# Three-Way Analysis of Facial Similarity Judgments

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

University of Regina, Regina, SK, S4S 0A2, CANADA
hepting@cs.uregina.ca, binamerh@cs.uregina.ca, yyao@cs.uregina.ca

**Abstract.** The card sorting problem involves the similarity judgments of pairs of photos, taken from a set of photos, by a group of participants. 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. In this paper, we present a framework for three-way analysis of judgments of 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 by at least 60% of participants as similar; a set of dissimilar pairs that are judged by at least 60% of participants 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 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.

**Keywords:** facial photograph, card sorting, three-way decision, probability

## 1 Introduction

In this paper, we focus on card sorting applied to a set of facial photographs. A fundamental task in the sorting of facial photographs is the modelling of card similarity. 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 [6,7], the main objective of this paper is to suggest a framework of three-way analysis of human judgments of similarity.

In a nutshell, the basic ideas of three-way decisions are thinking and problem-solving about threes [7], 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. To grasp the idea of thinking in threes, let us consider three examples. Marr [5] 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 [4] 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 toidentify 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. By extracting the basic components of trisection and action from these examples, Yao [7] 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.

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 paper, we report our preliminary results from applying the theory and performing experiments with three-way analysis of similarity based on human judgments.

## 2 A Trisecting-and-Acting Model of Three-way Decisions

An evaluation-based trisecting-and-acting model of three-way decisions consists of two components [6,7]. 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 $(\alpha, \beta)$, $\alpha, \beta \in \Re$, 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:

$$
\begin{array}{ll}
\text{(i)} & \mathrm{R}_l(v) \cap \mathrm{R}_m(v) = \mathrm{R}_l(v) \cap \mathrm{R}_h(v) = \mathrm{R}_m(v) \cap \mathrm{R}_h(v) = \emptyset, \\
\text{(ii)} & \mathrm{R}_l(v) \cup \mathrm{R}_m(v) \cup \mathrm{R}_h(v) = OB.
\end{array}
$$

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 $(\alpha, \beta)$. 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 [6]. 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 $(\alpha, \beta)$ 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., $income \leq \$35\text{k}$), middle income (i.e., $\$35\text{k} < income < \$120\text{k}$), and high income (i.e., $income \geq \$120\text{k}$), respectively. For the three levels of income, one typically devise different formulas or rates to compute tax.

## 3 Three-way Classification of Similarity Judgments of Facial Photographs

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

### 3.1 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 [3].

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.

For 356 cards, there are a total of 63,190 pairs, which 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 each 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.

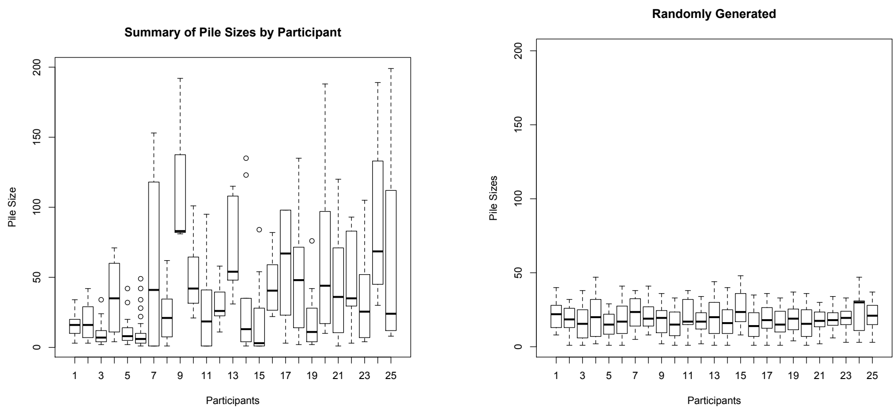
To show the possible utility of the judgments from the described procedure, we observe the diversity of behaviours from the 25 participants by comparing it with the randomly-generated judgments. For this purpose, a set of randomly-generated data for 25 hypothetical participants was created, which was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

In terms of number of piles, the 25 participants produced between 3 to 38 piles, which indicates a large variation. Figure 1(a) shows the participant behaviours in the card sorting study and Figure 1(b) gives the corresponding behaviours of the randomly-simulated participants. 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.

```
# 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 participant 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 1: Code to generate piles of photos to simulate participants behaving randomly.




(a) (b)

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

### 3.2 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 $p \in \mathbb{P}$,

$$v(p) = \text{the number of participants who put the pair in the same pile.} \tag{2}$$

Given a pair of threshold $(l, u)$ 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:

$$\begin{aligned} \mathrm{SIM}(v) &= \{p \in \mathbb{P} \mid v(p) > u\}, \\ \mathrm{UND}(v) &= \{p \in \mathbb{P} \mid l \leq v(p) \leq u\}, \\ \mathrm{DIS}(v) &= \{p \in \mathbb{P} \mid v(p) < l\}. \end{aligned} \tag{3}$$

They are called, respectively, the sets of similar, undecided, 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 [8], to determine pair of thresholds $(\alpha, \beta)$ with $0 < \beta < \alpha \leq 1$.

For the dataset used in this paper, 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.

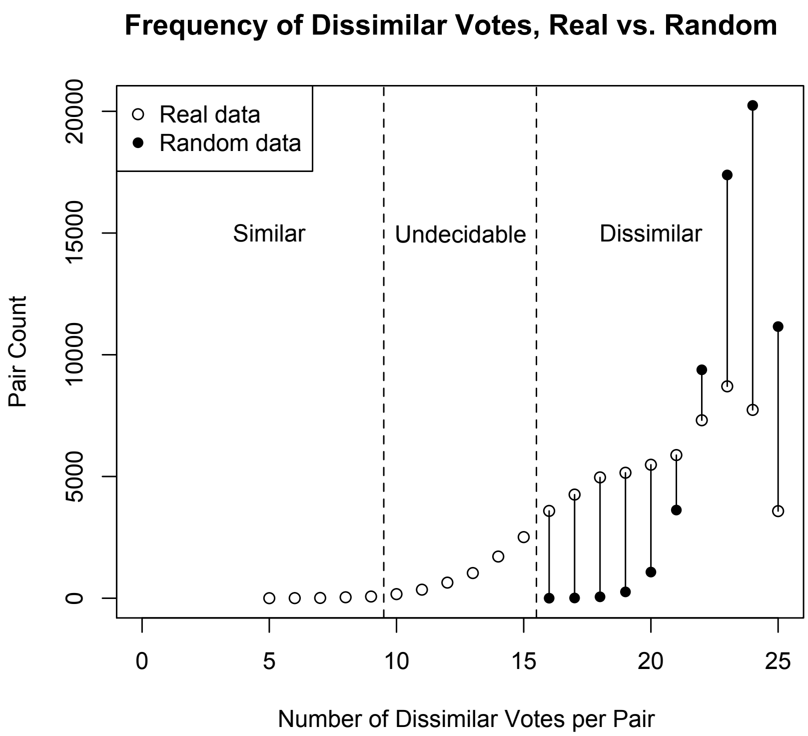



Fig. 2: A summary of pile sizes by participant: (a) real data from card sorting study and (b) randomly-simulated data. Notice that all pairs from the randomly-simulated data are classified as 'dissimilar'.

Based on the pair of thresholds $l = 10$ and $u = 15$, we have similar pairs, undecidable pairs, and dissimilar pairs. Figure 3 gives a few samples of similar, undecidable, and dissimilar pairs. 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 differ in some other aspects. As future work, it will be of

Table 2: Number of pairs in each group, observed from the real data and simulated by the random data.

| *Group* | Real | Random |
|---|---|---|
| Similar | 125 | 0 |
| Undecidable | 6,416 | 0 |
| Dissimilar | 56,649 | 63,190 |

interest to define quantitative measures to precisely describe our initial observations.

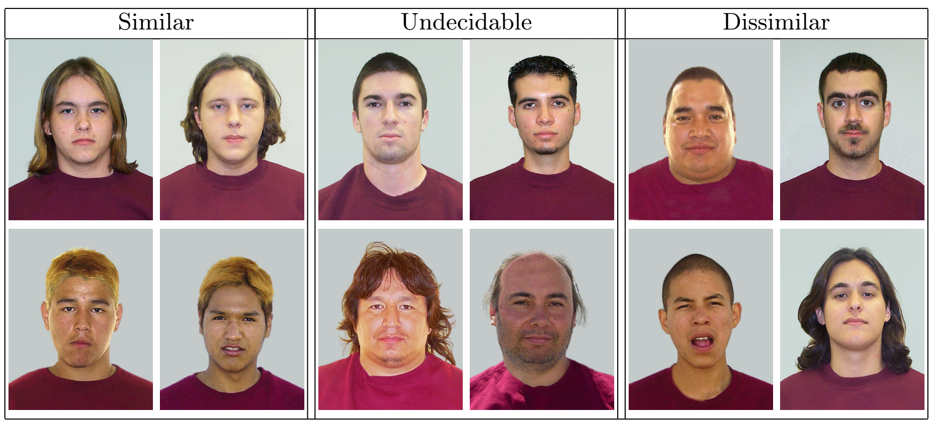



Fig. 3: The 6 pairs of photos shown here represent 2 samples from each of the similar, undecidable, and dissimilar groups.

### 3.3 A More 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. [3], 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 number of 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:

$$N_S = \sum_{i=1}^{M} \frac{n_i(n_i - 1)}{2}. \tag{4}$$

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:

$$P_S = \frac{\sum_{i=1}^{M} n_i(n_i - 1)}{356 * 355}, \tag{5}$$

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

$$P_D = 1 - P_S. \tag{6}$$

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.

## 4 Conclusions and Future Work

This paper 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 undecided pairs. Finding from this study may find practical applications. For example, the selected photo pairs (Figure 3) 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.

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.

## References


1. Gao, C., Yao, Y.: Actionable strategies in three-way decisions. Knowledge-Based Systems (2017)
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)
3. Hepting, D.H., Spring, R., Ślęzak, D.: A rough set exploration of facial similarity judgements. In: Peters, J.F., Skowron, A., Hiroshi, S., Chakraborty, M.K., Ślęzak, D., Hassanien, A.E., Zhu, W. (eds.) Transactions on Rough Sets XIV, Lecture Notes in Computer Science, vol. 6600, pp. 81–99. Springer (2011)
4. Kelly, J.: Computing, cognition and the future of knowing. Whitepaper, IBM Reseach (2015)
5. Marr, D.: Vision: A Computational Investigation into the Human Representation and Processing of Visual Information. W.H. Freeman and Company, New York (1982)
6. Yao, Y.: An Outline of a Theory of Three-Way Decisions, pp. 1–17. Springer Berlin Heidelberg, Berlin, Heidelberg (2012), https://doi.org/10.1007/978-3-642-32115-3_1
7. Yao, Y.: Three-way decisions and cognitive computing. Cognitive Computation 8(4), 543–554 (Aug 2016), https://doi.org/10.1007/s12559-016-9397-5
8. Yao, Y., Gao, C.: Statistical Interpretations of Three-Way Decisions, pp. 309–320. Springer International Publishing, Cham (2015), https://doi.org/10.1007/978-3-319-25754-9_28