Is Perceptual Similarity the underlying mechanism of categorical perception?

Tanja de Boer, Ignace Hooge and Mijke O. Hartendorp

Masterthesis Applied Cognitive Psychology
Tanja de Boer, mail@tanjadeboer.nl

Supervised by:
Mijke Hartendorp MSc, Utrecht University
Division of Experimental Psychology
Willem C. van Unnik Building
Heidelberglaan 2
3584 CS, Utrecht
The Netherlands

Dr. S.F. Donker, Utrecht University
Division of Experimental Psychology
Willem C. van Unnik Building
Heidelberglaan 2
3584 CS, Utrecht
The Netherlands

Keywords: Categorical Perception, Morphed Objects, Perceptual Similarity, Object recognition, Object Skeleton, Eye-Tracking, Area Of Interest.

Abstract

Based on previous research by Hartendorp, Van der Stigchel, Burnett, Jellema, Eilers & Postma (2010), morphed objects are perceived as their dominant category, which is known as categorical perception (CP). Subsequently, they found higher perceptual similarity ratings for CP morph series compared to non-CP series. In the current study, we try to predict CP from similarity ratings. To do so, in addition to the existing Hartendorp dataset, a new stimulus-dataset is developed and tested to be able to generalize the CP/similarity effect. In experiment 1, we conducted a verification task, showing that almost all stimuli were representative for their category, except for three stimuli. In Experiment 2, a free-naming task was conducted using the two datasets. We also recorded eye-tracking data during this task, to gain insight in focus patterns when detecting and categorizing an object.

Our similarity experiment revealed that similarity ratings did not differ with those in the Hartendorp study. Also, we found overlap between CP series in the Hartendorp study and current experiment with respect to the Hartendorp dataset. However, we did not find a correlation between perceptual similarity and CP series. We conclude that, based on this study, similarity does not seem to be the basic assumption of CP.

Concerning the eye tracking data, we found a strong correlation between focus and free-naming responses. This indicates that within-category, viewing patterns are consistent. When the category boundary is passed, focus also shifts to another area of interest (AOI). Based on these findings, we can conclude that human object recognition is preceded by efficient viewing patterns.

Index

Abstract 2

1. Introduction 5

Categorical perception 7

Object's skeleton 8

Experiment overview 9

Experiment 1: Verification task 10

2a. Method 10

Participants 10

Stimuli and materials 10

Procedure 11

2c. Results 11

2d. Discussion 12

Experiment 2: Similarity task 12

3a. Method 12

Participants 12

Stimuli and materials 13

Procedure 13

3c. Results 14

Results of the replicated similarity ratings 14

Stimuli selection for morphing 14

Experiment 3: Morphing task 15

4a. Method 15

Participants 15

Stimuli and materials 16

Procedure 16

4b. Data Analysis 17

Modelling morphing and CP patterns 17

Data assumptions for a CP pattern 18

Heat maps 19

4c. Results and discussion 21

CP testing 21

Matching the replicated and current study 25

Similarity as a predictor of CP 26

Matching Areas Of Interest with Free naming responses 28

5. General Discussion 28

Replication study 28

AOI's 29

6. References 30

7. Appendices

1. Introduction

Assigning detected objects into a correct category enables us to create a coherent percept of our world. In general, we are very skilled in interpreting the category in which new objects belong. In a fraction of a second, we are able to comprehend a novel object or image, in spite of degraded viewing conditions like partial obstruction, a uncommon viewpoint (rotation), distance (size on retina) or posture. To recognize an object, we primarily use shape, alongside size, color, texture, context and orientation to do so. For example, we don't need size to recognize a horse, we recognize a cow whether it is purple or brown (although we would probably be surprised and search for chocolate in case of a purple one) and a dog would be recognized as a dog whether it is bold or fuzzy, curled up or walking around. There is evidence provided for the importance of shape (Biederman & Ju, 1998) in that we are able to recognize many objects from basic line drawings or silhouettes, who eliminate a lot of information. Besides the use of these (bottom up) elements, semantic memory (top-down information) also supports object recognition performance, which will be discussed later on.

Moreover, we are able to generalize across object views very easily. Changes in viewing condition like viewpoint, illumination etc. do not hamper successful recognition, which is called 'invariance'. Theories about object recognition must provide an account of how observers compensate for a wide variety of changes in the image. These theories can roughly be categorized in viewpoint-invariant and viewpoint dependent approaches, who make different predictions regarding how invariance is achieved.

Viewpoint invariant theories assume that there are specific invariant cues in every object that are recovered under almost all viewing conditions. Examples of invariant cues are texture and colour, but also shape of object components and their interrelations. These invariants are thought to provide sufficient information to recognize the object. An example of a viewpoint invariant theory is proposed by Biederman (1987). Biederman's "Recognition by Components Theory" (RBC) proposes that visual input is matched with representations of objects in the brain. These structural representations consist of 'geons' (geometric icons) and their interrelations. According to the RBC theory, complex objects are represented as an arrangement of simple, convex, viewpoint-invariant 3D shape primitives, such as bricks, cylinders, wedges, and cones. Pattern recognition comprises of recognizing these components. Geons can be used to represent a large number of possible objects with very few components; e.g., 24 geons can be recombined to create over 10 million different two-geon objects. As long as we can extract two or more geons from an image that relate to each other, classification of this object is almost always successful. Drastic alteration of the object's silhouette or partial occlusion do not necessarily hamper recognition performance.

Viewpoint dependent theories (Tarr & Bülthoff, 1995) argue that no such invariants exist, but propose the storage of object representations, or mental images, in visual memory. These theories assume it is likely that these object representations are characteristic views of objects, so the mental image of a coffee machine is viewed from the front, and not viewed from the bottom. According to these theories, object recognition occurs by finding the closest match with the stored representations. The features visible in the input image are compared to features in object representations. This can be done by mentally rotating the image to the same viewing position as the stored representation. Viewpoint dependent theories assume we compute a statistical estimate of the match between the input image and stored representations.

Viewpoint dependent theories do not require a 3D shape in order to identify objects. A study by Tarr, Williams, Hayward & Gauthier (1998) suggests that three-dimensional object recognition can be viewpoint dependent, rather than independent. This argument follows partially from the observation that different regions of the brain respond to different viewpoints of a common object, for instance the human face. Since each region responds differently to different viewpoints of the same human face, it is unlikely that the brain uses a geon-based representation for face recognition. Object recognition probably relies on a viewpoint-dependent encoding schema, or mental image, where matches are determined by similarity. A study by Edelman and Bülthoff (1992) also supports this theory. They trained people to learn the shapes of non-existing objects that they created. Participants were tested on identification of the objects presented in the same or different viewpoint than seen during the training phase. They found that performance was much better for objects presented in the same viewpoint as presented during the training.

Viewpoint invariant theories (Biederman, 1987) assume that transformations do not alter object recognition, so changes in viewpoint or illumination (as long as for instance geons are recoverable) do not alter recognition performance. In contrast, viewpoint dependent theories are based on the thought that recognition is dependent on specific mental object representations. When a viewed object does not meet this viewing parameters, recognition performance decreases.

Another factor of object recognition humans are very skilled at, is the ability to recognize objects across different instances of an object category. Because objects of the same category should elicit the same response, an observer should be able to use their mental representation of that category to recognize new objects belonging to the same class. The lion we are staring at in the local zoo is unlikely to be the same as the one we see in the Kruger Park in South-Africa, but it would be fatal not to realize this is a lion as well. But, on the contrary, it is important not to confuse similar looking objects that belong to different categories. A peppermint or a medical pill for instance, should not be mixed up. These two goals of object recognition are termed by Marr and Nishihara (1978) as 'stability' and 'sensitivity'. A trade-off between these goals can be found.

The more one can generalize across objects because of viewing experience, the harder it may be to distinguish viewed objects from each other. On the other hand, when recognition of objects belonging to a certain category increases due to exposure to category members, abilities become more established and thus more sensitive, the worse an observer may be in deciding whether two objects belong to the same category.

Categorical perception

A requisite to assign an object to the correct category, is the existence of category boundaries. The boundaries of object categories are found to be quite strict. A technique that has been used to investigate these boundaries is 'morphing'. Morphing stands for the gradual change through a seamless transition of one image into another. When two objects are morphed with each other (see appendix 7.2.B for an example of two morphed objects), the switch point of the perceived category along the morphing continuum is very abrupt. At one point, objects are perceived as their first extreme, and when the category boundary is passed, the second extreme is the perceived object (Hartendorp, Stigchel, Burnett, Jellema, Eilers & Postma, 2010. This effect is described by a phenomenon that is known as categorical perception (CP; Harnad, 1987).

This phenomenon means that a change in sensory information along a continuum is perceived not as gradual but as instances of discrete categories. This perceptual invariance will last until the category boundary is passed, and will switch to another percept. Perception of different sensory modalities are found to be categorically perceived, for instance colour perception (Bornstein & Korda, 1984) or even higher order information like faces (Beale & Keil, 1995).

For instance, we take phonemes that belong to the same phonetic distinct and separate percept. Within a particular part of the continuum, the percepts are perceived as belonging to the same category, with an abrupt switch of perception at the position of the continuum where the category boundary is passed. Despite the gradual change of the information, percepts are not perceived as gradually changing. An example by Fitch, Miller, & Tallal, (1997) uses the consonants 'ba' and 'da'. By merging them together along a continuum from 'ba' to 'da', the percept shows a clear switch point at the moment the category border is passed.

Newell & Bülthoff (2002) and Hartendorp and colleagues (2010) conducted respectively a forced-choice and free-naming experiment to test morphed figures for CP. From the findings of these studies, the question raised why some series were perceived categorically and others not. In a subsequent experiment by Hartendorp et al., similarity ratings for the aspects of intrinsic part structure, shape, number of parts, semantics and phonology were compared for the morph series that were perceived categorically and non-categorically. The perceptual similarity aspects all showed higher similarity ratings for the CP-series than for the non-CP-series found by the free-naming task. It was observed that perceived similarity on all these aspects correlated positively with categorical perception. The restriction to observe categorical perception seems to be perceptual similarity between the two extremes of a morph series.

The goal of the current research is to determine the boundaries of object categorization using series of morphed figures. How do we detect the category an object belongs to, and how is information about the object rated in the allocation to a category. Can we alter the perception of the presented objects by affecting aspects like object contour, perspective and posture of the intrinsic parts? More specifically, we try to determine whether or not perceptual similarity is the underlying mechanism of categorical perception.

During this study, when the term (perceptual) similarity is used, it must be read as the intrinsic part structure of the object. In other words, the composition of parts of the two objects is similar. An example of two objects with a similar intrinsic part structure are a horse and a table; a horse usually has a horizontal body and four legs, as well as a table has a horizontal tabletop and four legs. We try to predict categorical perception by using sets of two images that are highly rated on similarity with respect to the intrinsic part structure. These objects are morphed with each other to test for CP.

Object's skeleton

Based on the similarity in features at the first half of the continuum of a morph series, objects are categorized into the same object class. However, along the continuum, more and more features are impaired when the second extreme becomes dominant. After the objects boundaries are passed, mostly halfway the continuum, the object no longer meets the criteria to fit in the non-dominant category, and the object is assigned to the new category. When we observe an abrupt switch between the percepts, we speak about CP. We expect that perceptual similar objects are more likely to show CP, because the so called skeleton (Blum & Nagel 1978, Feldman & Singh, 2006) of the two extremes is less impaired along the morphing continuum compared to two extremes that differ more from each other. After all, when the two images differ more from each other, the morphs have to bear more changes to merge together.

A skeletal representation of a visual shape shows the meaningful components of the object. Meaningful components are those that are perceived as distinct, essential components of the shape. This results in branches and sub branches representing the former object by emphasizing geometrical properties like connectivity, length and direction. Since the introduction of the theory of visual shape of the medial axis transform (MAT) by Blum and Nagel (1978), several models are proposed how to decompose shapes into parts. Feldman and Singh (2006) use a Bayesian model to compute the most effective skeletal representation of a visual shape.

If object recognition is based on the object's skeleton, it is likely that the rating of perceptual similarity between two objects is also based on the similarity of the skeletons. Following this theory, objects which are allocated to the same category have the same skeleton. For instance, two different cars have the same skeleton, as well as two different turtles have the same skeleton. So if skeletons form the basis of object recognition and categorization, than both series of cars and turtles should show categorical perception.

Experiment overview

In the current study, we try to validate the Hartendorp et al. (2010) study, by switching the CP experiment and similarity test. Hartendorp et al. (2010) first tested for CP and subsequently found a strong relation between CP and the perceived similarity for the extremes. If similarity is a predictor of CP, we should be able to combine objects with a similar intrinsic part structure and subsequently observe CP for these series. Objects with a low response on perceptual similarity should predict non-CP for the corresponding morphing series.

To test for this hypothesis, a new dataset of extremes was created. First, we will verify the identity of the new dataset consisting of 99 silhouettes in Experiment 1. Second, we will investigate how similar the extremes of morph series were rated on the aspect of intrinsic part structure (the perceptual similarity aspect most similar to the idea of a skeleton) in Experiment 2. We expect to find categorical perception for morph series that show high similarity ratings. The similarity ratings of the chosen extremes for the new dataset should be the same for the Hartendorp et al. (2010) dataset, so we expect to find CP for the same series where Hartendorp found CP. This would ground the idea that CP can be predicted by perceptual similarity ratings.

With this new dataset, besides the predictability, the robustness of the CP effect will be examined in Experiment 3. Does CP still occur using visual representations of the same category and the same similarity ratings, but with slightly changed visual features like posture or viewpoint, or between different members of the same category? This will ground the idea that skeletons are at the basis of object recognition, since the skeletons remained the same and only object contours are changed.

If the objects skeleton is used for object recognition, this results in the expectation that the entire image is scanned before determination of the objects category is completed. In Experiment 3, eye tracking was used to determine the focus of participants by measuring the point of gaze while detecting the object category to which the morphed object belonged. Do participants focus on details, like a face or object part, or do they use a global view of the object for categorization? Eye tracking data does not reveal what the observer has seen, since focus is no requirement for attention nor detection. Nonetheless, with the foregoing in mind, we assume that observers focus on the main information needed for object categorization. We expect to find eye-tracking gaze data that cover the whole object, since we expect to find that the whole skeleton of the stimulus is used to determine the object.

Experiment 1: Verification task

This experiment was used to verify the newly created dataset, to be able to select representative images for morphing later on.

2a. Method

Participants

Twenty students participated in the experiment. They all had normal or corrected to normal vision, and were fluent in Dutch. This experiment was combined with the second experiment. It took participants about half an hour for the two experiments, they received a fee or course credits for their participation.

Stimuli and materials

Stimuli pictures were obtained using common creative licensed photos on Flickr (www.flickr.com). They were selected with the only requirement that a large part of a high resolution picture was covered by the chosen object. Object categories were chosen based on the set of silhouette drawings used by Hartendorp et al. (2010). This set contained 30 images: a gorilla, dog, arm, banana, turtle, car, bow, bear, truck, peacock, pram, squirrel, bell, kettle, duck, church, cat, butterfly, bird, hat, man, lamp, gun, rabbit, apple, heart, airplane, crocodile, guitar and sea lion. These images were selected from a validated set of contour drawings, consisting of a wide range of living and nonliving objects (De Winter & Wagemans, 2004), which in turn were based on a set of line drawings by Snodgrass and Vanderwart (1980). The figure of a man was derived from another image set by Downing, Bray, Rogers, & Childs (2004). The validated status of the stimulus set came down on the already establishment identification rates of the images, so all images are representative for their category.

The viewpoint or posture of the object was not taken into account, since object skeletons should remain the same, despite the angle from which the object is viewed. Black silhouettes on a white background were drawn out of the pictures using Adobe Photoshop CS (8.0, Adobe

Systems, 2003). For each of the object categories, three to six images were created, loosely based on the number of object parts. So, simple objects, like a bell, were represented by three images and more complex objects, like a dog, were represented by five or six images. In total, 99 (30*(3, 4, 5 or 6)) new images were created and verified in this experiment. Every object category was represented by at least three images. Besides these newly created images, the extremes used as basis for the Hartendorp et al. (2010) morph series were also retested to make our replication as conscientious as possible. This made 129 silhouettes in total. The experiment was build using E-prime 1.1.4.1 (Psychology Software Tools Inc.), and was presented on a Philips Brilliance 202 P70 monitor.

Procedure

Experiment 1 consisted of 258 trials (99*2+30*2). Sixteen practice trials preceded the experiment consisting of object silhouettes that were not part of the actual experiment. Stimuli were presented randomly. The stimuli were preceded by a fixation cross in the middle of the computer screen for 500 milliseconds, followed by an object-category name in the middle of the screen for 1000 ms. Every stimulus was presented twice: once preceded by the corresponding object name, and once with an object name that belonged to another object from the stimulus set. At the start of the testing procedure, participants were instructed on screen to react on the 'question' (object name and question mark) presented, by pressing 'f' or 'j' on the keyboard, which corresponded respectively with 'no' or 'yes'. The 'f' and 'j' key were chosen because of the tactile aids that improve the accuracy and speed of pressing by reducing the need to look at the keyboard. Stimuli presentation lasted until the 'f' or 'j' key was pressed. There were no time restrictions for responding. A new trial followed after a blank screen that was presented 0.2 ms. The stimuli and object names that did not correspond were combined randomly. Participants were instructed to react as quickly as possible and therefore hold their index fingers on the 'f' and 'j' keys. They were also asked to visualize the named object to enhance a quick reaction time.

2c. Results

Based on the verification task, average reaction times (RTs) in milliseconds (ms) were calculated. These reaction times indicated the representativeness of an object for its category. Almost all average reaction times were situated between 450 and 700 milliseconds. Stimuli with large (above 1000 ms) reaction times were excluded from the dataset, since these reaction times were obvious outliers. Only two images met these criteria with reaction times of 1305.1 and

2192.4 ms. Furthermore, images that showed more than four wrong answers were perceived as less representative for the corresponding category. Besides the already excluded items (that also did not met this criterion), this rule excluded one more item.

When we compare the reaction times of both datasets, we find similar means (582.15 ms for the Hartendorp dataset and 587.10 ms for the new dataset). Since the Hartendorp dataset is based on a verified (by Snodgrass and Vanderwart, 1980) set of images, we assume the new dataset can be interpreted as representative for the corresponding categories as well.

2d. Discussion

In Experiment 1, a verification task was used to test whether or not stimuli are representative for their category. Both the Hartendorp and new dataset were tested. Participants were asked if an object name corresponded with an image. Reaction times as well as answers (yes/no) were recorded.

Based on the cohesive reaction times between images, it is concluded that all images (besides, of course, the three excluded images) can be used as representative for the corresponding category. However, based on the reaction times in the verification task, no conclusions about object recognition can be made. Because participants are primed by the object word that preceded the stimulus, reaction times are presumably shorter than the time span of object recognition by naive individuals. Besides, actual object recognition duration cannot be measured at all, since reaction times include lexical retrieval and verbal expression of the object name.

Experiment 2: Similarity task

3a. Method

Participants

Participants were the same as the ones participating in Experiment 1, since the similarity task directly followed the verification task.

Stimuli and materials

Because of the distinct nature of the verification and similarity task, presentation of the two tasks in this particular order has no influence on the outcome of the similarity task. For the verification task, naïve participants were necessary to gain trustworthy reaction times. For the similarity task however, recognition of the images would not interfere with the perceived similarity between images. All 129 images, from which 30 came out of the Hartendorp et al. (2010) dataset, and 99 out of the new dataset, were used during this task. Furthermore, all stimuli and materials were the same as used for Experiment 1.

Procedure

The current task followed after finishing the verification task. Participants were able to take a break before the introduction of the similarity task was started. In the stimulus set used by Hartendorp and colleagues (2010), two extremes from different categories were combined as constitution of a morph series. Different object images, but belonging to the same category as the pairs used by Hartendorp et al. were as well combined for testing in the present experiment. They were selected from the 99 newly created silhouettes. For example, three different newly created silhouettes of arms were combined with the newly created silhouettes of bananas (see Appendix 7.1). All possible combinations of the extremes from the new dataset, for instance all arm silhouettes with all banana silhouettes and all car silhouettes with all turtle silhouettes were tested. In the similarity task, the dataset used by Hartendorp et al. (2010) was retested with the same pairs of extremes as used in their study. This made 208 (15 Hartendorp series + 99 images combined based on the Hartendorp series) trials in total.

Participants were asked to rate the similarity (by looking at the intrinsic part structure) between the presented combinations on a seven-points-rating-scale with 1 referring to no similarity and 7 referring to strong similarity.

Participants were instructed to use the complete range between 1 and 7 to rate the similarity between images. They were given two example trials similar to the actual trials. An image of a boat and a flower and a table and a horse were presented on one screen. At the head of the page, the question (translated from Dutch) ‘How strong do you rate the similarity between the images concerning the intrinsic part structure?’. This question was followed by the rating scale (in Dutch):

Very week 1 2 3 4 5 6 7 very strong

In the two practice trials, the range was followed by a red note (a 2 for the boat-flower and a 6 for the table-horse combination). Previously, participants were instructed that their appraisal could differ from the rating given by the experimenter, and that no good or wrong answers existed. During the actual experiment the number keys from 1 to 7 were used to enter the rating. Visual feedback of the typed answer was given on the screen. Participants were able to modify their answer as long as the screen was presented. Participants were given as long time as needed to finish a trial, and were instructed to press the ENTER key to proceed.

3c. Results

The E-prime output per participant was copied to Microsoft Excel and the mean ratings were computed.

Results of the replicated similarity ratings

A paired samples t-test was conducted to compare the similarity ratings from the Hartendorp et al. (2010) testing and the replication of this study in the current experiment. It was found that these ratings did not differ significantly from each other $t(14) = -.65$, $p = .526$.

These results form a good basis for further testing, since it seems that so far, the replication of the Hartendorp study is successful. Based on these findings, we expect to find similar findings in the free-naming study as were found by Hartendorp et al. (2010).

Stimuli selection for morphing

The similarity ratings found for the extreme pairs of the new dataset were used to select extreme figures that would be morphed to create new morph series. The new morph series would resemble the old morph series as much as possible in their similarity ratings. Thus, pairs of extremes from the new dataset with an average ranking closest to the average given for the Hartendorp dataset were chosen. Consequently, similar findings of CP (and non-CP) were expected for the old and new morph series, since they matched strongly on similarity of the extreme figures. To illustrate the comparison of the similarity ratings found for the old and new pairs we use the crocodile-airplane pairs as an example. When the Hartendorp et al. (2010) pair was given a 4.55, and the combinations of extremes from the new dataset were given the numbers shown in Table 1, the last combination (4.45) was chosen to be used for morphing and further testing.

Table 1. Example of similarity ratings on a 7 point scale for all possible combinations of airplane-crocodile silhouette images.

airplane1 – crocodile1	airplane1 – crocodile2	airplane1 – crocodile3	airplane2 – crocodile1	airplane2 – crocodile2	airplane2 – crocodile3	airplane3 – crocodile1	airplane3 – crocodile2	airplane3 – crocodile3
2.74	3.55	2.7	3.85	5.15	4.1	3.75	4.79	4.45

Experiment 3: Morphing task

With this experiment, we try to find a replication of CP patterns for the Hartendorp et al. (2010) dataset. Also, we try to find CP patterns for the same series of the new dataset. For example, if Hartendorp et al. found a CP pattern for the banana-arm morph series, we expect to find a CP pattern for the banana-arm morph series of the new dataset due to the great overlap in similarity observed for the extreme figures.

Furthermore, we use eye-tracking gaze data to analyze focus patterns. We expect viewers to gaze over the complete object area to recognize it, based on the skeleton theory by Hartendorp et al. (2010) in contrast to focusing on one particular part of the object.

4a. Method

Participants

Twenty students participated in the experiment. It took participants about half an hour and they received a fee or course credits for their participation. All participants had normal or corrected to normal vision and were fluent in Dutch. Two participants also participated in Experiment 1 and 2, but this was not expected to influence the results, because these participants had only seen the extremes of the morph series and were unfamiliar with the morphs. Possible priming effects would be ruled out by the random presentation of the images.

Stimuli and materials

In the morphing task, the paired categories were identical to the series in the Hartendorp et al. (2010) experiment. The images to be morphed were chosen based on the similarity task (experiment 2). The paired objects were morphed using Sqirlz-Morph software (Xiberpix, version 2.0). These interpolations were made by positioning markers on the boundaries of the two combined silhouettes. Each morph series consisted of 19 interpolations and two extremes, to obtain 5% change between stimuli. Similar to the Hartendorp et al. (2010) experiment, the 0%100%, 20%80%, 30%70%, 40%60%, 50%50%, 60%40%, 70%30%, 80%20% and 100%0% images were selected for use. These 9x15 images were edited using Adobe Photoshop CS software (see appendix7.2.A and 7.2.B. for example series). Grey spots, that were a result of the technique used by the morphing software, were replaced by black or white, depending on their shade of grey, so only black silhouettes on a white background remained in the images. A total of 270 trials were tested (15 morph series from the Hartendorp dataset x 9 figures per series + new dataset 15 morph series from the new dataset x 9 figures per series).

Out of the 30 (2x15) morph series, four pairs consisted of two living objects, six pairs consisted of two non-living objects, and twenty pairs of a living and non-living object (see Appendix 7.3 and 7.4).

The experiment was designed using E-prime version 1.2.1.844, and the additional TET (Tobii EyeTracking) package version 1.0.3.0. The eye tracker used for this experiment was a Tobii 1750 which looked like a plain monitor. Input devices were a keyboard and the E-Prime Stimulus Response box (SR-box). A microphone was connected to the SR-box. This microphone was used to end stimulus presentation by a sound cue and present the next slide. No records were made. A chinrest at a 52 cm distance of the screen was used to stabilize the position of the eyes in relation to the eye tracker to gain optimal eye-tracking results. Also, the distance between the mouth in relation to the microphone was kept stable. Port one, two, three and four of the SR-box were used to code the answers as the correct extreme, the incorrect extreme, an alternative answer (an answer that could not be subscribed to one of the two extremes) or an inaccurate answer that needed to be rejected as a result of blowing in the microphone or other flaws, respectively. During the experiment, the experimenter sat behind a second computer screen that showed the answer codes belonging to the image presented to the participant.

Procedure

Prior to the experiment, the Tobii eye tracker was calibrated using the Tobii software 'ClearView' (version 2.7.1). Participants were asked to place their head in the chinrest, and visually follow a black circle that moved over the entire screen. Once this was done, instructions about the experiment were given. Once they had read the instructions, the experimenter emphasized the possibility that stimuli might look indefinable, but that it was important that participants would say their first impression out loud. Participants were asked to avoid sighs or other noise, since the microphone was very sensitive to sound. Four practice trials preceded the experiment. These trials consisted of non-morphed images, but followed the same procedure as the actual experiment. During the actual experiment, 270 images in the same number of trials were presented, since all 15 morph series consisted of 9 images, and 2 datasets were tested. All stimuli were presented randomly, with the restriction that two figures of the same series could not be presented successively.

Participants were instructed that the experiment was about their interpretation of the objects, meaning that no incorrect answers could be given. They were asked to say their answers loudly, and use one word that described the object as accurate as possible. There were no time restrictions, but they were asked to answer as quickly as possible.

During the task, the experimenter stayed in the testing room. Answers given by the participant were immediately labeled by the experimenter making use of the SR-box. On the screen in front of the experiment the stimulus was presented similar as to the participant together with the correct and incorrect answers. The experimenter coded the given answer by pressing the buttons on the SR-box. Four categories were used: correct answers received code 1, answers that belonged to the other end of the morph continuum received code 2, completely wrong answers were given code 3 and errors due to sound recording errors such as blowing in the microphone were given code 4.

4b. Data Analysis

Modelling morphing and CP patterns

The E-prime data file was rearranged to make analyses of the data possible. Participants were placed at the vertical axis, and the 270 images were presented horizontally. Code 4 answers (errors) were excluded from analyses. The remaining responses were re-coded as 1 for one extreme (e.g. rabbit to a target of a rabbit-gun series), 2 for the other extreme (e.g. gun to a target of a rabbit-gun series), or a score of 3 for alternative answers (e.g. cat or monkey to a target of a rabbit-gun series). This data matrix was generated to compute frequency percentages using SPSS 16.0.2 (SPSS, Inc., Chicago IL). Subsequently, the same statistical analyses as for the Hartendorp et al. (2010) experiment was executed, using the Birnbaum model. A brief description of this model will be outlined below, a more detailed explanation can be found in Hartendorp et al. (2010).

The Birnbaum model (1968) or four parametric model is designed to imitate data and to be faithful to the data as good as possible. The goal of the Birnbaum model in our case is to measure the steepness and the moment of switching from one percept to another. Multiple answers can be given by the participant: dominant and non-dominant responses, but also alternative interpretations are possible. Due to these errors, no logistic model can be used, but the Birnbaum model can give a good fit to our data:

$$p(x)= \gamma+(1- \gamma- \delta)\ 1/1+exp[-\alpha(x- \beta)]$$

This model fits the data in s-shaped curves with varying ranges, slopes and midpoints. Four parameters have to be estimated, namely α, the slope, which determines the steepness of the curve (abrupt switch or not from one percept to another), the morphing level β (the moment of switching) and the expected fraction of errors made γ (δ). Since the latter are probabilities, their values have to be between 0 and 1.

Estimating parameters for the Birmbaum model by maximum likelihood can only be done by numerical optimization. A Microsoft Excel spreadsheet developed for analysis of the Hartendorp et al. (2010) experiment was reused for analysis of the current dataset. The Excel add-on 'Solver' was used to set the starting estimates for the parameters. The default set was stated at $\alpha = 10$, $\beta = 0.5$ and $\gamma = \delta = 0$.

Data assumptions for a CP pattern

An ideal CP pattern is probably not observed due to alternative responses and individual differences in response patterns. Therefore, we speak about an approximation of CP. A pattern of CP is observed when two assumptions are met. First, X50 refers to the morphing level β and is a measure that indicates the switch point of perception of the two extremes of a morphing series. X50 is based on the morphing level at which the y-axis has reached 0.50. This measure should be as close as possible to 0.50, but at least in-between 0.40 and 0.60. When X50 is at the midpoint of the continuum, we observe a response pattern where the fifth (and thus middle) image out of the nine-images morphing series indicates the switch point from perception of the one extreme to the other (50%/50%).

The other measure that has to be met for a CP pattern is a steep slope α. To make this term more comprehensible, the 'Gap' is computed using the formula:

$$G=4(1 - \gamma - \delta)/ \alpha.$$

This 'Gap' is a measure of the steepness of the slope, and thus abruptness of the switch from one percept to the other. An ideal pattern of CP would be a small G, and thus a steep slope, but this might be hard to find, since observers show individual differences in their responses. In addition, alternative (other than the two extremes) answers could have been given due to the free-naming experiment. We speak of an approximation of CP when the found G is between 0.00 and 0.24.

Heat maps

To provide insight into the eye-tracking data, a graphical representation of this data in the form of heat maps was generated. A heat map is a visual display of the most frequently scanned areas of an image by using a color gradient overlay. This color gradient indicates the average viewing pattern by using red tones for the most thoroughly viewed spots of the image. The more blue the less participants gazed over the area.

Heat maps of the morph series are based on participants' view of the stimuli without time restriction. Some heat maps cover a longer viewing time than others, as a consequence of a longer decision process until the object was categorized (i.e. an answer in the free-naming task was given). Heat maps were generated by the use of custom-made software (I. Hooge, 2010). All heat maps can be found in Appendix 7.6.

To prepare the heat maps for analysis, Areas of Interest (AOI's) of the extremes of the morph series were created using Adobe Photoshop CS4. AOI's were placed on the red fixation areas of the heat maps. The AOI's of both extremes were copied as a layer on the entire range of a morph series (for an example, see Appendix 7.7) A selection of heat maps was used for further analysis. Extremes with more than one AOI were ruled out, because the relative coverage with AOI's of the image was too large for reliable analysis. Series with two extremes with partial or completely overlapping AOI's could not be used either, since allocating the AOI's of the morphs to one of the two extremes was impossible due to the overlap. The remaining eight series showed two AOI's (one per extreme) on all nine images of the morph series. Out of the Hartendorp dataset, the arm-banana, lamp-man, cat-butterfly and bear-bow series were used. Out of the new dataset the dog-gorilla, duck-church, lamp-man and peacock-truck series were selected. Unforeseen, four of these eight series came out of the Hartendorp dataset and four from the new dataset.

After this, the images were coded in SPSS 17.0 with 100%, 50% or 0%. A 100% score was given to morphs with a similar heat map as the reference extreme. A similar heat map was diagnosed when the red parts of the heat map of the morph had at least ¾ overlap with the AOI of the reference extreme. A 0% score was given when the heat map of the morph had at least ¾ overlap with the AOI of the other extreme. A 50% score was given when the heat map did not correspond with the heat map of one of the two extremes. This was the case when the red parts of the heat map fell in neither one of the two AOI's, or when one or both AOI's were partially covered. Each series was coded twice, once the first extreme was used as the reference, and once the second extreme was used as the reference. An example of the resulting table is shown in Table 2. The percentages correct of the free-naming task are also copied into this table.

Table 2. Example of SPSS file used to compare the Free naming data with AOI scores.

	Filename	Free-naming score correct	AOI score
Gorilla used as reference extreme	Dogrilla100%0%	100,0	100,0
	Dogrilla80%20%	95,0	100,0
	Dogrilla70%30%	100,0	100,0
	Dogrilla60%40%	94,4	100,0
	Dogrilla50%50%	95,0	100,0
	Dogrilla40%60%	52,6	50,0
	Dogrilla30%70%	5,0	0,0
	Dogrilla20%80%	0,0	0,0
	Dogrilla0%100%	0,0	0,0
Dog used as reference extreme	Dogrilla100%0%	0,0	0,0
	Dogrilla80%20%	5,0	0,0
	Dogrilla70%30%	0,0	0,0
	Dogrilla60%40%	5,6	50,0
	Dogrilla50%50%	5,0	100,0
	Dogrilla40%60%	47,4	100,0
	Dogrilla30%70%	95,0	100,0
	Dogrilla20%80%	100,0	100,0
	Dogrilla0%100%	100,0	100,0

4c. Results and discussion

CP testing

Tables for the 60 (2*15*2) individual ranges of extremes within a morph series were created using Microsoft Excel. Based on a table presenting answer percentages (already created using SPSS) for all nine targets in a morph series, a graph was created in the datasheet. Subsequently, the add-on 'Solver' was used to fit a logistic curve within the range of the observed data. This curve with the best fit (see Fig. 1 for an example) visualized an estimation of the fraction positive answers per extreme and was plotted in the data graph. The fraction positive answers stood for the percentage of answers that corresponded with the category of the extreme. The percentage alternative answers and answers corresponding with the other extreme were subtracted from this number. Based on the default parameters, the Gap and X50 were calculated for each extreme (see Fig. 1). These values were compared to the ideal Delta Gap and X50 values for a pattern of CP. Based on the goodness-of-fit of the estimated curve compared to an ideal CP pattern, the decision was made whether or not series showed a CP pattern or no distinct pattern (CP = Gap<.024 & 0.40<X50<0.60).

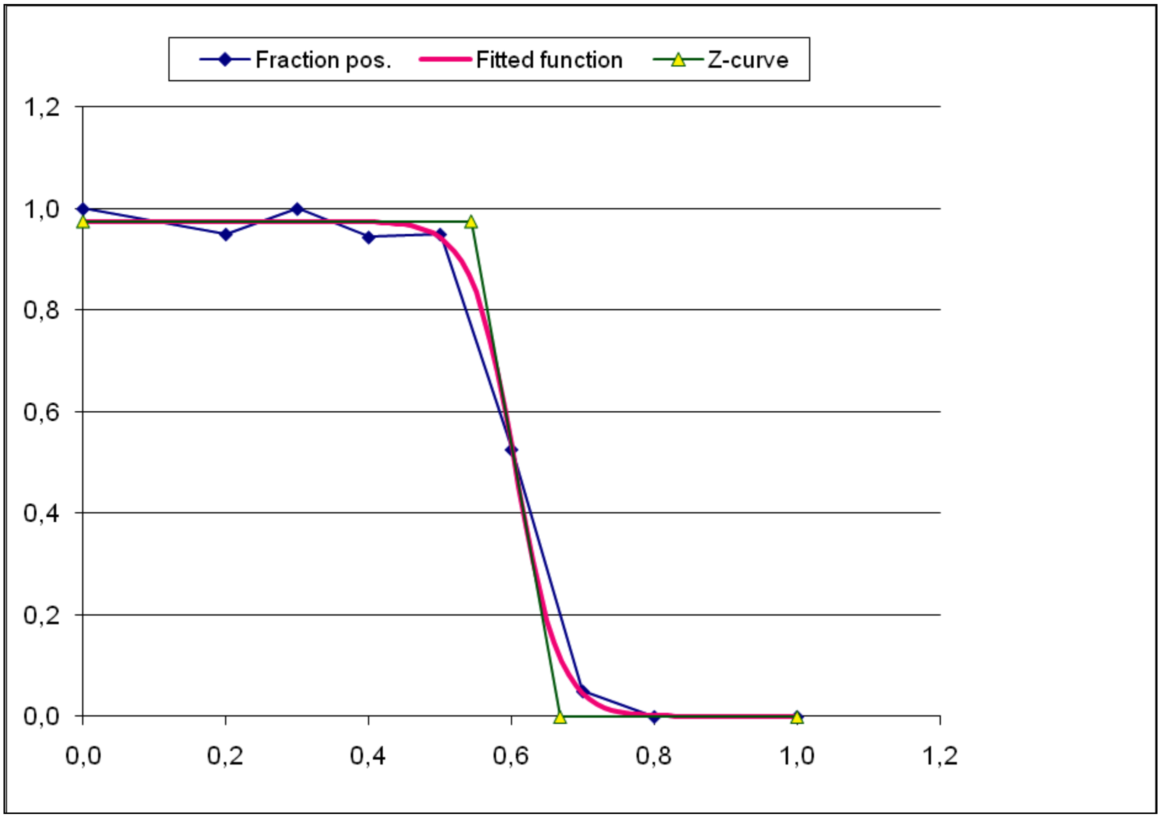

Fig.1. An example of the fraction positive answers for one of the two extremes of a series, and the fitted curve based on this data.

After allocating the series into the CP or non-CP category, the new dataset was compared to the Hartendorp study and the current retest. As can be seen in Table 3, considering CP patterns, no perfect agreement between the results of the two datasets tested in the current experiment was obtained. Five series showed a CP-pattern for both datasets and five series showed no CP-pattern for both datasets. Two series did show a CP-pattern for the Hartendorp dataset, but failed to show a CP-pattern for the new dataset and three series showed no CP-pattern for the Hartendorp dataset, but did show a CP-pattern for the new dataset.

Considering the retest of the Hartendorp dataset and the Hartendorp et al. (2010) experiment, no perfect agreement was found either. Five series showed a CP-pattern for both tests and six series showed no CP-pattern for both tests. Two series did show a CP-pattern for the Hartendorp retest, but failed to show a CP-pattern for the Hartendorp experiment and two series showed no CP-pattern for the Hartendorp retest, but did show a CP-pattern for the Hartendorp experiment.

Table 3. The series marked with green show a cp pattern, for the series marked with red one of the two or both extremes did not met the cp criteria.

	Hartendorp et al. (2010) study	Hartendorp dataset X50 (0.40-0.60)	Hartendorp dataset Delta (max0.24)		New dataset X50	New dataset Delta
banana		0,54	0,35		0,38	0,45
arm		0,44	0,42		0,59	0,46
bell		0,52	0,22		-0,17	1
kettle		0,66	0,33		0,39	0,19
crocodile		0,38	0,24		0,48	0,14
airplane		0,5	0,31		0,42	0,24
cat		0,29	0,19		0,42	0,21
butterfly		0,55	0,31		0,5	0,01
pram		0,55	0,26		0,59	0,03
squirrel		0,48	0,23		0,5	0,03
car		0,54	0,22		0,51	0,09
turtle		0,54	0,22		0,27	0,32
rabbit		0,7	0,22		0,4	0,13
gun		0,48	0,18		0,43	0,26
apple		0,38	0,22		0,72	0,23
hart		0,38	0,22		0,72	0,23

lamp		0,56	0,23		0,59	0,02
man		0,53	0,23		0,56	0,11
gorilla		0,47	0,24		0,6	0,12
dog		0,47	0,24		0,6	0,12
bow		0,47	0,21		0,46	0,22
bear		0,45	0,21		0,5	0,05
hat		0,49	0,4		0,26	0,28
bird		0,68	0,37		0,35	0,18
guitar		0,51	0,03		0,4	0,15
sea lion		0,52	0,04		0,49	0,02
truck		0,54	0,09		0,48	0,17
peacock		0,54	0,08		0,47	0,18
duck		0,45	0,09		0,39	0,15
church		0,47	0,12		0,49	0,03

Matching the replicated and current study

Since the morphing dataset used by Hartendorp et al. (2010) was retested in the current experiment, at least for this dataset we expected to find CP where Hartendorp and colleagues had found CP, and non-CP where they had found non-CP. Based on the previously stated criteria, seven out of fifteen morph series from the Hartendorp dataset showed a CP pattern. These were the series car-turtle, lamp-man, gorilla-dog, bow-bear, guitar-sea lion, truck-peacock and duck-church. Out of the newly created dataset, eight out of fifteen series showed a CP pattern. These were the series crocodile-airplane, cat-butterfly, pram-squirrel, lamp-man, gorilla-dog, bow-bear, guitar-sea lion and truck-peacock.

When we look at the data (see Table 3), we notice differences between Hartendorp et al. and the current data for whether a series is showing a CP pattern or not. Also, different CP/non-

CP patterns are observed for the Hartendorp dataset (tested in the current experiment) and the new dataset. For example, the crocodile-airplane series showed CP in the original Hartendorp experiment, but failed to show CP in the replication study, which was an exact copy of the original study but with a smaller sample size. For this crocodile-airplane series, the new dataset however did show a CP pattern.

A Cohen's kappa has been executed to test for overlap between the CP/non-CP arrangement of the Hartendorp series (tested in the Hartendorp and current experiment). This statistical measure tests for inter-rater agreement that takes the agreement occurring by chance into account. In this case, testing was executed by only one rater, but our motive to use this statistical measure was to test for homogeneity between two participant samples that executed the same test. The first sample of participants was imported as rater one, and the second as rater two. A value of 0.60 was found for the overlap between the Hartendorp and current experiment, which is a reasonable score. Differences could be due to the smaller amount of tested participants in the current study (20) compared to the Hartendorp experiment (83). Probably, alternative answers and individual answer patterns will have larger influence on a smaller population than on a large one. This will be explained more extensively in the General Discussion.

Similarity as a predictor of CP

We also tested for homogeneity between the two tested datasets, Hartendorp and newly created, in the current experiment. A Cohen's kappa was executed and a value of 0.20 was found for the overlap between the two datasets tested in the current experiment. This indicated that there is only a slight agreement between the two datasets, despite the completely equal testing environment, the same rater and participants.

Based on the finding in the Hartendorp et al. (2010) experiment, perceptual similarity is thought to be a predictor of CP, since high similarity ratings are strongly associated with categorically perceived series. Correlations, based on the average delta gap (α), between the two datasets and similarity ratings are computed. The smaller the Delta Gap is, the more categorically perceived the series are. To confirm our hypothesis, correlations should be negative, so the smaller Delta Gap is, the higher perceptual similarity ratings are given. A Pearson correlation was conducted for the Hartendorp dataset, $r = 0.41$ and $p > .05$, and the new dataset, $r = 0.51$ and $p > .05$. However, they are both found not to be significant. Based on these findings, it can be concluded that perceptual similarity and CP are not correlated. This indicates that the finding of CP is not caused by perceptual similarity of the extremes.

During further testing, the kettle-bell series were excluded from analysis due to off target answer patterns in the morphing task: this series showed that –despite the verification task- the bell-extreme was not recognized as a bell by most participants. This inconsistent pattern was probably due to presentation of the response option and comparison of the image with this concept in the verification task, instead of the free-naming design in Experiment 3.

Subsequently, we conducted a 2x2 Repeated Measures ANOVA, with the within-subject factor ‘CP’ (CP-series vs. non-CP-series) and the between-subject variable ‘dataset’ (Hartendorp vs. new). We expected to find higher similarity ratings for CP-series than for non-CP series. A significant main effect for the variable ‘CP’ was observed, $F(1, 19) = 8.23$, $p = .01$, with higher similarity ratings for the non-CP series than for the CP series. We also found a significant main effect for the variable ‘dataset’, $F(1, 19) = 5.5$, $p = .03$, which indicates higher similarity ratings for the Hartendorp dataset. However, no interaction effect between CP and similarity was found $F(1, 19) = 2.18$, $p = .16$. Instead we observe a trend (see Fig. 2) of higher similarity ratings for non-CP series. These findings however are not significant, so similarity is not found to be a predictor of CP nor non-CP.

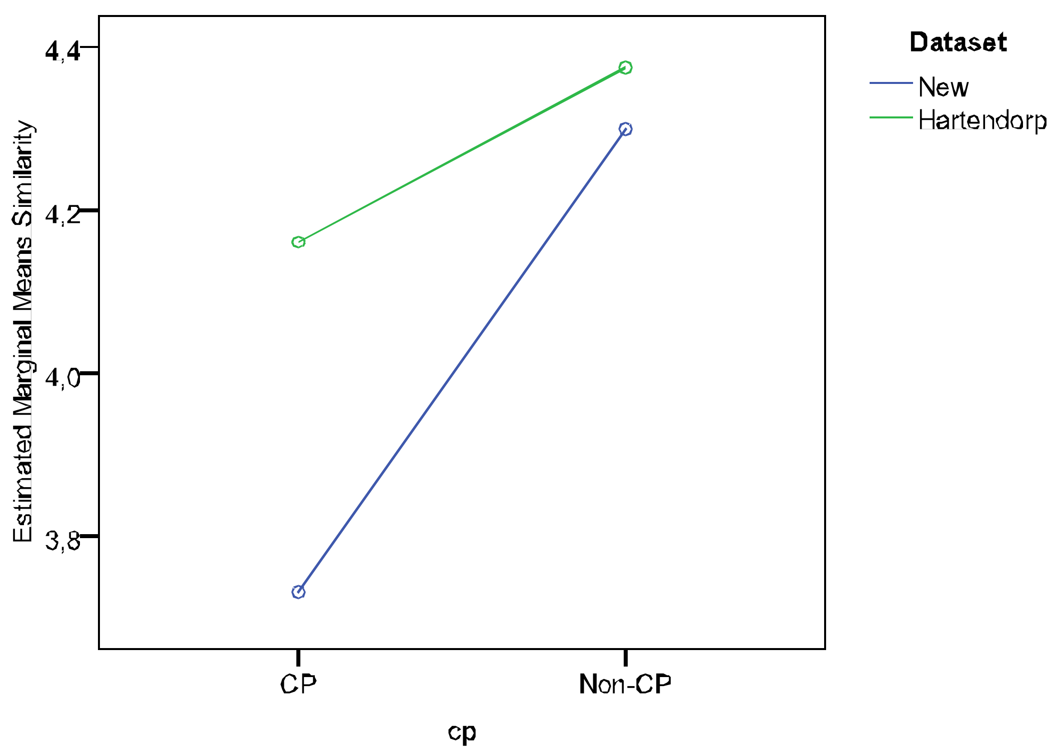

Fig. 2. Results of the 2x2 Repeated Measures ANOVA, as an illustration of the (not significant) trend towards higher similarity ratings corresponding with non-CP.

Matching Areas Of Interest with Free naming responses

Table 2 was used to match the AOI's with the free naming responses using a bivariate correlation in SPSS 17.0. A strong Pearson correlation of $r(142) = .87$, $p < .001$ between AOI and Free naming responses was found. The reason for using N=144 is because we used 8 morph series out of 30 series, all existing of 9 figures. Each series was used twice, since both extremes from one series were coded.

5. General Discussion

The goal of this study was twofold: on one hand, we tried to predict categorical perception of morphed objects based on perceptual similarity. On the other hand, we tried to support the idea that object recognition occurs by using the object's skeleton, instead of single object's parts. First, we started verifying a newly created dataset in Experiment 1, by measuring the representativeness of images for their category. Second, in Experiment 2, a similarity task was used to select images that would be morphed, based on similarity ratings. During a free-naming experiment with image datasets, besides testing for CP, we also recorded eye-tracking data to gain insight in visual focus patterns when detecting and categorizing an object.

Replication study

In the last experiment of this study, we tried to predict CP by using perceptual similarity ratings found in Experiment 2. Based on the finding of high similarity ratings correlating with CP series, and low similarity ratings with non-CP series in the Hartendorp et al. (2010) experiment (see also Newell & Bülthoff, 2002), we expected similar findings in the current experiment. However, the Hartendorp study tested for CP first, and subsequently linked the results to perceptual similarity. The current study switched CP-testing and similarity rating to test the predictability of CP by the similarity ratings. In addition to the existing Hartendorp dataset, a new stimulus-dataset was developed and tested to generalize the CP/similarity effect beyond the Hartendorp dataset.

Our similarity experiment revealed that similarity ratings did not differ significantly from those in the Hartendorp study. Also, we found overlap between CP series in the Hartendorp study and current experiment with respect to the Hartendorp dataset. Unfortunately, positive, but not significant correlations between similarity and CP (delta gap) were found for both datasets in the current experiment. In other words, we found higher similarity ratings for non-CP than for CP series. However, since these findings were not significant, we did not find perceptual similarity being a predictor of CP nor for non-CP for neither one of the two datasets. Our findings do not support the theory that CP could be predicted using similarity ratings, since we did not found a connection between perceptual similarity and CP. We conclude that, based on this study, similarity does not seem to be the basic assumption of CP.

There are a couple of possible explanations for the different outcome of the current study compared to the outcome of the Hartendorp study. First, due to a smaller sample of participants for the current experiment compared to the Hartendorp et al. (2010) study, the differences for the replicated parts can be explained. First, the delta gap is based on the percentage a particular response is given. Each participant sees each figure just once, meaning that the percentage a particular response is given is based on just twenty participants in case of the current study and on more than eighty participants in the Hartendorp free-naming study. Therefore, if three participants give another response than the dominant interpretation (or correct one), this will affect the outcome more in case of twenty participants than in case of eighty participant. Subsequently, with less participants, the Delta Gap will increase considerably, and as a result, series will turn out to be less (or not) categorically perceived. We argue, therefore, that the great difference in sample size is underlying difference between the findings of the previous and the current study. Second, another possibility due to individual differences is shifting of the X50 switch from the middle of the series to the one or other side of the morphing continuum, with the lack of CP as a possible result. Increase of the testing population size will rule out these effects for the actual CP series, or amplify them for actual non-CP series. Further research, with a larger test group should be done to verify this presumption.

AOI's

In the first instance eye tracking data was collected to test for the use of the object skeleton for object recognition. Progressively we got to the insight that eye-tracking data could not be used in such a manner, since viewing does not necessarily mean seeing. Eye tracking data does not reveal what the observer has noticed, since focus is no requirement for attention nor detection. As a consequence, the findings mentioned below occurred more or less serendipitous. Nonetheless, the results of this data are interesting, and can be used for further testing.

Based on the images' heat maps, areas of interest (AOI) of the morph series' extremes were positioned. Subsequently, the images were coded, based on the position of the AOI on the heat map of the morph compared to the AOI's of the extremes. Subsequently, we looked at the resemblance between the position of the AOI of a morphed image and the corresponding response in the free-naming task. A strong correlation between the position of the AOI's and free-naming responses was found. This indicates that within-category, viewing patterns are consistent. In other words, images from the same category demonstrate similar heat maps. When the category boundary is passed, and another interpretation is given to a figure, visual focus also shifts to another area of interest (AOI). It can be concluded that the viewers' focus of images within-category is consistent, and corresponds with the final categorization of an object. When the category boundary is passed, the object is allocated into a new category and the focus shifts to another AOI. Based on these findings, we can conclude that human object recognition is associated with very efficient viewing patterns.

Furthermore, this data shows a relation between categorization and eye-movement patterns. However, this data does not reveal the order of this relation: do eye-movement patterns precede object categorization, or does object categorization precedes human eye-movement patterns? This would be an interesting topic for further research, since these findings imply that it might be possible to predict object categorization based on eye-movement patterns. Another possibility for further research lies in the use of the current dataset, since not all information of this dataset has been used in the current study. For instance, it would be interesting to analyze and compare viewing patterns of living and non-living objects.

6. References

Beale, J.M., & Keil, F.C. (1995). Categorical effects in the perception of faces. *Cognition, 57*, 217-239.

Biederman, I. (1987) Recognition-by-components: a theory of human image understanding. Psychol Rev. 1987 Apr;94(2):115-47.

Biederman, I., & Ju, G. (1988). Surface versus edge-based determinants of visual recognition. *Cognitive Psychology, 20*, 38-64.

Birnbaum, A. (1968). Some latent trait models. In F.M. Lord & M.R. Novick, (Eds.), Statistical theories of mental test scores. Reading, MA: Addison-Wesley.

Bornstein, M.H., & Korda, N.O. (1984). Discrimination and matching within and between hues measured by reaction times: Some implications for categorical perception and levels of information processing. *Psychological Research, 46*, 207-222.

Blum, H. and Nagel, R. 1978. Form description using weighted symmetric axis features. *Pattern Recognition, 10,* 167-180.

Edelman, S., Bülthoff, H. H. (1992). Orientation dependence in the recognition of familiar and novel views of three-dimensional objects. *Vision Research,* 32, 2385-2400.

Fitch, R.H., Miller, S., & Tallal, P. (1997). Neurobiology of speech perception. *Annual Review of Neuroscience, 20*, 331-353.

De Winter, J., & Wagemans, J. (2004). Contour-based object identification and segmentation: Stimuli, norms and data, and software tools. *Behavior Research Methods, Instruments, & Computers, 36* (4), 604-624.

Downing, P.E., Bray, D., Rogers, J., & Childs, C. (2004). Bodies capture attention when nothing is expected. Cognition, 93, B27–B38.

Feldman, J. & Singh, M. (2006). Bayesian estimation of the shape skeleton. *PNAS, 103* (47),18014-18019.

Harnad, S. (1987). Psychophysical and cognitive aspects of categorical perception: A critical overview. *Chapter 1 of: Harnad, S. (ed.) (1987) Categorical Perception: The Groundwork of Cognition.* New York: Cambridge University Press.

Hartendorp M.O., van der Stigchel, S., Burnett, H.G., Jellema, T., Eilers, P.H.C. & Postma, A. (2010), Categorical Perception of Morphed Objects Using a Free-Naming Experiment.

I. Hooge. (2010). Heat map generating software for the Tobii eye-tracker.

Marr, D. and Nishihara, H. K. (1978). Representation and recognition of the spatial organization of three dimensional structure. *Proceedings of the Royal Society of London B*, 200:269-294.

Newell, F.N., & Bulthoff, H.H. (2002). Categorical perception of familiar objects. *Cognition, 85*, 113-143.

Snodgrass, J.G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory, 6* (2), 174-215.

SPSS for Windows, Rel. 11.0.1. 2001. Chicago: SPSS Inc.

Tarr, M., & Bülthoff, H. (1995). Is human object recognition better described by geon structural descriptions or by multiple views? Comment on Biederman and Gerhardstein (1993). Journal of Experimental Psychology: Human Perception & Performance , 21, 1494-1505.

Tarr, M. J.; Williams, P.; Hayward, W. G.; Gauthier, I. (1998) Three-dimensional object recognition is viewpoint dependent. Nature Neuroscience; 1:275-277.