective and provoke analytical processing, participants had shared preferences. Certain gambles had a consensus in the ratings. There was more idiosyncratic variance when participants were asked to

view faces, where the variance in answers depended on both individual differences and on some shared judgments. The highest idiosyncratic component was seen for the artwork judgments, in which the preferences of raters were largely not shared at all. From these results, it appears to be the case that judgments across both stimuli groups can both be reliable; however, gamble judgments are more so reliable. There is an idiosyncratic component in preferences between individual respondents; and sharedness varies depending on whether the preferences are based on which of the stimuli groups, with more shared preferences for the gamble stimuli. These results reflect inter-rater correlations, where there is highest inter-rater agreement for gambles, intermediate for faces, and lowest for abstract art. That is, each participant tended to have high agreement with the group for judgments on gambles, intermediate agreement with the group for faces, and lowest group agreement for the artwork stimuli. For intra-rater correlations, gambles and faces had similarly high correlations while abstract art had the lowest.

Implications and discussion

The stability of preferences is assumed in many theoretical analyses (Loewenstein & Agnner, 2003; Lichtenstein & Slovic, 2006), but there is a distinct lack of relevant behavioral research examining the reliability of preferences within the domains explored as well as less research on the variance partitioning of gambles, faces, and artwork. The present research also provides an important methodological foundation for past research on the accuracy or usefulness of preference judgments (Halberstadt, 2010; McMackin & Slovic, 2000; Wilson & Schooler, 1991; and Wilson, Lisle, Schooler, Hodges, Klaaren, & LaFleur, 1993). The reliability of preference judgments provides an important base line for the assessment of accuracy. If the preference judgments are unreliable, individuals will be unable to forecast their future preferences, and the possibility defining rational choices is called into question.

By contributing behavioral research to this topic, further analysis of other domains may be explored. Fields such as art therapy or museum curation can directly utilize this particular research to advance practices. The practical value of assessing the reliability and idiosyncrasy of consumers' preferences is obvious in applications to marketing. Overall, the present research will help expand our knowledge on how people make decisions in many domains, by providing precise general methods to make cross-domain comparisons of two important properties of all decisions: reliability or stability and idiosyncrasy or sharedness.

References

Asahel D. Woodruff (1952) The Roles of Value in Human Behavior, The Journal of Social Psychology, 36:1, 97-107, DOI: 10.1080/00224545.1952.9919093

Austin, W., & Williams, T. A. (1977). A Survey of Judges' Responses to Simulated Legal Cases: Research Note on Sentencing Disparity. *The Journal of Criminal Law and Criminology (1973-)*, *68*(2), 306–310. https://doi.org/10.2307/1142852

Barberis, N. (2006). Individual Preferences, Monetary Gambles, and Stock Market Participation: A Case for Narrow Framing. *The American economic review*, *96*(4),.

Bateman, I., Dent, S., Peters, E., Slovic, P., & Starmer, C. (2007). The affect heuristic and the attractiveness of simple gambles. *Journal of Behavioral Decision Making*, *20*(4), 365–380. https://doi.org/10.1002/bdm.558

Belke, B., Leder, H., and Augustin, D. (2006). Mastering style: effects of explicit style-related information, art knowledge and affective state on appreciation of abstract paintings. *Psychol. Test Assess. Model.* 48, 115–134.

Brinkmann, H., Commare, L., Leder, H., and Rosenberg, R. (2014). Abstract art as a universal language?

Cheek, N. N., Blackman, S. F., & Pronin, E. (2021). Seeing the subjective as objective: People perceive the taste of those they disagree with as biased and wrong. *Journal of Behavioral Decision Making*, *34*(2), 167-182. https://doi.org/10.1002/bdm.2201

Cunningham, M. R., Roberts, A. R., Barbee, A. P., Druen, P. B., & Wu, C.-H. (1995). "Their ideas of beauty are, on the whole, the same as ours": Consistency and variability in the cross-cultural perception of female physical attractiveness. *Journal of Personality and Social Psychology*, *68*(2), 261–279. https://doi.org/10.1037/0022-3514.68.2.261

DeScioli, P., Massenkoff, M., Shaw, A., Petersen, M. B., & Kurzban, R. (2014). Equity or equality? Moral judgments follow the money. *Proceedings. Biological sciences*, *281*(1797), 20142112. https://doi.org/10.1098/rspb.2014.2112

Edward L. Deci & Richard M. Ryan (2000) The "What" and "Why" of Goal Pursuits: Human Needs and the Self-Determination of Behavior, Psychological Inquiry, 11:4, 227-268, DOI: 10.1207/S15327965PLI1104_01

Fink, B., Grammer, K., & Thornhill, R. (2001). Human (Homo sapiens) facial attractiveness in relation to skin texture and color. *Journal of comparative psychology (Washington, D.C. : 1983)*, *115*(1), 92–99. https://doi.org/10.1037/0735-7036.115.1.92

Gainsbury, S. M., Tobias-Webb, J., & Slonim, R. (2018). Behavioural economics and gambling: A new paradigm for approaching harm-minimization. Gaming Law Review, 22(10). https://doi.org/10.1089/glr2.2018.22106

Goldstein, H., Browne, W., & Rasbash, J. (2002). Partitioning variation in multilevel models. *Understanding Statistics*, *1*(4), 223–231. https://doi.org/10.1207/S15328031US0104_02

Ha-Brookshire, J.E., & Norum, P.S. (2011). Willingness to pay for socially responsible products: case of cotton apparel. *Journal of Consumer Marketing, 28,* 344-353.

Halberstadt, J. B. (2010). Intuition: Dumb but lucky. Fortuitous affective cues and their disruption by analytic thought. Social and Personality Psychology Compass, 4, 64– 76. https://doi.org/10.1111/j.1751-9004.2009.00242.x

Hastie, R., & Dawes, R. M. (2016). *Sage Publications, Inc.*

Homburg, C., Hoyer, W.D. and Koschate, N. (2005) Customers' Reactions to Price Increases: Do Customer Satisfaction and Perceived Motive Fairness Matter? Journal of the Academy of Marketing Science, 33, 36-49. http://dx.doi.org/10.1177/0092070304269953

Hönekopp J. (2006). Once more: is beauty in the eye of the beholder? Relative contributions of private and shared taste to judgments of facial attractiveness. *Journal of experimental psychology. Human perception and performance*, *32*(2), 199–209. https://doi.org/10.1037/0096-1523.32.2.199

Kim, B. E., Seligman, D., & Kable, J. W. (2012). Preference Reversals in Decision Making Under Risk are Accompanied by Changes in Attention to Different Attributes. *Frontiers in neuroscience*, *6*, 109. https://doi.org/10.3389/fnins.2012.00109

Kurosu, A., & Todorov, A. (2017). The shape of novel objects contributes to shared impressions. *Journal of Vision, 17*(13), Article 14. https://doi.org/10.1167/17.13.14

Langlois, J. H., & Roggman, L. A. (1990). Attractive faces are only average. *Psychological Science, 1*(2), 115–121. https://doi.org/10.1111/j.1467-9280.1990.tb00079.x

Leder, H., Belke, B., Oeberst, A., & Augustin, D. (2004). A model of aesthetic appreciation and aesthetic judgments. *British journal of psychology (London, England : 1953), 95*(Pt 4), 489–508. https://doi.org/10.1348/0007126042369811

Leder, H., Gerger, G., Dressler, S. G., and Schabmann, A. (2012). How art is appreciated. *Psychol. Aesthet. Creat. Arts* 6, 2–10. doi: 10.1037/a0026396

Leder, H., Goller, J., Rigotti, T., & Forster, M. (2016). Private and Shared Taste in Art and Face Appreciation. *Frontiers in human neuroscience, 10*, 155. https://doi.org/10.3389/fnhum.2016.00155

Leder, H., & Nadal, M. (2014). Ten years of a model of aesthetic appreciation and aesthetic judgments : The aesthetic episode - Developments and challenges in empirical aesthetics. *British journal of psychology (London, England : 1953), 105*(4), 443–464. https://doi.org/10.1111/bjop.12084

Lichtenstein, S. & Slovic, Paul. (2006). The construction of preference: An overview. The Construction of Preference. 1-40.

Lin, C. H., Song, T. J., Chen, Y. Y., Lee, W. K., & Chiu, Y. C. (2013). Reexamining the validity and reliability of the clinical version of the iowa gambling task: evidence from a normal subject group. *Frontiers in psychology*, *4*, 220. https://doi.org/10.3389/fpsyg.2013.00220

Little, A. C., & Perrett, D. I. (2002). Putting beauty back in the eye of the beholder. *The Psychologist, 15*(1), 28–32.

Loewenstein, G., & Angner, E.. (2003). Predicting and indulging changing preferences. Time and Decision: Economic and Psychological Perspectives on Intertemporal Choice. 351-391.

Loewenstein, G. F., Weber, E. U., Hsee, C. K., & Welch, N. (2001). Risk as feelings. *Psychological Bulletin, 127*(2), 267–286. https://doi.org/10.1037/0033-2909.127.2.267

Lopez-Persem, A., Domenech, P., & Pessiglione, M. (2016). How prior preferences determine decision-making frames and biases in the human brain. *eLife, 5*, e20317. https://doi.org/10.7554/eLife.20317

Martinez, J.E., Funk, F. & Todorov, A. Quantifying idiosyncratic and shared contributions to judgment. *Behav Res* 52, 1428–1444 (2020). https://doi.org/10.3758/s13428-019-01323-0

McMackin, J., & Slovic, P. (2000). When does explicit justification impair decision making? *Applied Cognitive Psychology, 14*(6), 527–541. https://doi.org/10.1002/1099-0720(200011/12)14:6<527::AID-ACP671>3.0.CO;2-J

Oh, D., Walker, M., & Freeman, J. B. (2020, December 9). Person knowledge shapes face identity perception. https://doi.org/10.31234/osf.io/sdup8

Palmer, S. E., Schloss, K. B., & Sammartino, J. (2013). Visual aesthetics and human preference. *Annual review of psychology*, *64*, 77–107. https://doi.org/10.1146/annurev-psych-120710-100504

Penton-Voak, I. S., Little, A. C., Jones, B. C., Burt, D. M., Tiddeman, B. P., & Perrett, D. I. (2003). Female condition influences preferences for sexual dimorphism in faces of male humans (Homo sapiens). *Journal of comparative psychology (Washington, D.C. : 1983)*, *117*(3), 264–271. https://doi.org/10.1037/0735-7036.117.3.264

Peters, E., Västfjäll, D., Gärling, T., & Slovic, P. (2006). Affect and Decision Making: A "Hot" Topic. *Journal of Behavioral Decision Making*, *19*(2), 79–85. https://doi.org/10.1002/bdm.528

Peterson, J. C., Uddenberg, S., Griffiths, T. L., Todorov, A., & Suchow, J. W. (2022). Deep models of superficial face judgments. *Proceedings of the National Academy of Sciences of the United States of America*, *119*(17), e2115228119. https://doi.org/10.1073/pnas.2115228119

Rehavi, M. & Starr, S. (2014). "Racial Disparity in Federal Criminal Sentences. Journal of Political Economy. 122. 10.1086/677255.

Rhodes G. (2006). The evolutionary psychology of facial beauty. *Annual review of psychology*, *57*, 199–226. https://doi.org/10.1146/annurev.psych.57.102904.190208

Sandelands, L. (1988). The concept of work feeling. Journal for the Theory of Social Behavior. 18. 437-457.

Salkind, L., & Salkind, N. J. (1997). Gender and Age Differences in Preference for Works of Art. *Studies in Art Education*, *38*(4), 246–256. https://doi.org/10.2307/1320524

Schepman, A., Rodway, P., Pullen, S. J., & Kirkham, J. (2015). Shared liking and association valence for representational art but not abstract art. *Journal of vision*, *15*(5), 11. https://doi.org/10.1167/15.5.11

Song, J., Kwak, Y., & Kim, C. Y. (2021). Familiarity and Novelty in Aesthetic Preference: The Effects of the Properties of the Artwork and the Beholder. *Frontiers in psychology*, *12*, 694927. https://doi.org/10.3389/fpsyg.2021.694927

Slovic, P., Finucane, M., Peters, E., & MacGregor, D. G. (2002). The affect heuristic. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 397–420). Cambridge University Press. https://doi.org/10.1017/CBO9780511808098.025

Sorensen, T., Sarnikar, S. & Oaxaca, R. (2014). Do You Receive a Lighter Prison Sentence Because You Are a Woman or a White? An Economic Analysis of the Federal Criminal Sentencing Guidelines. *The B.E. Journal of Economic Analysis & Policy*, *14*(1), 1-54. https://doi.org/10.1515/bejeap-2012-0057

Stetzka, R.M., & Winter, S. (2021). How rational is gambling? *Journal of Economic Surveys*.

Swami, V., & Furnham, A. (2008). *The psychology of physical attraction.* Routledge/Taylor & Francis Group.

Vartanian, O., Navarrete, G., Chatterjee, A., Fich, L. B., Leder, H., Modroño, C., Nadal, M., Rostrup, N., & Skov, M. (2013). Impact of contour on aesthetic judgments and approach-avoidance decisions in architecture. *Proceedings of the National Academy of Sciences of the United States of America*, *110 Suppl 2*(Suppl 2), 10446–10453. https://doi.org/10.1073/pnas.1301227110

Vessel, E. A., Isik, A. I., Belfi, A. M., Stahl, J. L., & Starr, G. G. (2019). The default-mode network represents aesthetic appeal that generalizes across visual domains. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(38), 19155–19164. https://doi.org/10.1073/pnas.1902650116

Vessel, E. A., & Rubin, N. (2010). Beauty and the beholder: highly individual taste for abstract, but not real-world images. *Journal of vision*, *10*(2), 1–14. https://doi.org/10.1167/10.2.18

Vohs, K. D., Mead, N. L., & Goode, M. R. (2006). The psychological consequences of money. *Science (New York, N.Y.)*, *314*(5802), 1154–1156. https://doi.org/10.1126/science.1132491

Wilson, T. D., Hodges, S. D., & LaFleur, S. J. (1995). Effects of introspecting about reasons: Inferring attitudes from accessible thoughts. *Journal of Personality and Social Psychology*, *69*(1), 16–28. https://doi.org/10.1037/0022-3514.69.1.16

Wilson, T. D., & Schooler, J. W. (1991). Thinking too much: introspection can reduce the quality of preferences and decisions. *Journal of personality and social psychology*, *60*(2), 181–192. https://doi.org/10.1037//0022-3514.60.2.181

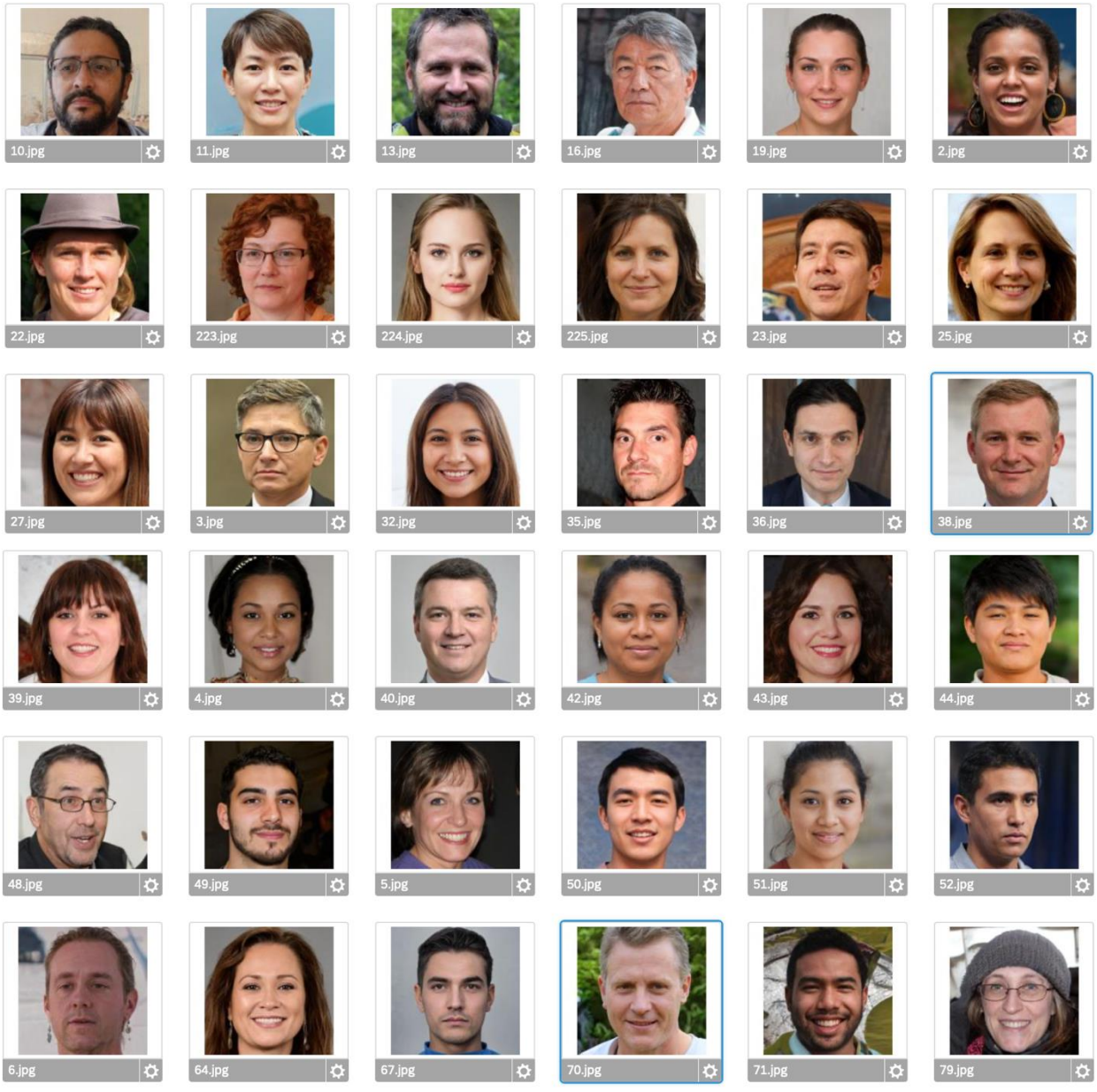
Appendix 1

Gamble stimuli

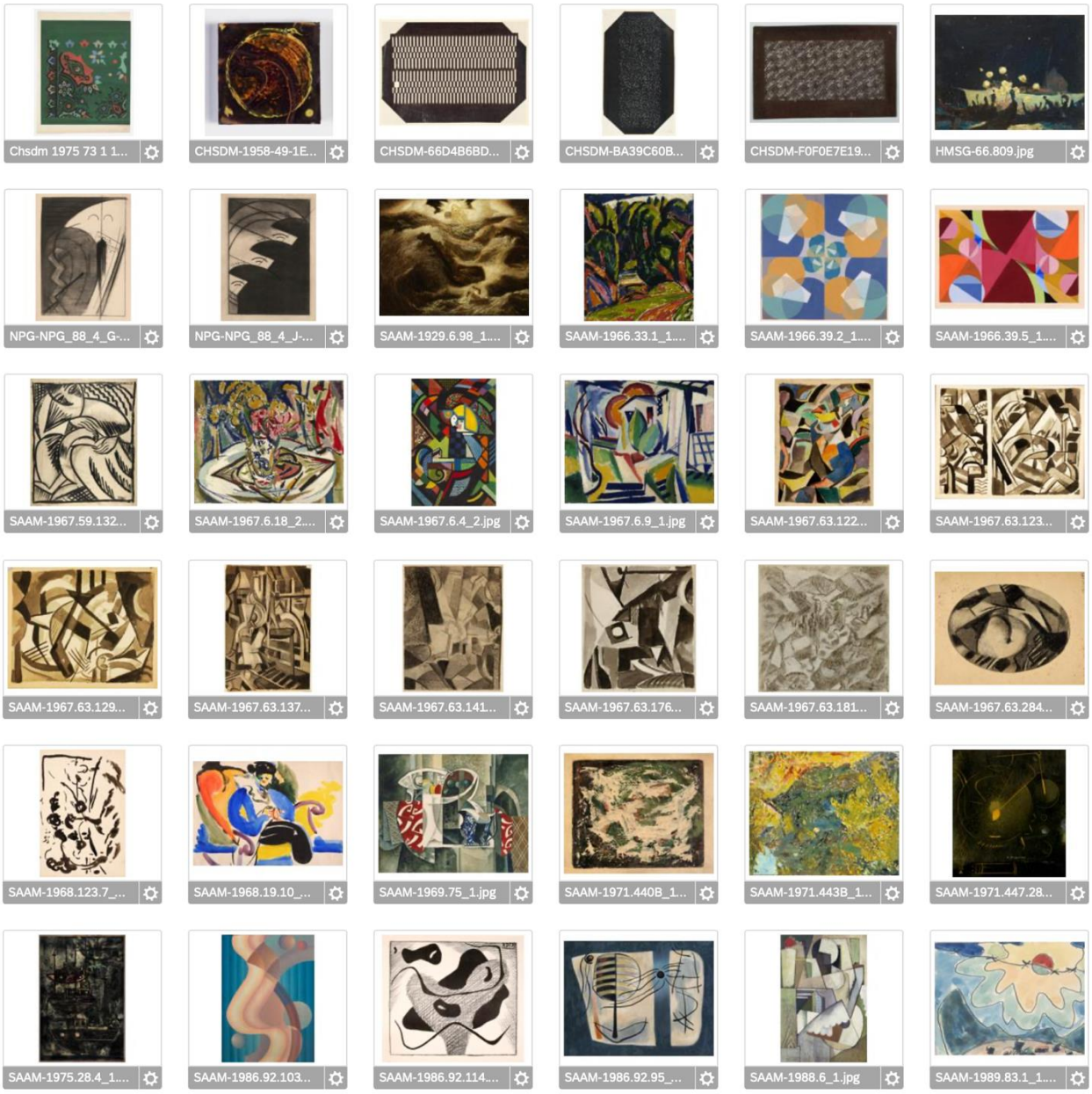
ID	WIN	LOSS	pLOSS	pWIN	pNULL	EV
1	3	-4	0.75	0.25	0	-2.3
2	3	-4	0.5	0.5	0	-0.5
3	3	-4	0.25	0.75	0	1.3
4	3	-4	0.5	0.17	0.33	-1.5
5	3	-4	0.34	0.33	0.33	-0.37
6	3	-4	0.17	0.5	0.33	0.82
7	3	-2	0.75	0.25	0	-0.75
8	3	-2	0.5	0.5	0	0.5
9	3	-2	0.25	0.75	0	1.8
10	3	-2	0.5	0.17	0.33	-0.49
11	3	-2	0.34	0.33	0.33	0.31
12	5	-2	0.17	0.5	0.33	1.2
13	5	-6	0.75	0.25	0	-3.3
14	5	-6	0.5	0.5	0	-0.5
15	5	-6	0.25	0.75	0	2.3
16	5	-6	0.5	0.17	0.33	-2.2
17	5	-6	0.34	0.33	0.33	-0.39
18	5	-6	0.17	0.5	0.33	1.5
19	5	-4	0.75	0.25	0	-1.8
20	5	-4	0.5	0.5	0	0.5
21	5	-4	0.25	0.75	0	2.8
22	5	-4	0.5	0.17	0.33	-1.1
23	5	-4	0.34	0.33	0.33	0.29
24	5	-4	0.17	0.5	0.33	1.8
25	5	-2	0.75	0.25	0	-0.25
26	5	-2	0.5	0.5	0	1.5
27	5	-2	0.25	0.75	0	3.3
28	5	-2	0.5	0.17	0.33	-0.15
29	5	-2	0.34	0.33	0.33	0.97
30	5	-2	0.17	0.5	0.33	2.2
31	7	-6	0.75	0.25	0	-2.8
32	7	-6	0.5	0.5	0	0.5
33	7	-6	0.25	0.75	0	3.8
34	7	-6	0.5	0.17	0.33	-1.8

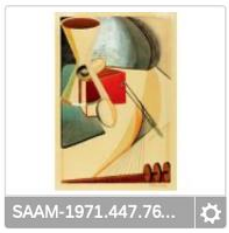
35	7	-6	0.34	0.33	0.33	0.27
36	7	-6	0.17	0.5	0.33	2.5
37	7	-4	0.75	0.25	0	-1.3
38	7	-4	0.5	0.5	0	1.5
39	7	-4	0.25	0.75	0	4.3
40	7	-4	0.5	0.17	0.33	-0.81
41	7	-4	0.34	0.33	0.33	0.95
42	7	-4	0.17	0.5	0.33	2.8

Face Stimuli



Artwork Stimuli




SAAM-1971.447.76...


SAAM-1966.39.6_1....


SAAM-1993.22.1_1....


SAAM-1993.22.2_1....


SAAM-2005.5.25_1....


SAAM-1968.123.11...


HMSG-HMSG_86.7...