Facial Similarity Analysis: A Three-Way Decision Perspective

Daryl H. Hepting, Hadeel Hatim Bin Amer, and Yiyu Yao

1 Introduction

A fundamental task in the sorting of facial photographs is the modelling of pairwise facial similarity. A three-way analysis of facial similarity described in this work uses data obtained from card sorting of a set of facial photographs, done by a group of participants. Participants were asked to sort the photographs into an unrestricted number of piles, using their own judgements of similarity to place similar photos into the same pile. Photos placed into different piles are considered to be dissimilar. In particular, each participant compared the photo to be sorted with the last photo placed on top of each pile. The decision faced by each participant is to add the photo to an existing pile or to create a new pile. Given the lack of an objective standard for judging similarity, different participants may be using different strategies in judging the similarity of photos. It could be very useful to identify and study these strategies.

An overall evaluation of similarity can be obtained by synthesizing judgments from the set of participants. A two-way analysis classifies a pair of photos as either similar or dissimilar. This may be too restrictive. Motivated by the three regions in rough set theory, in this work we present a framework for three-way analysis of judgments of facial similarity. Based on judgments by the set of participants, we divide all pairs of photos into three classes: a set of similar pairs that are judged primarily as similar; a set of dissimilar pairs that are judged primarily as dissimilar; and a set of undecidable pairs that have conflicting judgments. A more refined three-way classification method is also suggested based on a quantitative description of

Daryl H. Hepting
Department of Computer Science, University of Regina, Regina, S4S 0A2, Canada
e-mail: hepting@cs.uregina.ca

Hadeel Hatim Bin Amer
Department of Computer Science, University of Regina, e-mail: binamerh@cs.uregina.ca

Yiyu Yao
Department of Computer Science, University of Regina, e-mail: yyao@cs.uregina.ca

the quality of similarity judgments. The classification in terms of three classes provides an effective method to examine the notions of similarity, dissimilarity, and disagreement. Within the framework, we examine the following two issues:

- Connections between rough sets, shadowed sets, and three-way decisions. We will argue that three-way decisions provide a generalization of rough sets and shadowed sets and offer a new perspective for facial similarity analysis.
- Three-way quantitative analysis of human similarity judgments, beyond a focus on the entire group of participants. We also consider a three-way analysis of individuals, as well as groupings of individuals.

The results shed new lights on our understanding of similarity in both a subjective and an objective setting.

There are at least two possible classes of methods for studying similarity. One class is based on similarity measures on facial photographs. The other class relies on human judgments. Although the former class can be easily automatized, the selection of a semantically meaningful similarity measure is a challenge. On the other hand, the latter class explores human perception of similarity. A semantically meaningful similarity measure must be informed by the human perception of similarity, therefore studies of human judgments of similarity are critical in formulating a semantically meaningful similarity measure.

Unlike tasks of sorting based on shapes, colours of shapes, or numbers of shapes (as in the Wisconsin Card Sorting Test [2]) which have objectively correct answers, the judgment of facial similarity is more subjective. With the sorting of facial photographs, a desired outcome is information about how similarity between faces was judged. Because different participants in the card sorting task may apply different strategies, it may be difficult to identify the information about the similarity judgments. Based on a recently proposed theory of three-way decisions [10, 11], the main objective of this work is to suggest a framework of three-way analysis of human judgments of similarity.

Suppose a group of participants provides similarity judgments of a set of facial photos. Based on their judgments, we divide all pairs of photos into three classes: a set of similar pairs that are judged by 60% of the participants as similar, a set of dissimilar pairs that are judged by 60% of the participants as dissimilar, and a set of undecidable pairs about which the participants have disagreed (less than 60% of the participants judged the pair as similar and less than 60% of the participants judge the pair as dissimilar). Based on this three-way classification, we can identify several research problems, such as: human perception of similarity versus dissimilarity, comparative studies of similar and dissimilar photos, and many more. In this work, we report our preliminary results from applying the theory and performing experiments with three-way analysis of similarity based on human judgments.

The remainder of the Chapter is organized as follows. Section 2 presents a trisecting-and-acting model of three-way decisions as a general framework. Section 3 explores the connections between rough sets and three-way decisions and interprets probabilistic rough sets and shadowed sets in terms of three-way decisions. Section 4 presents the analysis of facial similarity judgments as an application of the trisecting-and-acting model of three-way decisions. Section 5 presents results from analyzing a dataset obtained from card sorting. Conclusions and opportunities for future work are identified in Section 6.

2 A Trisecting-and-Acting Model of Three-Way Decisions

In a nutshell, the basic ideas of three-way decisions are thinking and problem-solving about threes [11], representing a commonly and widely used human practice in perceiving and dealing with complex worlds. In other words, we typically divide a whole or a complex problem into three relatively independent parts and design strategies to process each of the three parts. Yao [11] proposed a trisecting-and-acting (T&A) model of three-way decisions. By following this basic idea of three-way decisions, we put forward here a model of three-way analysis of facial photos.

To grasp the idea of thinking in threes, let us consider three examples. Marr [6] suggested that an in-depth understanding of any information processing can be understood at three levels: the computational theory level, the representation and algorithm level, and the hardware implementation level. Each level focuses on a different aspect of information processing. Kelly [5] presented a three-era framework for studying the past, present, and future of computing: the tabulating era (1900s-1950s), the programming era (1950s-present), and the cognitive era (2011-). This framework helps us to identify the main challenges and objectives of today's computing. Many taxation systems categorically classify citizens as low, middle, or high income, with different taxation methods applied to each.

An evaluation-based trisecting-and-acting model of three-way decisions consists of two components [10, 11]. According to the values of an evaluation function, we first trisect a universal set into three pair-wise disjoint regions. The result is a weak tri-partition or a trisection of the set. With respect to a trisection, we design strategies to process the three regions individually or jointly. The two components of trisecting and acting are both relatively independent and mutually supportive. Effectiveness of the strategies of action depends on the appropriateness of the trisection; a good trisecting method relies on knowledge about how the resulting trisection is to be used. It is important to search for the right combination of trisection and strategies of action.

Let OB denote a finite set of objects. Suppose $v : OB \longrightarrow \Re$ is an evaluation function over OB, where $\Re$ is the set of real numbers. For an object $x \in OB$, $v(x)$ is called the evaluation status value (ESV) of x. Intuitively, the ESV of an object quantifies the object with respect to some criteria or objectives. To obtain a trisection of OB, we require a pair of thresholds (α, β), $\alpha, \beta \in \Re$, with $\beta < \alpha$. Formally, three regions of OB are defined by:

$$\begin{aligned} \mathrm{R}_l(v) &= \{x \in OB \mid v(x) \leq \beta\}, \\ \mathrm{R}_m(v) &= \{x \in OB \mid \beta < v(x) < \alpha\}, \\ \mathrm{R}_h(v) &= \{x \in OB \mid v(x) \geq \alpha\}. \end{aligned} \tag{1}$$

They correspond to subsets of objects with low, middle, and high ESVs, respectively. The three regions satisfy the following properties:

(i) $\mathrm{R}_l(v) \cap \mathrm{R}_m(v) = \mathrm{R}_l(v) \cap \mathrm{R}_h(v) = \mathrm{R}_m(v) \cap \mathrm{R}_h(v) = \emptyset$,
(ii) $\mathrm{R}_l(v) \cup \mathrm{R}_m(v) \cup \mathrm{R}_h(v) = OB$.

It should be noted that one or two of the three regions may in fact be the empty set. Thus, the family of three regions is not necessarily a partition of *OB*. We call the triplet $(\mathrm{R}_l(v), \mathrm{R}_m(v), \mathrm{R}_h(v))$ a weak tri-partition or a trisection of *OB*.

A trisection is determined by a pair of thresholds (α, β). Based on the physical meaning of the evaluation function v, we can formulate the problem of finding a pair of thresholds as an optimization problem [10]. In order words, we search for a pair of thresholds that produces an optimal trisection according an objective function.

Once we obtain a trisection, we can devise strategies to act on the three regions. We can study properties of objects in the same region. We can compare objects in different regions. We can form strategies to facilitate the movement of objects between regions [1]. There are many opportunities working with a trisection of a universal set of objects.

Let us use the example of a taxation system again to illustrate the main ideas of the trisecting-and-acting model of three-way decisions. In this case, *OB* is the set of all citzens who pay tax. The evaluation function is the income of a citizen in dollars. Suppose that a pair of thresholds (α, β) is given in terms of these dollars, say \$35k and \$120k, respectively. The three regions $\mathrm{R}_l(v)$, $\mathrm{R}_m(v)$, and $\mathrm{R}_h(v))$ represent the low income (i.e., $income \leq$ \$35k), middle income (i.e., \$35k $<$ *income* $<$ \$120k), and high income (i.e., $income \geq$ \$120k), respectively. For the three levels of income, one typically devises different formulas or rates to compute tax.

3 Rough Sets, Fuzzy Sets, Shadowed Sets, and Three-Way Decisions

Three-way decisions provide a general framework to unify ideas from several theories for modelling uncertainty. Although the introduction of the concept of three-way decisions was motivated by the notion of the three regions of rough sets, its recent developments are far beyond rough sets. To gain insights into three-way analysis of facial similarity, we interpret probabilistic rough sets and shadowed sets in terms of three-way decisions.

3.1 Probabilistic Rough Sets as Three-Way Decisions

In formulating probabilistic rough sets [13, 14], we start with an equivalence relation E on the set of objects OB, namely, E is reflexive (i.e., $\forall x \in OB, xEx$), symmetric (i.e., $\forall x, y \in OB, xEy \Longrightarrow yEx$), and transitive (i.e., $\forall x, y, x \in OB, xEy \wedge yEz \Longrightarrow xEz$). The equivalence relation divides the set of objects into a family of pair-wise disjoint equivalence classes. Let $[x]_E$, or simply $[x]$ when E is understood, denote the equivalence class containing x:

$$[x]_E = \{y \in OB \mid xEy\}. \tag{2}$$

It can be seen that our card sorting problem can, in fact, be modelled by an equivalence relation for each individual participant. That is, piles made by a participant can be viewed as a family of pair-wise disjoint equivalence classes. Given a subset of objects $X \subseteq OB$, we define a conditional probability by [7],

$$Pr(X|[x]) = \frac{|X \cap [x]|}{|[x]|}, \tag{3}$$

where $|\cdot|$ denotes the cardinality of a set. In the framework of three-way decisions, the condition probability may be viewed as an evaluation function about the probability that an object belongs to X given that the object belongs to $[x]$. For a pair of thresholds (α, β) with $0 \leq \beta < \alpha \leq 1$, according to Equation (1), we have probabilistic three regions of rough sets:

$$\begin{aligned} \mathrm{POS}_{(\alpha,\beta)}(X) &= \{x \in OB \mid Pr(X|[x]) \geq \alpha\}, \\ \mathrm{BND}_{(\alpha,\beta)}(X) &= \{x \in OB \mid \beta < Pr(X|[x]) < \alpha\}, \\ \mathrm{NEG}_{(\alpha,\beta)}(X) &= \{x \in OB \mid Pr(X|[x]) \leq \beta\}. \end{aligned} \tag{4}$$

That is, the three regions of rough sets are interpreted within the framework of three-way decisions.

The three regions of rough sets deals with the classification of objects based on information about the equivalence of objects. Equivalence is a special type of similarity. Our study of facial similarity uses a notion that may be considered as a generalization of equivalence relations. More specifically, we consider three levels: similar, undecidable, and dissimilar. It will be interesting to study approximations under a three-level of similarity, rather than a two-level equivalence.

3.2 Shadowed Sets as Three-Way Approximations of Fuzzy Sets

A fuzzy set, proposed by Zadeh [15], models a concept with an unsharp boundary. A fuzzy set is defined by a membership function $\mu_{\mathscr{A}} : OB \longrightarrow [0,1]$. The value $\mu_{\mathscr{A}}(x)$ is called the membership grade of x. Pedrycz [8, 9] argues that humans are typically insensitive to detailed membership grades. While we can easily comprehend membership grades close to the two end points of 0 and 1, we cannot make distinction about membership grades in the middle. For this reason, he introduces the notion of shadowed sets as three-way approximations of fuzzy sets.

In the framework of three-way decisions, the membership function $\mu_{\mathscr{A}}$ is an evaluation function. Given a pair of thresholds (α, β) with $0 \leq \beta < \alpha \leq 1$, according to Equation (1), we have a three-way approximation as follows:

$$\begin{aligned} \mathrm{CORE}_{(\alpha,\beta)}(\mu_{\mathscr{A}}) &= \{x \in OB \mid \mu_{\mathscr{A}}(x) \geq \alpha\}, \\ \mathrm{SHADOW}_{(\alpha,\beta)}(\mu_{\mathscr{A}}) &= \{x \in OB \mid \beta < \mu_{\mathscr{A}}(x) < \alpha\}, \\ \mathrm{NULL}_{(\alpha,\beta)}(\mu_{\mathscr{A}}) &= \{x \in OB \mid \mu_{\mathscr{A}}(x) \leq \beta\}. \end{aligned} \tag{5}$$

It might be pointed out the formulation is slightly different from Pedrycz's formulation. Pedrycz uses three values $0, [0,1]$, and 1 as membership grades for the three regions.

For modelling similarity, we need to consider a fuzzy relation $\mu_{\mathscr{R}} : OB \times OB \longrightarrow [0,1]$. A fuzzy similarity relation [16], as a generalization of equivalence relation, is a fuzzy relation that is reflexive (i.e., $\forall x \in OB, \mu_{calR}(x,x) = 1$, symmetric (i.e., $\forall x, y \in OB, \mu_{\mathscr{R}}(x,y) = \mu_{\mathscr{R}}(y,x)$, and max-min transitive (i.e., $\forall x, z \in OB, \mu_{\mathscr{R}}(x,z) \geq \max_{y \in OB} \min(\mu_{\mathscr{R}}(x,y), \mu_{\mathscr{R}}(y,z))$. The membership grade $\mu_{\mathscr{R}}(x,y)$ may be interpreted as the degree to which x is similar to y.

In the framework of three-way decisions, the fuzzy similarity relation $\mu_{\mathscr{R}}$ is an evaluation function on $OB \times OB$. Given a pair of thresholds (α, β) with $0 \leq \beta < \alpha \leq 1$, according to Equation (1), we have a three-way approximation of the similarity relation:

$$\begin{aligned} \mathrm{SIM}_{(\alpha,\beta)}(\mu_{\mathscr{R}}) &= \{(x,y) \in OB \times OB \mid \mu_{\mathscr{R}}(x,y) \geq \alpha\}, \\ \mathrm{UND}_{(\alpha,\beta)}(\mu_{\mathscr{R}}) &= \{(x,y) \in OB \times OB \mid \beta < \mu_{\mathscr{R}}(x,y) < \alpha\}, \\ \mathrm{DIS}_{(\alpha,\beta)}(\mu_{\mathscr{R}}) &= \{(x,y) \in OB \times OB \mid \mu_{\mathscr{R}}(x,y) \leq \beta\}, \end{aligned} \tag{6}$$

where SIM, UND, and DIS denote, respectively, the sets of similar, undecidable, and dissimilar pairs of objects. To be consistent with later discussions, we rename the three regions to better reflect their semantics.

When applying three-way decisions, it is necessary use semantically sound and meaningful notions and terms for naming and interpreting an evaluation function and the resulting three regions. In the rest of this paper, we give details about how to obtain a similarity evaluation by pooling together information from a group of raters. While each of the raters uses an equivalence relations, a synthesis of a family of equivalence relations is a similarity relation.

4 Three-Way Classification of Similarity Judgments of Facial Photographs

This section describes an application of the trisecting-and-acting model of three-way decisions in analyzing facial similarity judgments.

4.1 A Simple Three-Way Classification

Let $\mathbb{P}$ denote the set of unordered pairs of photos from a set of photos. Let N denote the number of participants. Based on the results of sorting, we can easily establish an evaluation function, $v :\longrightarrow \{0, 1, \ldots, N\}$ regarding the similarity of a pair of photographs, that is, for $p \in \mathbb{P}$,

$$v(p) = \text{the number of participants who put the pair in the same pile.} \tag{7}$$

Given a pair of thresholds (l, u) with $1 \leq l < u \leq N$, according to Equation (1), we can divide the set of pairs $\mathbb{P}$ into three pair-wise disjoint regions:

$$\begin{aligned} \text{SIM}(v) &= \{p \in \mathbb{P} \mid v(p) > u\}, \\ \text{UND}(v) &= \{p \in \mathbb{P} \mid l \leq v(p) \leq u\}, \\ \text{DIS}(v) &= \{p \in \mathbb{P} \mid v(p) < l\}. \end{aligned} \tag{8}$$

They are called, respectively, the sets of similar, undecidable, and dissimilar pairs. Alternatively, we can consider a normalized evaluation function $v_n(p) = v(p)/N$, which gives the percentage of participants who consider the pair p is similar. This provides a probabilistic interpretation of the normalized evaluation function. With such a transformation, we can apply a probabilistic approach, suggested by Yao and Cong [12], to determine the pair of thresholds (α, β) with $0 < \beta < \alpha \leq 1$.

4.2 A Refined Interpretation Based on Quality of Judgments

The simple three-way classification discussed in the previous subsection is based on the analyses reported by Hepting et al. [4], in which they did not consider the quality of judgments made by each participant. In this subsection, we look at ways to quantify the quality of the judgments by different participants. Intuitively speaking, both the number of piles and the sizes of individual piles provide hints on the quality and confidence of a participant. If a participant used more piles and, in turn, smaller sizes of individual piles, we consider the judgments to be more meaningful. Consequently, we may assign a higher weight to the participant.

Consider a pile of n photos. According to the assumption that a pair of photos in the same pile is similar, it produces $n(n-1)/2$ pairs of similar photos. Suppose a participant provided M piles with the sizes, respectively, of $n_1, \ldots n_M$. The total number of similar pairs of photos is given by:

$$N_S = \sum_{i=1}^{M} \frac{n_i(n_i-1)}{2}. \tag{9}$$

Since the total number of all possible pairs is $\frac{356*355}{2}$, the probability of judging a random pair of photos to be similar by the participant is given by:

$$P_S = \frac{\sum_{i=1}^{M} n_i(n_i-1)}{356*355}, \tag{10}$$

and the probability of judging a random pair of photos to be dissimilar is given by:

$$P_D = 1 - P_S. \tag{11}$$

Thus, we have a probability distribution $(P_S, 1-P_S)$ to model the similarity judgment of the participant. Returning to Figure 2, having fewer than 16 dissimilar votes for a pair (placing it either in the similar or undecidable groups) is highly unlikely.

The intuition is that the smaller the probability, the greater the confidence of the participant in that judgment. For most pairs of photos, some participants have rated them similar and some have rated them dissimilar. There are no photo pairs that were rated similar by all participants, but there are some photo pairs that were rated as dissimilar by all participants. Based on the probabilities calculated: for the real data there are 0 all-similar pairs and 232 all-dissimilar pairs expected and for the simulated data there are likewise 0 all-similar pairs and 11490 all-dissimilar pairs expected.

5 Three-Way Analysis of Human Similarity

Based on the proposed model, we report results from analyzing a dataset obtained from card sorting.

5.1 Facial Similarity Judgments through Card Sorting

We briefly review a procedure used to obtain similarity judgments on a set of facial photographs through a technique known as card sorting. The details have been reported elsewhere [4].

There were 25 participants who judged the similarity of a set of facial photographs. Each photograph was presented on a separate card. Each participant was given a randomly-ordered stack of 356 facial photographs and asked to sort the facial photographs into an unrestricted number of piles based on perceived similarity.

It was explained to the participants that photographs in the same pile are considered to be similar and photographs in different piles are considered to be dissimilar. Figure 1(a) shows the participant behaviours in the card sorting study.

The total of 63,190 pairs from the 356 cards is a very large number. It was impossible to ask a participant to exhaustively consider all pairs. Instead, the following procedure was used so that a participant made direct judgments on a small fraction of all possible pairs. Each participant drew a single photo successively from the stack of photos. Once a photo was placed in a pile, it could not be moved. When a new photo was drawn from the stack, a participant only compared the newly-drawn photo with the very top photo on each existing pile. The new photo could be placed on an existing pile, or a new pile could be created.

To show the possible utility of the judgments from the described procedure, we observe the diversity of behaviours from the 25 participants by comparing it with the randomly-generated judgments. For this purpose, a set of randomly-generated data for 25 hypothetical participants was created, which was generated according to the code in Table 1. Figure 1(b) presents the randomly-simulated participants.

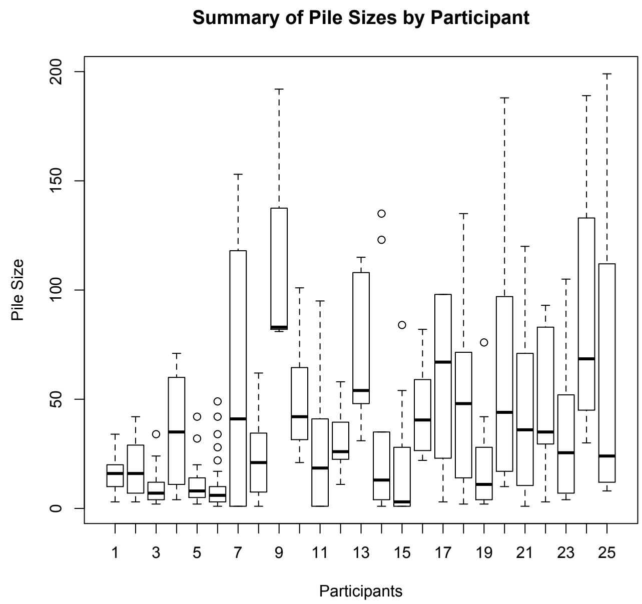
In terms of number of piles, the 25 participants produced between 3 to 38 piles, which indicates a large variation. It can be observed that the participant judgments in terms of sizes of different piles are significantly different from those in the randomly-generated data. This suggests that the restricted procedure does generate useful human similarity judgments. We hypothesize that the variability in the number of piles (between 3 and 38) and the pile size (1 and 199) reflects some variability in the confidence of the participants' judgments. The interpretation that some participants judge similarity "correctly" and others judge it "incorrectly" cannot be applied here because there is no objective standard against which each participant's ratings can be judged.

5.2 Three-Way Analysis Based on the Simple Model

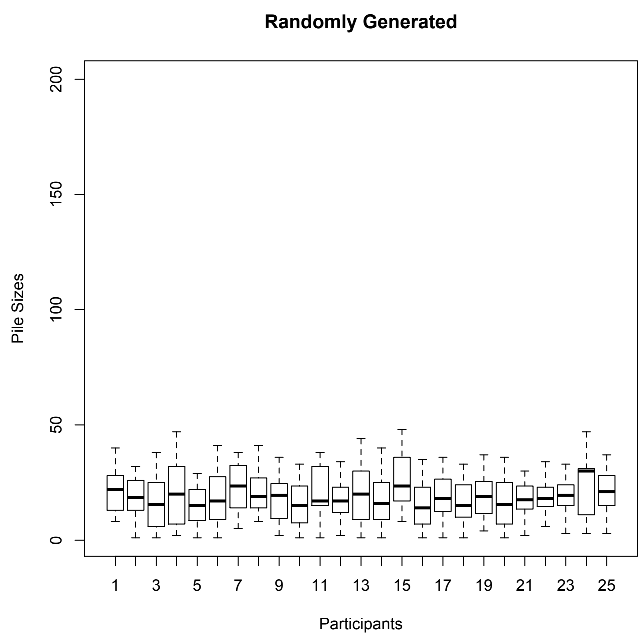
For the dataset used in this work, we have $N = 25$. We set $l = 10$ and $u = 15$. Specifically, we consider a pair of photographs to be similar if more than 15 participants out of 25 put them in the same pile, or equivalently, more than $15/25 = 60\%$ participants put them in the same pile. We consider a pair of photographs to be dissimilar if less than 10 participants out of 25 put them in the same pile, or equivalently, less than $10/25 = 40\%$ participants put them in the same pile. Otherwise, we view that the judgments of the 25 participants are inconclusive to declare similarity or dissimilarity of the pair of photos.

Figure 2 shows the effects of these thresholds on the real and random data. Based on the pair of thresholds $l = 10$ and $u = 15$, we have similar pairs, undecidable pairs, and dissimilar pairs. Table 2 summarizes the numbers of pairs in each region, between the observed and randomly-simulated data.

Figure 3 shows two samples of Similar pairs (S1 and S2 refer to the left and right pairs, respectively). For both S1 and S2, 19 participants put the pair into the same



(a)



(b)

Fig. 1: A summary of pile sizes by participant: (a) real data from card sorting study and (b) randomly-simulated data.

Table 1: Code, written in the python language, to generate piles of photos to simulate participants behaving randomly.

```
# assign photos randomly to piles
import sys, itertools, random
# dictionary of photos
from photos_dict import photos
# seed random number generator
random.seed()
# get the list of photo labels
photonames = list(photos.keys())

# for each participant (p)
for p in range(25):
  # initialize dictionary
  randpiles = {}
  # start with 0 piles
  pilecount = 0
  # randomly shuffle the photo names
  random.shuffle(photonames)
  # for each photo (ph)
  for ph in range(356):
    # choose a pile for photo, at random
    cp = int(round(random.random() * pilecount))
    # append photo to chosen pile (initialize if needed)
    if cp not in randpiles:
      randpiles[cp] = []
      (randpiles[cp]).append(photonames[ph])
      pilecount += 1
    else:
      (randpiles[cp]).append(photonames[ph])
  # write out the simulated data into a separate file
  with open('rand/'+str(p+1).zfill(2)+'.txt','w') as outf:
    for rk in sorted(randpiles.keys()):
      # concatenate same-pile photo names for output
      ostr = ""
      for pl in range(len(randpiles[rk])-1):
        ostr += str((randpiles[rk])[pl]) + " "
      ostr += str((randpiles[rk])[len(randpiles[rk])-1])
      ostr += "\n"
      outf.write(ostr)
```

Table 2: Number of pairs in each region.

Region	Real	Random
Similar (SIM)	125	0
Undecidable (UND)	6,416	0
Dissimilar (DIS)	56,649	63,190

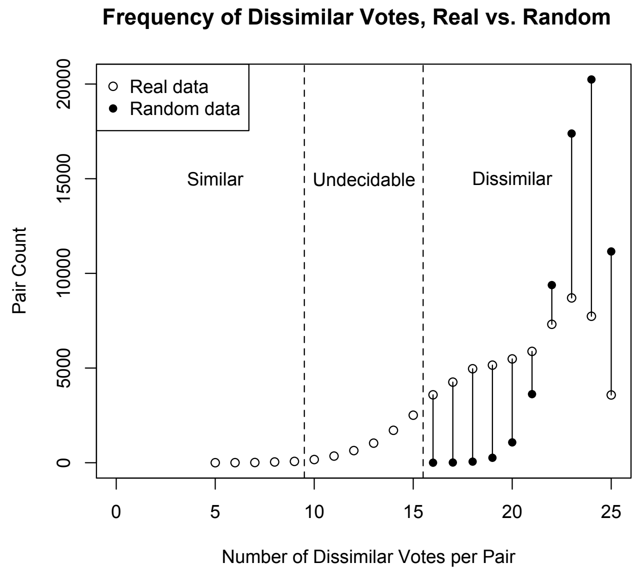


Fig. 2: A summary of ratings by participant from real data from card sorting study and randomly-simulated data.

pile. Figure 4 shows two samples of Undecidable pairs (U1 and U2 refer to the left and right pairs, respectively). For both U1 and U2, 13 participants put the pair into the same pile. Figure 5 shows two samples of Dissimilar pairs (D1 and D2 refer to the left and right pairs, respectively). For D1, 4 participants put the pair into the same pile and for D2, only 2 participants put the pair into the same pile.

Fig. 3: The 2 pairs of photos shown here (S1 left, S2 right) represent samples from the similar (SIM) region.

Fig. 4: The 2 pairs of photos shown here (U1 left, U2 right) represent samples from the undecidable (UND) region. Pairs U1 and U2 were highlighted in the study by Hepting and Almestadi [3].

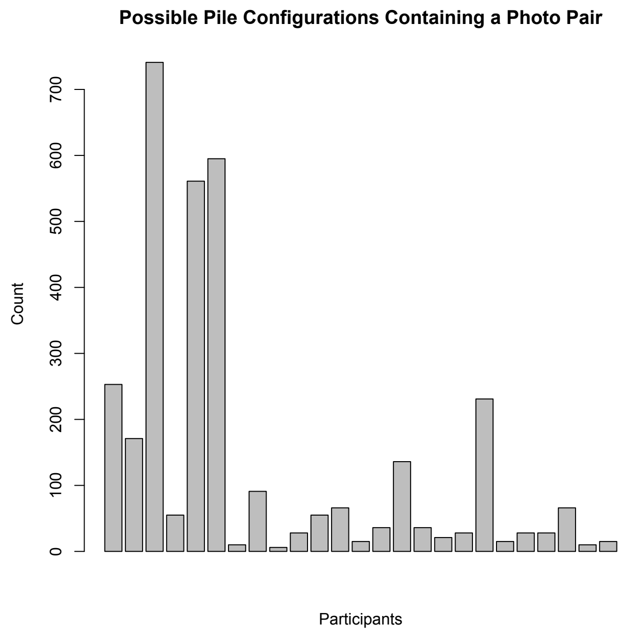
Fig. 5: The 2 pairs of photos shown here (D1 left, D2 right) represent samples from the dissimilar (DIS) region.

An inspection of the final three-way classification confirms that pairs in the similar set are indeed similar, pairs in the dissimilar set are very different, and pairs in the undecidable set share some common features while differing in some other aspects.

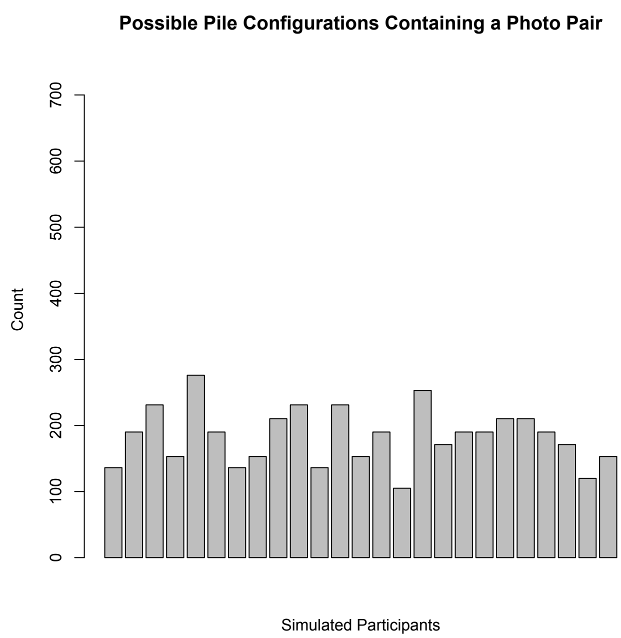
5.3 Three-Way Analysis Based on the Refined Model

A more refined approach is possible by looking at the number of photos that are considered along with the photos in any particular pair. If a participant made M piles, the number of possible configurations for the participant is $M + \binom{M}{2}$. Figure 6 compares the variability in observed participant data (min=6, max=741) with that of simulated participants (min=105, max=276). These plots summarize the number of possible pile configurations that may contain a particular photo pair, by participant. Higher numbers of possible configurations correspond to more piles of smaller size.

Figure 7 summarizes the number of photos in the piles that contain each of the photo pairs in Figures 3 - 5. When the pair is judged to be dissimilar (N) by a



(a)



(b)

Fig. 6: Number of possible pile configurations that may contain a particular photo pair, by participant. Real participants on the left and simulated participants on the right.

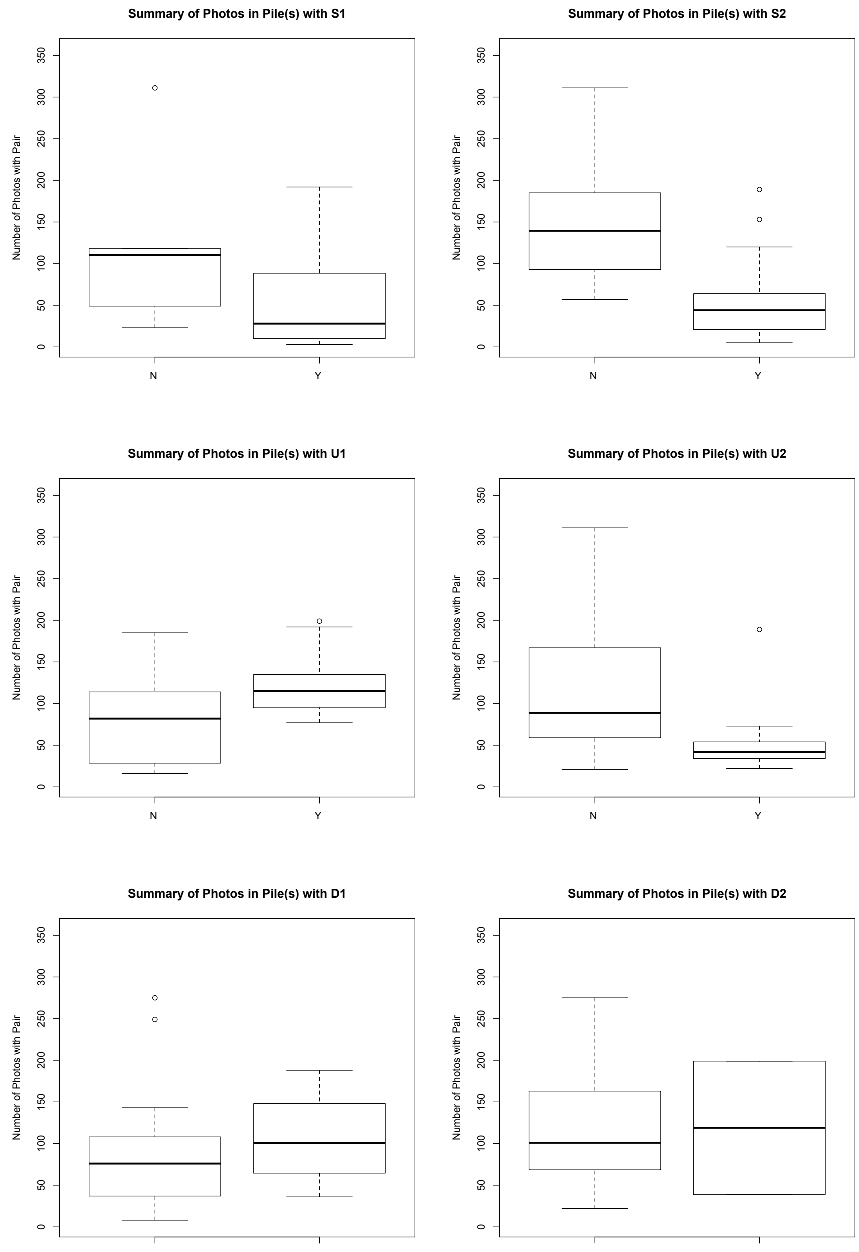


Fig. 7: Summary of pile configuration sizes for the sample pairs (see Figures 3 - 5). The bold lines indicate the median sizes.

participant, the number of photos associated with the pair is the sum of the sizes of the 2 different piles that each contain one of the photos in the pair. When the pair is judged to be similar (Y) by a participant, the number of photos associated with the pair is size of the single pile that contains both photos.

Figure 8 summarizes by relative rank the number of photos associated with each of the photo pairs in Figures 3 - 5. Regardless of the number of possible pile configurations that may contain the pair of interest, the smallest of these configurations has a relative rank approaching 0 and the largest of these configurations has a relative rank of 1. The relative rank can be transformed into a similarity score according to Equation 12.

$$S_r(A,B) = \begin{cases} \frac{2-\text{relative rank}(P_{AB})}{2}, & A \text{ and } B \text{ are in the same pile,} \\ \frac{\text{relative rank}(P_A+P_B)}{2}, & A \text{ and } B \text{ are in two different piles.} \end{cases} \tag{12}$$

This score is computed for each rating of each pair of photos. From the card sorting study, 63,190 scores can be computed for each of the 25 participants. As an example, Participant 21 made 7 piles of photos with sizes: 2, 19, 36, 36, 56, 86, and 120 (355 photos rated). This leads to 28 configurations of piles, some with the same size. Please see Table 3 for details of the calculations and Figure 9 for a plot of the results. In order to create a single similarity score for a pair of photos, we sum the score from each rating and divide by the number of raters (N), according to Equation 13.

$$S(A,B) = \frac{1}{N}\sum_{r=1}^{N} S_r(A,B). \tag{13}$$

Figure 10 summarizes the similarity scores, sorted into increasing order, for each rating of each sample pair. The scores are determined by the relative rank of the configuration that contains the pair. Similarity scores for pairs rated as dissimilar (not placed in the same pile) will be in the range $(0, 0.5]$ and scores for pairs rated as similar (placed in the same pile) will be in the range $(0.5, 1.0)$. A score near 0 occurs when the photo pair is rated as dissimilar, but the combined size of the piles containing the photos is very small. A score near 1 occurs when the photo pair is rated as similar and the size of that pile is the very small. The similarity scores of the sample pairs are, for S1: 0.7377; for S2: 0.7230; for U1: 0.5607; for U2: 0.5742; for D1: 0.4015; and for D2: 0.4421. In Section 5.2, we began with $\alpha_0 = 0.6$ and $\beta_0 = 0.4$. We notice that S1, S2, U1, and U2 remain in their original regions. However D1 and D2 are now both in region UND. Let us examine the selection of α and β more closely.

Figure 11 considers all similarity scores from all ratings of photopairs. The boxplot summarizes 1,267,785 dissimilar (N) ratings and 304,186 similar (Y) ratings. From this analysis, we chose two sets of thresholds.

- $\alpha_1 = 0.7000$ (median score for pairs in same pile), $\beta_1 = 0.4367$ (median score for pairs in different piles). The application of this threshold set is illustrated in Figure 13.

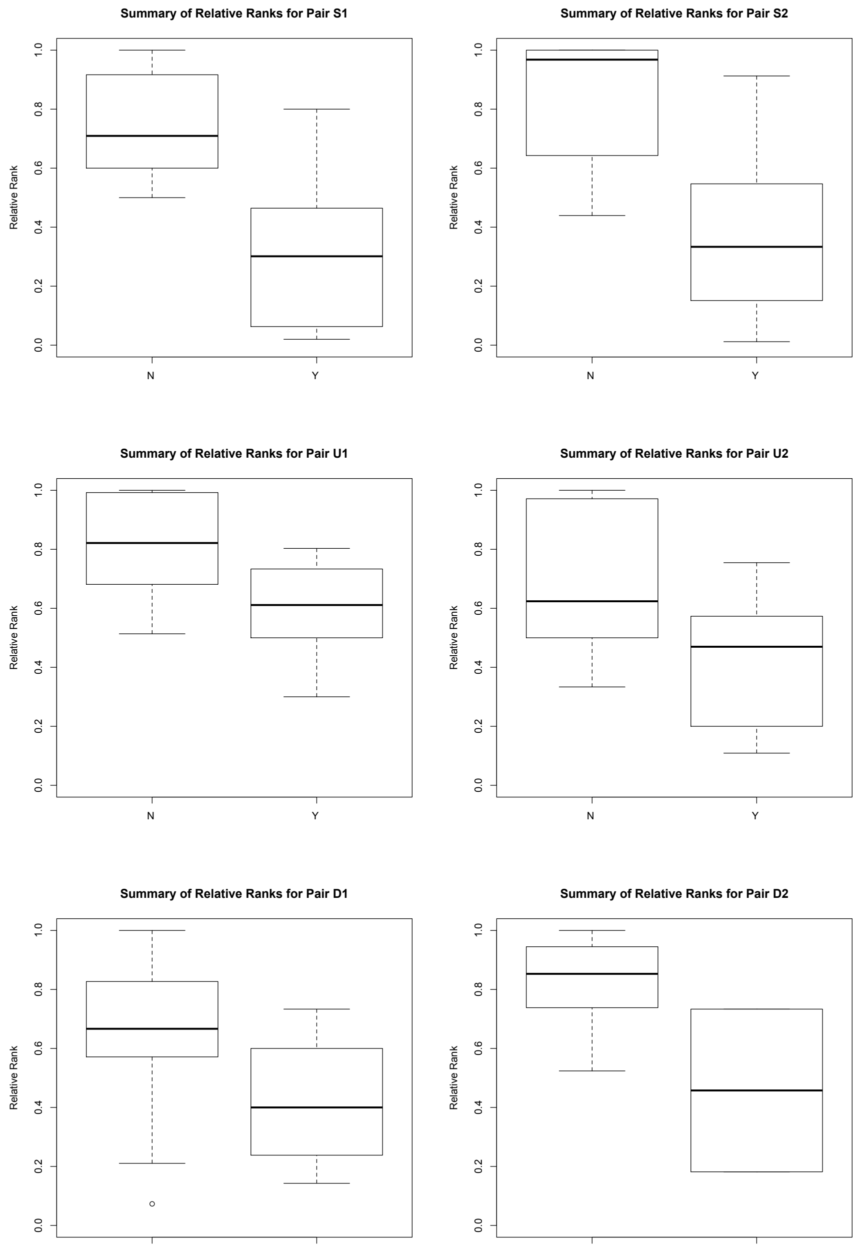


Fig. 8: Summary of relative ranks for the sample pairs (see Figures 3 - 5). The bold lines indicate the median relative ranks.

Table 3: Calculations from Equation 12 carried out for Participant 21, who made 7 piles of photos with sizes: 2, 19, 36, 36, 56, 86, and 120. Similar ratings are indicated by bold type.

Size	Rank	Relative Rank	Similarity Score
2	**1**	**0.0357**	**0.9821**
19	**2**	**0.0714**	**0.9643**
$2+19=21$	3	0.1071	0.0536
36	**4**	**0.1429**	**0.9286**
$2+36=38$	6	0.2143	0.1071
$19+36=55$	8	0.2857	0.1429
56	**10**	**0.3571**	**0.8214**
$2+56=58$	11	0.3929	0.1964
$36+36=72$	12	0.4286	0.2143
$19+56=75$	13	0.4642	0.2321
86	**14**	**0.5000**	**0.7500**
$2+86=88$	15	0.5357	0.2679
$36+56=92$	16	0.5714	0.2857
$19+86=105$	18	0.6429	0.3214
120	**19**	**0.6786**	**0.6607**
$2+120=122$ $36+86=122$	20	0.7143	0.3571
$19+120=139$	23	0.8214	0.4107
$56+86=142$	24	0.8571	0.4286
$36+120=156$	25	0.8929	0.4464
$56+120=176$	27	0.9643	0.4821
$86+120=206$	28	1.0000	0.5000

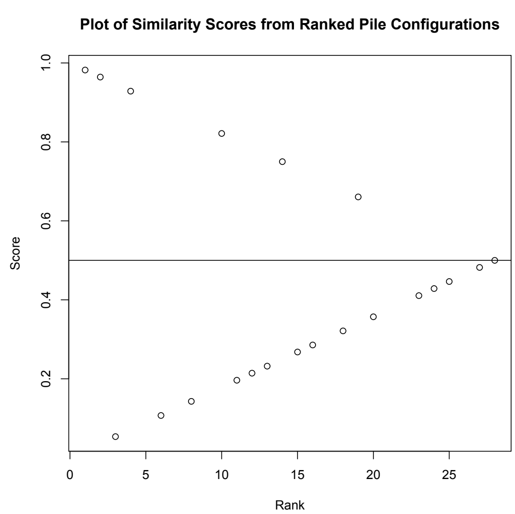


Fig. 9: Plot of similarity scores from rank of pile configurations for Participant 21. See Table 3 for the calculations.

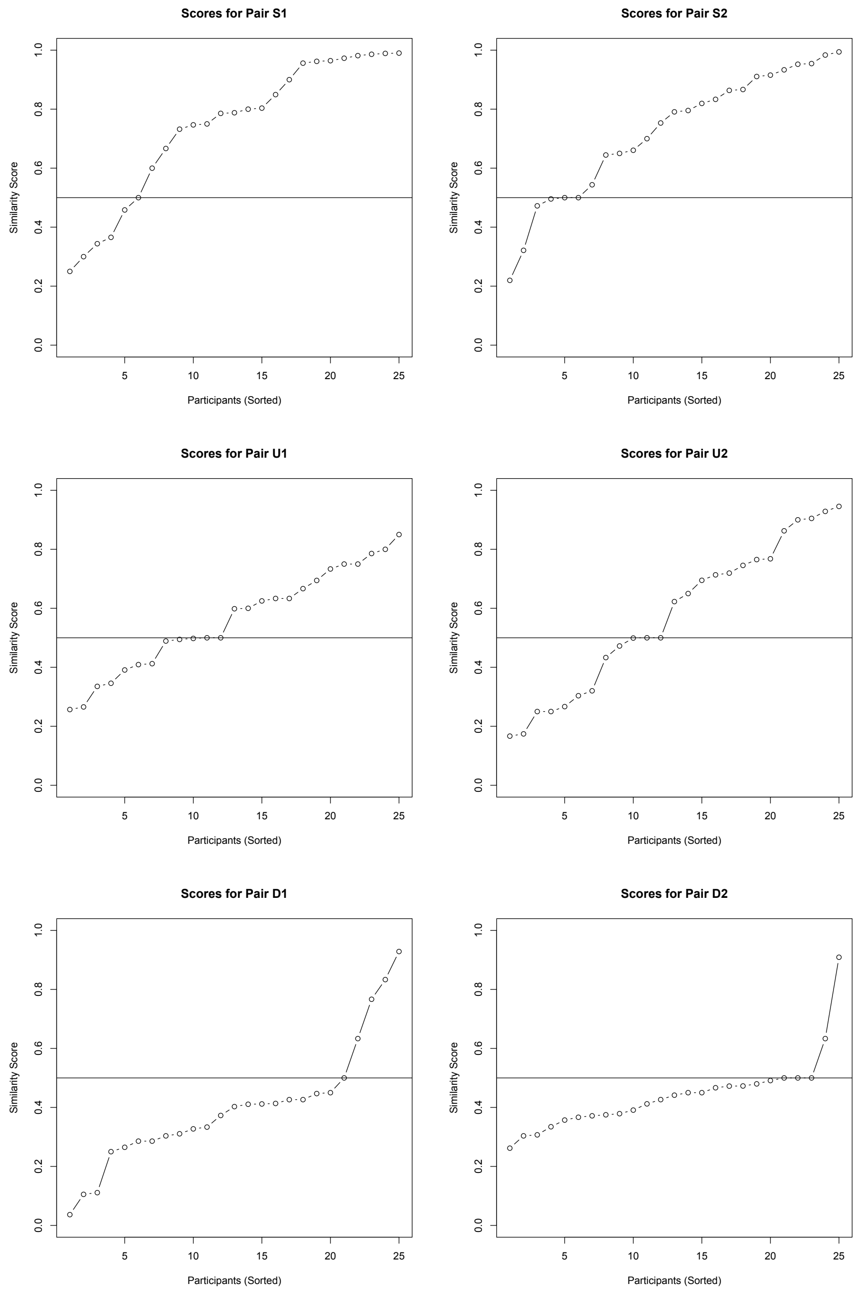


Fig. 10: Summary of similarity scores, sorted into ascending order, for each rating of the sample pairs (see Figures 3 - 5).

- $\alpha_2 = 0.6389$ (25th percentile of scores for pairs in same pile), $\beta_2 = 0.4824$ (75th percentile of scores for pairs in different piles). The application of this threshold set is illustrated in Figure 14.

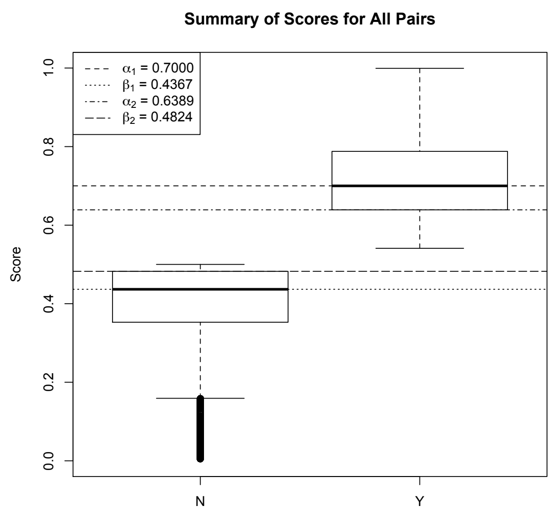


Fig. 11: Summary of similarity scores for Dissimilar (N) and Similar (Y) ratings for all 63,190 pairs, computed according to Equation 13. Two pairs of thresholds, (α_1, β_1) and (α_2, β_2), are also indicated.

Table 4: Number of pairs classified for different threshold pairs. The first line of data is repeated from Table 2.

Thresholds	Dissimilar (DIS)	Undecidable (UND)	Similar (SIM)
$(u = 15, l = 10)$	56,649	6,416	125
$(\alpha_0 = 0.6000, \beta_0 = 0.4000)$	2,782	60,018	390
$(\alpha_1 = 0.7000, \beta_1 = 0.4367)$	16,472	46,714	4
$(\alpha_2 = 0.6389, \beta_2 = 0.4824)$	43,469	19,649	72

In Figure 12, the trilinear plot summarizes an exploration for values of α and β. Each plotted point represents the fraction of pairs in the DIS, UND, and SIM regions by different choices for α and β. Points at a vertex indicate 100% of the pairs are assigned to the region indicated by vertex label. In this Figure, each point represents the assignment of all 63,190 pairs to the 3 regions. It is also possible to consider the assignment of a pair's individual ratings to those regions and obtain more finely-grained information about the pair's similarity. Figures 13 and 14 illustrate the assignment of individual ratings amongst the DIS, UND, and SIM regions.

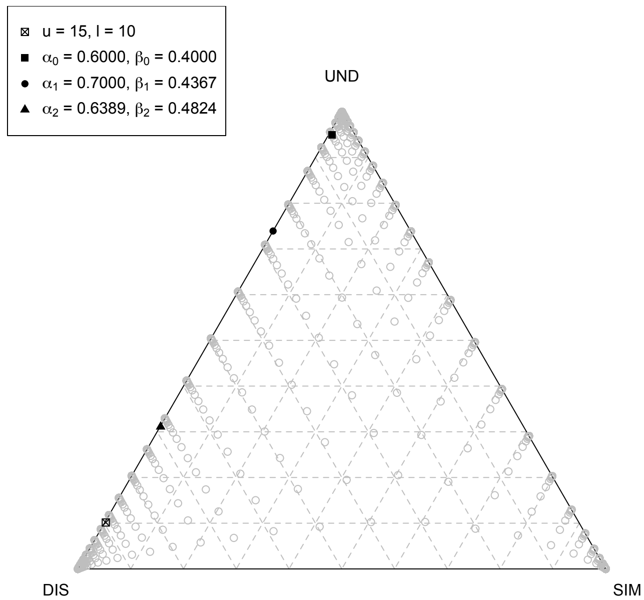


Fig. 12: This trilinear plot summarizes an exploration for values of α and β taken from [0,1] at increments of 0.01 such that $\alpha > \beta$. Each point plotted in grey represents a choice of α and β. Plotted in black are the points corresponding to Table 4.

6 Conclusions and Future Work

This work presents a three-way classification of human judgments of similarity. The agreement of a set of participants leads to both a set of similar pairs and a set of dissimilar pairs. Their disagreement leads to undecidable pairs. Findings from this study may find practical applications. For example, the selected photo pairs (Figures 3 - 5) may provide a firm foundation for the development of understanding of the processes or strategies that different people use to judge facial similarity. We anticipate that it may be possible to use the correct identification of strategy to create presentations of photos that would allow eyewitness identification to have improved accuracy and utility.

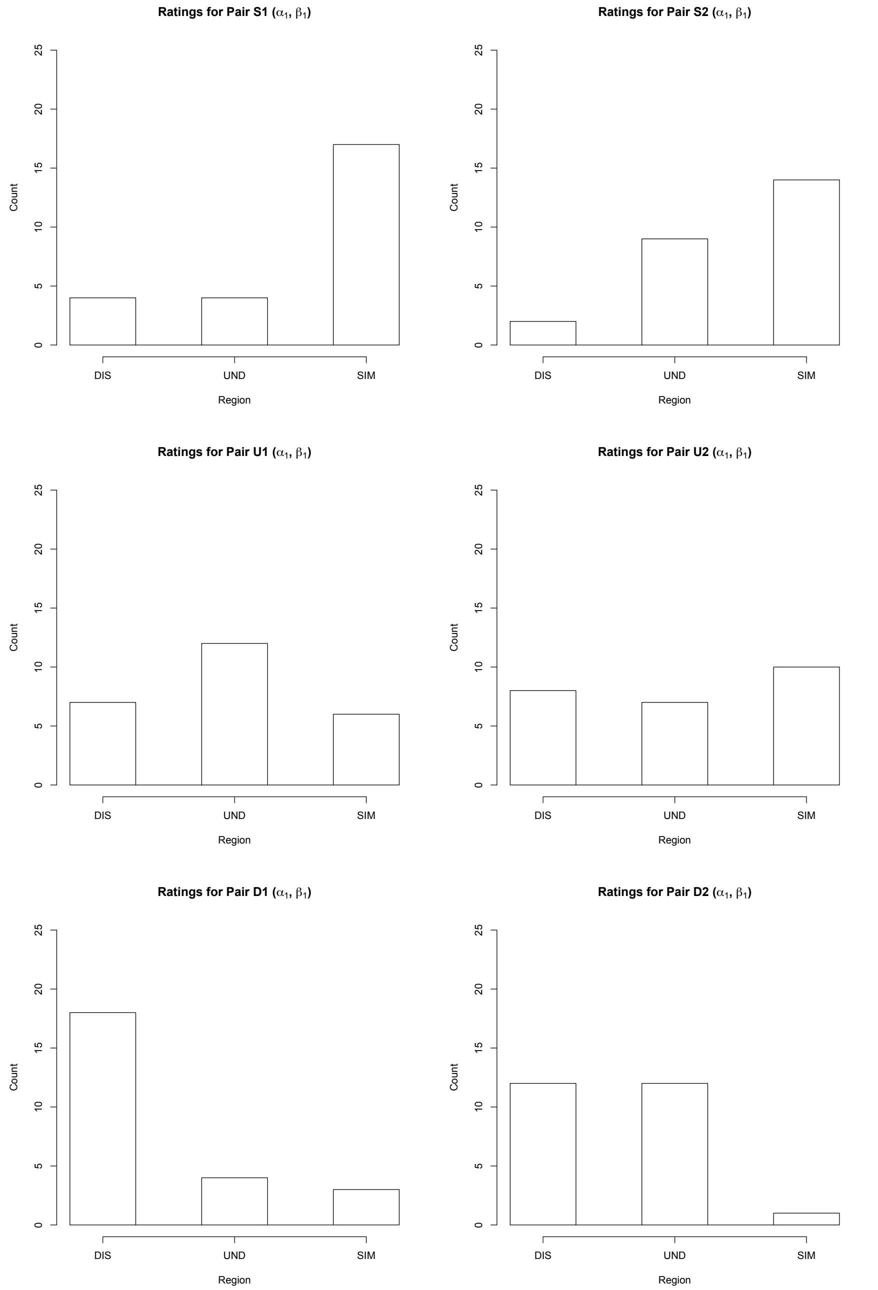


Fig. 13: Classification as one of Dissimilar, Undecidable, or Similar. These decisions are based on thresholds $\alpha_1 = 0.7000$ and $\beta_1 = 0.4367$.

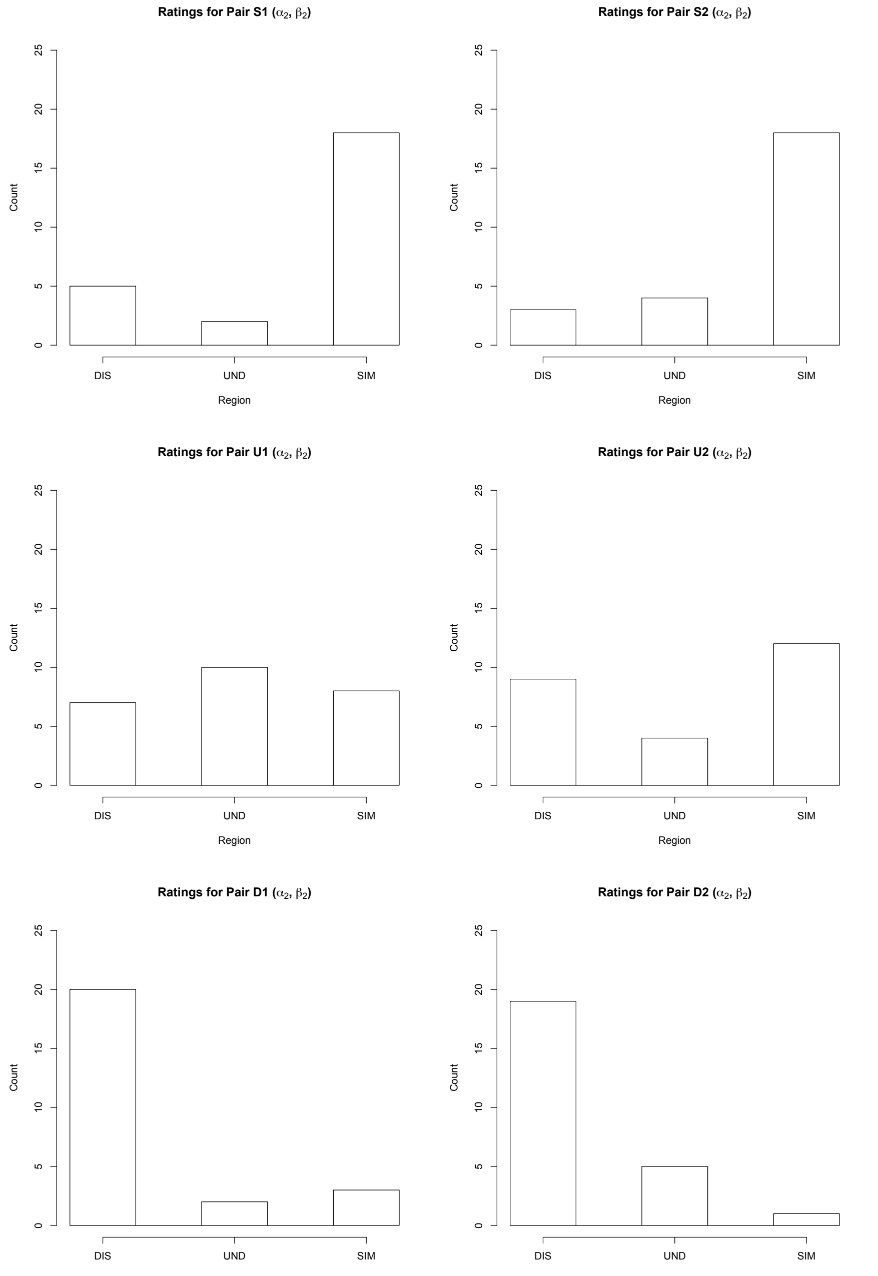


Fig. 14: Classification as one of Dissimilar, Undecidable, or Similar. These decisions are based on thresholds $\alpha_2 = 0.6389$ and $\beta_2 = 0.4824$.

As future work, a three-way classification suggests two types of investigation. By studying each class of pairs, we try to identify features that are useful in arriving at a judgment of similarity or dissimilarity. By comparing pairs of classes, for example, the class of similar pairs and the class of dissimilar pairs, we try to identify features that enable the participants to differentiate the two classes. It will also be of interest to define quantitative measures to precisely describe our initial observations.

Acknowledgements The authors thank the editors, Rafael Bello, Rafael Falcon, and José Luis Verdegay, for their encouragement and the anonymous reviewers for their constructive comments. This work has been supported, in part, by two NSERC Discovery Grants.

References

1. Gao, C., Yao, Y.: Actionable strategies in three-way decisions. Knowledge-Based Systems (2017). DOI 10.1016/j.knosys.2017.07.001
2. Grant, D.A., Berg, E.: A behavioral analysis of degree of reinforcement and ease of shifting to new responses in Weigl-type card-sorting problem. Journal of Experimental Psychology **38**(4), 404–411 (1948). DOI 10.1037/h0059831
3. Hepting, D.H., Almestadi, E.H.: Discernibility in the analysis of binary card sort data. In: D. Ciucci, M. Inuiguchi, Y. Yao, D. Ślęzak, G. Wang (eds.) Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing. RSFDGrC 2013, *Lecture Notes in Computer Science*, vol. 8170, pp. 380–387. Springer, Berlin (2013). DOI 10.1007/978-3-642-41218-9_41
4. Hepting, D.H., Spring, R., Ślęzak, D.: A rough set exploration of facial similarity judgements. In: J.F. Peters, A. Skowron, S. Hiroshi, M.K. Chakraborty, D. Ślęzak, A.E. Hassanien, W. Zhu (eds.) Transactions on Rough Sets XIV, *Lecture Notes in Computer Science*, vol. 6600, pp. 81–99. Springer (2011). DOI 10.1007/978-3-642-21563-6_5
5. Kelly, J.: Computing, cognition and the future of knowing. Whitepaper, IBM Reseach (2015)
6. Marr, D.: Vision: A Computational Investigation into the Human Representation and Processing of Visual Information. W.H. Freeman and Company, New York (1982)
7. Pawlak, Z., Wong, S.K.M., Ziarko, W.: Rough sets: Probabilistic versus deterministic approach. Int. J. Man-Mach. Stud. **29**(1), 81–95 (1988). DOI 10.1016/S0020-7373(88)80032-4
8. Pedrycz, W.: Shadowed sets: Representing and processing fuzzy sets. Trans. Sys. Man Cyber. Part B **28**(1), 103–109 (1998). DOI 10.1109/3477.658584
9. Pedrycz, W.: Rough-Fuzzy Hybridization: A New Trend in Decision Making, 1st edn., chap. Shadowed Sets: Bridging Fuzzy and Rough Sets, pp. 179–199. Springer-Verlag New York, Inc., New York (1999)
10. Yao, Y.: An Outline of a Theory of Three-Way Decisions, pp. 1–17. Springer Berlin Heidelberg, Berlin, Heidelberg (2012). DOI 10.1007/978-3-642-32115-3_1
11. Yao, Y.: Three-way decisions and cognitive computing. Cognitive Computation **8**(4), 543–554 (2016). DOI 10.1007/s12559-016-9397-5
12. Yao, Y., Gao, C.: Statistical Interpretations of Three-Way Decisions, pp. 309–320. Springer International Publishing, Cham (2015). DOI 10.1007/978-3-319-25754-9_28
13. Yao, Y., Greco, S., Słowiński, R.: Probabilistic Rough Sets, pp. 387–411. Springer Berlin Heidelberg, Berlin, Heidelberg (2015). DOI 10.1007/978-3-662-43505-2_24
14. Yao, Y.Y.: Probabilistic approaches to rough sets. Expert Systems **20**(5), 287–297 (2003). DOI 10.1111/1468-0394.00253
15. Zadeh, L.A.: Fuzzy sets. Information and Control **8**(3), 338 – 353 (1965). DOI 10.1016/S0019-9958(65)90241-X
16. Zadeh, L.A.: Similarity relations and fuzzy orderings. Information Sciences **3**(2), 177 – 200 (1971). DOI 10.1016/S0020-0255(71)80005-1