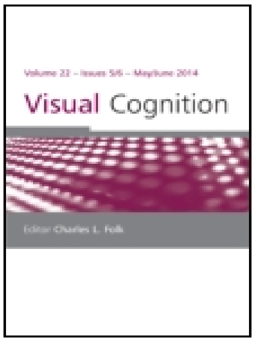

Visual Cognition

Routledge
Taylor & Francis Group

ISSN: 1350-6285 (Print) 1464-0716 (Online) Journal homepage: http://www.tandfonline.com/loi/pvis20

Visual categorization of social interactions

Stephan de la Rosa, Rabia N. Choudhery, Cristóbal Curio, Shimon Ullman, Liav Assif & Heinrich H. Bülthoff

To cite this article: Stephan de la Rosa, Rabia N. Choudhery, Cristóbal Curio, Shimon Ullman, Liav Assif & Heinrich H. Bülthoff (2014) Visual categorization of social interactions, Visual Cognition, 22:9-10, 1233-1271, DOI: 10.1080/13506285.2014.991368

To link to this article: http://dx.doi.org/10.1080/13506285.2014.991368

Published online: 06 Feb 2015.

Visual Cognition, 2014
Vol. 22, Nos. 9–10, 1233–1271, http://dx.doi.org/10.1080/13506285.2014.991368

Visual categorization of social interactions

Stephan de la Rosa[1], Rabia N. Choudhery[1], Cristóbal Curio[1], Shimon Ullman[2], Liav Assif[2], and Heinrich H. Bülthoff[1,3]

[1]Department of Human Perception, Cognition and Action, Max Planck Institute for Biological Cybernetics, Tübingen, Germany
[2]Department of Computer Science and Applied Mathematics, The Weizmann Institute of Science, Rehovot, Israel
[3]Department of Brain and Cognitive Engineering, Korea University, Seoul, South Korea

(*Received 27 July 2012; accepted 19 November 2014*)

Prominent theories of action recognition suggest that during the recognition of actions the physical patterns of the action is associated with only one action interpretation (e.g., a person waving his arm is recognized as waving). In contrast to this view, studies examining the visual categorization of objects show that objects are recognized in multiple ways (e.g., a VW Beetle can be recognized as a car or a beetle) and that categorization performance is based on the visual and motor movement similarity between objects. Here, we studied whether we find evidence for multiple levels of categorization for social interactions (physical interactions with another person, e.g., handshakes). To do so, we compared visual categorization of objects and social interactions (Experiments 1 and 2) in a grouping task and assessed the usefulness of motor and visual cues (Experiments 3, 4, and 5) for object and social interaction categorization. Additionally, we measured recognition performance associated with recognizing objects and social interactions at different categorization levels (Experiment 6). We found that basic level object categories were associated with a clear recognition advantage compared to subordinate recognition but basic level social

Please address all correspondence to Stephan de la Rosa or Heinrich H. Bülthoff, Department of Human Perception, Cognition and Action, Max Planck Institute for Biological Cybernetics, Spemannstrasse 38, 72076 Tübingen, Germany. E-mail: delarosa@tuebingen.mpg.de; or heinrich.buelthoff@tuebingen.mpg.de

We thank Betty Mohler, Stephan Streuber, Kathrin Kaulard, Johannes Schultz, and Julia Gatzek, for their help in creating some of the stimuli. We thank the anonymous reviewers for their helpful comments. This research was supported by the EU Project "TANGO" [ICT-2009-C 249858] and by the WCU (World Class University) program funded by the Ministry of Education, Science and Technology through the National Research Foundation of Korea [R31-10008] awarded to HHB. SU, LA were supported by ERC Advanced [grant number 269627] Digital Baby.

interaction categories provided only a little recognition advantage. Moreover, basic level object categories were more strongly associated with similar visual and motor cues than basic level social interaction categories. The results suggest that cognitive categories underlying the recognition of objects and social interactions are associated with different performances. These results are in line with the idea that the same action can be associated with several action interpretations (e.g., a person waving his arm can be recognized as waving or greeting).

Keywords: Visual categorization; Basic level; Sub-ordinate level; Objects; Social interactions.

Humans are social beings and the interaction with other humans is an integral part of human life. Here, we define social interactions as human non-verbal physical actions directed towards another human, e.g., when two people are kissing. The visual recognition of social interactions informs the observer about the ongoing social events in the social environment. This information is important for selecting appropriate action responses for the observed social environment (e.g., avoiding a street in which people are fighting). The visual recognition of social interactions is, therefore, critical for humans to navigate through both their physical and social environment. So far, prominent theories of action recognition assume that during action recognition the physical pattern of an action is associated with one particular action interpretation (e.g., Fleischer, Caggiano, Thier, & Giese, 2013; Giese & Poggio, 2003; Rizzolatti, Fogassi, & Gallese, 2001). Here we examine whether humans associate different semantic interpretations with the same action.

Visual recognition can be described as the process of mapping visual information onto semantic knowledge (Freedman & Miller, 2008). This process associates sensory information with meaning and provides knowledge about the observed stimulus (Freedman & Miller, 2008; Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976). Importantly, research on object recognition suggests that sensory information can be mapped onto semantic knowledge at different levels of abstraction. Rosch and colleagues (Rosch et al., 1976) suggested three levels of abstraction for object recognition: the sub-ordinate (e.g., recognizing an object as a German Shepherd), basic (e.g., recognizing an object as a dog), and superordinate level (e.g., recognizing an object as an animal) (see also Jolicoeur, Gluck, & Kosslyn, 1984). The process of mapping of visual information onto a particular level of abstraction is also referred to as visual categorization.

Rosch suggested that the formation of abstraction levels is based on cue category validity. Cue category validity (hereafter simply referred to cue validity) refers to the probability of a cue being a unique identifier for objects of a category on a given level of abstraction (e.g., motor movement cues such as the physical movement of petting a dog) (Corter & Gluck, 1992; Rosch et al., 1976). For a cue to be unique for all objects of a category (and hence cue validity to be high), objects of the same category need to share as many cues as possible while objects of contrasting categories should share as few cues as possible. As for object recognition, cues include shape and motor movements associated with the interaction of an object (Rosch et al., 1976). Rosch and colleagues (1976) showed that the cue validity of motor and shape cues is highest for objects at the basic level and lower for other levels of abstraction. For example, the cue validity at the sub-ordinate level (e.g., "German Shepherd") is lower because objects share many cues with other sub-ordinate categories (e.g., "Poodle") at this abstraction level. Only objects at the basic level share many cues with other objects of the same category (e.g., "dog") while at the same time these cues are dissimilar from objects of other contrasting basic level categories (e.g., "bird"). Several pieces of evidence support the existence of three categorization levels. However, visual object categorization performance is subject to learning as evidenced by the influence of experience on visual recognition performance (Tanaka & Taylor, 1991) and depends on the typicality of the recognized object (e.g., penguin is an atypical example for a bird; Jolicoeur et al., 1984).

Little is known about the abstraction levels of visual categorization beyond object categorization (Macé, Joubert, Nespoulous, & Fabre-Thorpe, 2009; Rosch et al., 1976; Tanaka & Taylor, 1991; VanRullen & Thorpe, 2001) and scene categorization (Boucart, Moroni, Thibaut, Szaffarczyk, & Greene, 2013; Fabre-Thorpe, Delorme, Marlot, & Thorpe, 2001; Thorpe, Fize, & Marlot, 1996; Thorpe, Gegenfurtner, Fabre-Thorpe, & Bülthoff, 2001). However, examining visual categorization outside of object categorization is important to gain a better understanding about the processes underlying visual categorization. Specifically, one can assess the universality of the known properties of visual categorization by comparing object categorization with the categorization of other types of stimuli. Moreover, the examination of visual categorization in social interaction recognition makes important contributions to action recognition research. Current prominent theories of action recognition have not considered the possibility that humans might be able to recognize actions at several levels, e.g., recognizing a handshake as a handshake or a greeting (Giese & Poggio, 2003; Rizzolatti et al., 2001). Examining whether different abstraction levels exist for action recognition is important to more fully understand the cognitive architecture supporting action recognition.

Hierarchical aspects of the cognitive structure underlying action recognition has received little attention so far. Previous research has been concerned with the human ability to distinguish between several actions along the temporal domain (action segmentation) or the examination of biological motion cues for the recognition of actions. These investigations have shown that humans use statistical visual regularities in the movement patterns of actions to discriminate actions within an action sequence (Baldwin, Andersson, Saffran, & Meyer, 2008). Other research has shown that humans use visual information from the movement of the human body to gather socially relevant information including telling friends apart from strangers (Loula, Prasad, Harber, & Shiffrar, 2005), recognizing a person's emotions (Roether, Omlor, Christensen, & Giese, 2009), and recognizing the identity of a person (Troje, Westhoff, & Lavrov, 2005) (for a more comprehensive review see Blake & Shiffrar, 2007). With regards to social interactions, humans are able to recognize social interactions from stimuli deprived of structural information (Dittrich, 1993) and the viewpoint of social interaction seems to be critical for the recognition of social interactions (de la Rosa, Mieskes, Bülthoff, & Curio, 2013). However, all of these studies have focused on the recognition of actions at only one level of abstraction. The examination of different cognitive categories of the same action, and in particular social interactions, has received little or no attention. This piece of information is important to understand how visual information about social interactions is associated with semantic knowledge.

In the present study, we were interested in examining the different recognition levels when viewing social interactions. Social interactions are prevalent in the human environment and their visual understanding allows humans to interact better with their physical and social environment. We were interested in participants' categorization behaviour of static social interactions, i.e., snapshots of social interactions. We examined the existence of different abstraction levels for social interactions and the cues that might promote the emergence of social interaction categories. In particular, we examined participants' grouping behaviour of social interactions to examine possible abstraction levels of visual social interaction recognition. Moreover, we investigated the similarity of visual (e.g., shape) and motor movements across different social interactions to assess the potential of these cues in explaining the emergence of social interaction categories. Finally, we assessed whether the found abstraction levels correlate with different recognition performance.

The study consists of six experiments. The purpose of Experiments 1 and 2 was to explore how humans visually categorize social interactions in order to shed light onto possible social interaction abstraction levels. Hierarchical cluster analysis was used to examine different levels of social interaction categorization. Three additional experiments examined the perceived similarity of motor movements (Experiment 3) and visual appearance (Experiments 4 and 5) of social interactions at different levels of cluster hierarchy. The examination of visual and motor movement similarity at different levels of the cluster hierarchy provides insights into the extent to which these cues drive social interaction categorization. Finally, we examined whether recognition performance is dependent on the abstraction level upon which the social interaction is recognized (Experiment 6).

We used static images of social interactions to investigate the cognitive hierarchy of social interaction recognition. Static images of social interactions occur manifold in the human environment, e.g., newspapers, photos, etc. The choice of static images over movies of social interactions was motivated to minimize the influence of other factors such as the temporal duration or motion energy on action categorization. At the same time, the recognition of static images still requires the mapping of visual social interaction information onto the semantic social interaction knowledge. For example, static keyframe information about an action is considered to be a constituent component for the recognition of dynamic actions (Barraclough, Ingham, & Page, 2012; Giese & Poggio, 2003). Visual categorization of static social interactions should therefore be appropriate to examine the cognitive hierarchy of social interaction recognition and be a seed for further research on dynamic social action recognition.

We also probed object categorization for two reasons. Firstly, we wanted to validate the experimental methods used in this study by replicating previously reported results about visual object categorization. Secondly, the results of visual object categorization tasks also served as a guideline for the interpretation of the social interaction categorization results. For example, comparing the similarity of motor movements across different levels of abstractions for object and social interaction categorization is useful to examine whether motor movements can explain social interaction categorization to the same degree as object categorization.

EXPERIMENT 1

Experiment 1 examined visual categorization of objects and social interactions by means of an image grouping task in order to shed light onto possible abstraction levels of object and social interaction recognition. According to the principle of cue validity, items of the same basic level category should be more similar than to items of other basic level categories. To reveal possible basic categories we instructed participants to group images of social interactions into self-defined groups in such a way that items within a given social interaction group were more similar to each other than to items of contrasting social interaction groups. We then used hierarchical cluster analysis to visualize the grouping results. Additionally, we asked participants to give each image and self created group a name to explore the naming of these groups. The main purpose of Experiment 1 was to explore the cognitive structure underlying social interaction categorization.

To our knowledge, little is known about the visual categorization of social interactions and the ability of a visual grouping task to reflect cognitive categories underlying visual recognition. To assess how well cognitive categories are captured by a visual grouping task, we also conducted a visual object grouping task for which the categorization levels are relatively well established.

Method

Participants

Fifteen participants from the community of Tübingen (age range: 20–28 years) gave their written informed consent prior to the experiment and received €8/hour for their participation. All participants were naïve with regards to the research question. All had corrected or corrected-to-normal vision. The experiment was approved by the ethics committee of the University of Tübingen and was conducted in accordance with the Declaration of Helsinki.

Apparatus and stimuli

Stimuli were presented on a LCD Monitor with a refresh rate of 60 Hz using the PICASA™ software. All images were static grey-scale images (300 × 300 pixels) sampled from the Internet. We chose 12 types of objects (e.g., Poodle) and 12 types of social interactions (e.g., waving). See Figure 1 for one example image for each of the 12 object and social interaction types. We sampled 10 images for each type for a total of 240 images. One author chose the specific social interaction and object types. The selection of social interaction types was based on pilot studies using a large set of social interaction images that were grouped by other participants. The choice of object types was guided by previous reports of sub-ordinate object levels that form a basic object level (Grill-Spector & Kanwisher, 2005; de la Rosa, Choudhery, & Chatziastros, 2011).

Procedure

Participants were given the following information. For the grouping task, participants grouped 120 images (object or social interaction images depending on the condition) from a folder labelled “image basket” from within PICASA®. To do so participants created and labelled new folders, and then tallied images from the image basket into them. There were no restrictions with respect to the number or labelling of the folders except that folder names should describe the folder content. Participants were given the instruction that images within a folder should be more similar to each other than to images of another folder. For the naming task, participants choose a name for each object image that described the displayed object (in the object naming task). Likewise participants labelled each social interaction image with a name that described the displayed social interaction (in the social interaction naming task). The naming was supposed to be as precise as possible using only a single word. Participants were allowed to use the same one-word name for several images. Furthermore, the image names were allowed to be identical to the folder name if desired. Participants were free to choose the order of executing the grouping and the naming task. No time restrictions were given and the testing order of stimulus type (object, social interaction) was counterbalanced across participants.

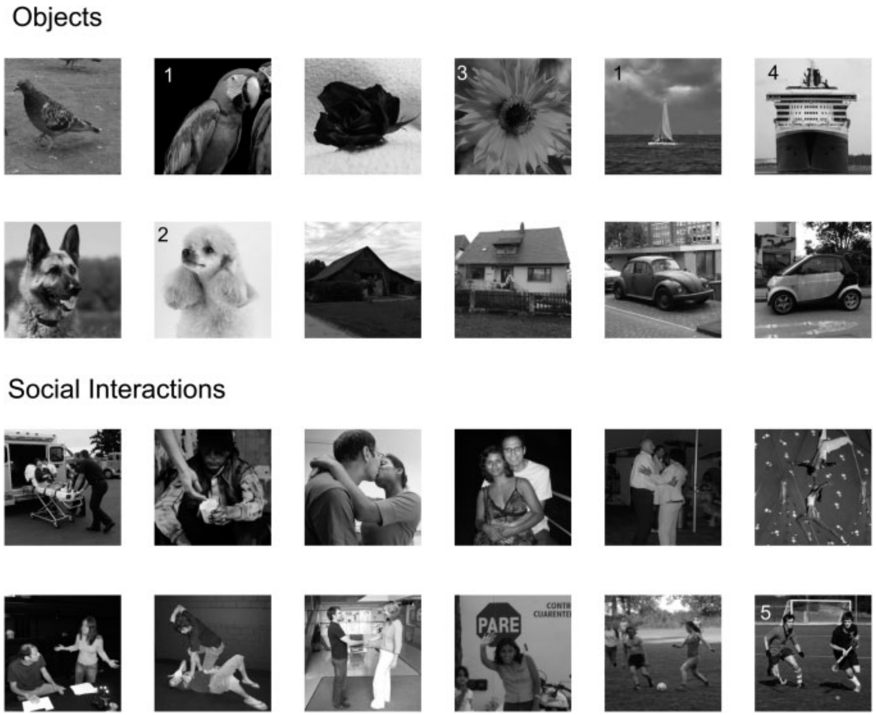

Figure 1. Example stimuli. Examples for each of the 12 objects and social interactions types employed in Experiment 1. Numbers within an image acknowledge the photographer ([1]Arvind Balaraman, [2]Luigi Diamanti, [3]Susie B, [4]Bernie Condon, [5]Tina Phillips; all at www.freedigitalphotos.net).

Results and discussion

We analyzed and present the results of the object categorization and social interaction categorization tasks separately.

Objects

Image naming. We first analyzed the one-word naming of object images for each object image separately in order to derive names that can be used to refer to object images. Table 1 lists the most frequent image description along with its relative frequency (in percent) for each object image separately. Overall, there seems to be good agreement between participants with regards to the naming of object images. On average 67.12% of the participants used the same image description for a given object image (min = 40%; max = 86.67%; *SD* = 14.51%). Images showing the same object type (e.g., VW Beetle) were referred to as with the same one-word description (Table 1). The single word descriptions are in good agreement with previous reports of sub-ordinate object levels.

Grouping task. We summarized the results of the grouping task in a confusion matrix which measured how often (in percentage) one image was put into the same group with another image across all participants. The results are shown in Figure 2. Grouping occurred for some image pairings but not for others. In particular images displaying the same object type were most often grouped as indicated by white squares along the diagonal. In particular images of German Shepherds, Poodles, VW Beetle, Smart cars, houses, barns/shacks, roses, sunflowers, sail boats, cruise ships, pigeons, and parrots were grouped. Furthermore, bright off-diagonal squares indicate frequent grouping of German Shepherds and Poodles, VW Beetles and Smart cars, houses and barns/shacks, roses and sunflowers, sail boats and cruise ships, and pigeons and parrots. Participants less frequently grouped German Shepherds, Poodles, pigeons, and parrots. Apart from these grouping patterns little grouping occurred.

To more formally assess participants' object image grouping behaviour we conducted a hierarchical cluster analysis. The purpose of a hierarchical cluster analysis is to identify subsets of similar items (groups) in a set of items. Here, it served the additional purpose of providing a statistical evaluation of the observed patterns in Figure 2. We used the inverse of the relative frequency of the above confusion matrix (i.e., 100-relative frequency) as a distance measure for the cluster analysis and treated them as correlational distances. Clusters were formed based on Ward's method. In short, Ward's algorithm clusters any two items that have minimal error sums of squares (ESS) to their centroid (Ward, 1963). Because ESS measures the sum of squared distances between cluster items and their centroid, it is a measure of cluster item closeness. Items of a cluster with a small ESS are closer to each other than items of a cluster with a larger ESS. We used the function *pvclust* of the statistical package R to calculate the

TABLE 1
Most frequent object descriptions listed for each object image separately

Image ID	*Description*	*Percent*	*Image ID*	*Description*	*Percent*	*Image ID*	*Description*	*Percent*
1	Parrot	66.67	41	Smart	80.00	81	Rose	86.67
2	Parrot	73.33	42	Smart	80.00	82	Rose	85.71
3	Parrot	53.33	43	Smart	80.00	83	Rose	86.67
4	Parrot	80.00	44	Smart	80.00	84	Rose	86.67
5	Parrot	66.67	45	Smart	80.00	85	Rose	86.67
6	Parrot	66.67	46	Smart	80.00	86	Rose	86.67
7	Parrot	66.67	47	Smart	80.00	87	Rose	86.67
8	Parrot	66.67	48	Smart	80.00	88	Rose	86.67
9	Parrot	73.33	49	Smart	80.00	89	Rose	86.67
10	Parrot	66.67	50	Smart	80.00	90	Rose	84.62
11	Pigeon	80.00	51	Beetle	46.67	91	Sunflower	80.00
12	Pigeon	80.00	52	Beetle	42.86	92	Sunflower	73.33
13	Pigeon	80.00	53	Beetle	53.33	93	Sunflower	73.33
14	Pigeon	80.00	54	Beetle	46.67	94	Sunflower	80.00
15	Pigeon	80.00	55	Beetle	46.67	95	Sunflower	73.33
16	Pigeon	80.00	56	Beetle	46.67	96	Sunflower	80.00
17	Pigeon	78.57	57	Beetle	46.67	97	Sunflower	73.33
18	Pigeon	78.57	58	Beetle	46.67	98	Sunflower	80.00
19	Pigeon	80.00	59	Beetle	46.67	99	Sunflower	86.67
20	Pigeon	78.57	60	Beetle	46.67	100	Sunflower	80.00
21	Cruise ship	53.33	61	German shepherd	71.43	101	Shack	53.33
22	Cruise ship	53.33	62	German shepherd	73.33	102	Barn	40.00
23	Cruise ship	53.33	63	German shepherd	73.33	103	Shack	46.67
24	Cruise ship	53.33	64	German shepherd	73.33	104	Shack	53.33
25	Cruise ship	53.33	65	German shepherd	73.33	105	Shack	40.00
26	Cruise ship	53.33	66	German shepherd	73.33	106	Shack	40.00
27	Cruise ship	50.00	67	German shepherd	73.33	107	Shack	40.00
28	Cruise ship	53.33	68	German shepherd	66.67	108	Shack	40.00
29	Cruise ship	53.33	69	German shepherd	73.33	109	Barn	40.00
30	Cruise ship	53.33	70	German shepherd	53.33	110	Barn	40.00
31	Sailboat	66.67	71	Poodle	80.00	111	House	53.33
32	Sailboat	66.67	72	Poodle	80.00	112	House	53.33
33	Sailboat	66.67	73	Poodle	80.00	113	House	53.33
34	Sailboat	66.67	74	Poodle	80.00	114	House	53.33
35	Sailboat	66.67	75	Poodle	80.00	115	House	46.67
36	Sailboat	66.67	76	Poodle	80.00	116	House	46.67
37	Sailboat	66.67	77	Poodle	80.00	117	House	53.33
38	Sailboat	66.67	78	Poodle	80.00	118	House	57.14
39	Sailboat	66.67	79	Poodle	80.00	119	House	60.00
40	Sailboat	66.67	80	Poodle	80.00	120	House	53.33

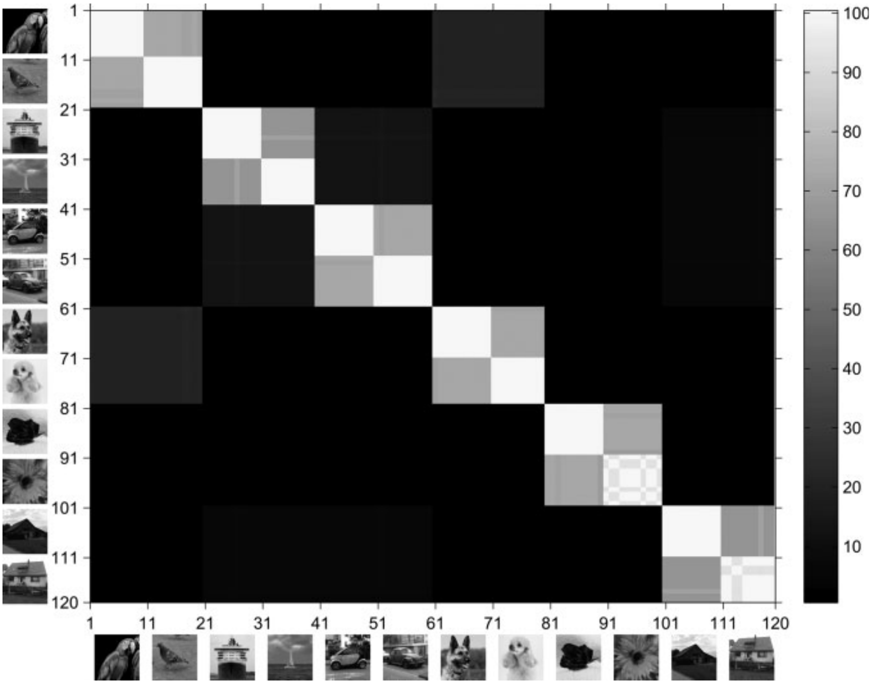

Figure 2. Confusion matrix of the object grouping task. The confusion matrix shows the relative frequency with which participants grouped different object images. Each index of vertical and horizontal axis refers to a different object image (for index-image assignment see Table 1). Images of the same object type are placed adjacent to each other. For sake of clarity, one example image of each object type is shown along the two axes. Relative frequency is grey-scale coded (see legend). Higher grey values indicate higher relative frequency.

Approximately Unbiased (AU) *p*-value for each cluster (10,000 bootstrap replications), which measures the probability of a cluster being a true cluster (Shimodaira, 2002). The *p*-value ranges between 0 and 1 with higher *p*-values suggesting that the cluster is more strongly supported by the data. We consider an AU *p*-value of equal or more than .95 as a cluster supported by the data. The results of the cluster analysis on the object data are shown in Figure 3.

Figure 3 shows 12 major clusters at the bottom of the cluster hierarchy that consists of images showing the same object type (e.g., poodles). We will refer to these 12 major clusters as first level clusters. These clusters have in common that the distance between items composing each cluster is perceived as very small as indicated by small ESS (mean ESS = 0.0002, *SD* = 0.0009). Moreover each of these 12 clusters is well supported by the data as indicated by large AU *p*-values. The following objects were put into separate clusters: Poodle, German Shepherd, parrot, pigeon, rose, sunflower, barn/shack, house, Smart car, Beetle, sailboat, cruise ship.

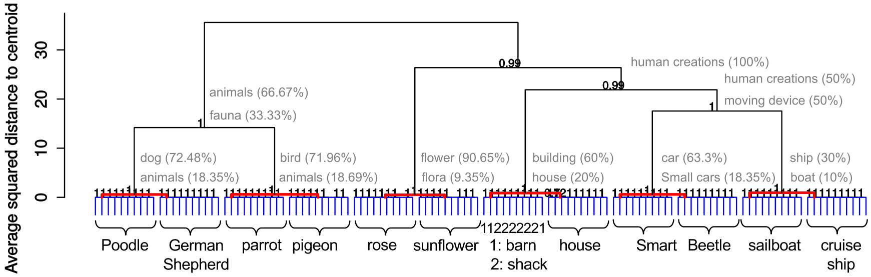

Figure 3. Dendrogram of the object cluster analysis. For sake of clarity we collapsed object images names that have the same name along the x-axis. The y-axis shows the ESS. The black numbers indicate the Approximately Unbiased (AU) probability. First level clusters are marked in blue and second level cluster are marked in red. The light grey to the right of the vertical lines text gives the two most frequent group names that corresponded to the cluster along with its relative frequency (in%).

Figure 3 further shows that any two first level clusters form a higher level, i.e., more general, cluster (hereafter referred to as second level cluster). In particular, roses and sunflowers, barns/shacks and houses, Smart cars and Beetles, sailboats and cruise ships form separate second level clusters. Items within these clusters are still perceived as similar as indicated by intermediate ESS (mean ESS = 0.692; *SD* = 0.180). Moreover the existence of these clusters is well supported by the data as suggested by large AU *p*-values ($p \geq .95$).

Finally, there is also evidence for higher order clusters as indicated by a large AU *p*-value. Because we only reliably identify first and second level clusters in the social interactions task (see below), cluster levels above the second level cluster are not further discussed here.

Group naming. We were interested in the names associated with second level clusters of Figure 3. We counted the frequency of names given to groups that corresponded to these clusters. These names and their relative frequency are shown in grey in Figure 3. These names are in good agreement with previous reports of basic object level names.

Overall, the image and group naming task are in good agreement with previously reported sub-ordinate and basic object levels, respectively. Specifically, the composition and naming of first level clusters is similar to sub-ordinate object levels while the composition and naming of second levels clusters has much in common with basic object levels. Hence, the visual categorization and naming task of objects reveals grouping and naming patterns in accordance with previously reported sub-ordinate and basic object levels and seems to be an appropriate method for assessing cognitive categories.

Social interactions

Image naming. Table 2 shows the most frequent one-word descriptions along with its relative frequency for each social interaction image separately.

There is agreement across participants with regards to their one-word social interaction descriptions (mean relative frequency: 46.21%, maximum: 80%, minimum: 20%, *SD* = 14.52). Some social interactions, e.g., dancing or begging, are referred to with more similar one-word descriptions than others, e.g., arguing. Overall, similar social interactions were described in a similar manner by different participants.

Grouping task. Participants' social interaction grouping behaviour is shown by means of a confusion matrix in Figure 4. The bright squares along the diagonal suggest that participants most frequently grouped social interactions of the same type (e.g., handshake). Furthermore, the lighter grey squares off the diagonal suggest that participants frequently grouped fighting and arguing, waving and handshaking, and soccer, hockey, and acrobatics. Participants less frequently grouped hug and kisses with each other and dance, soccer, hockey and acrobatics with each other.

We analyzed the grouping pattern more formally by means of a hierarchical cluster analysis (for details see the section "Objects" above). Figure 5 shows the dendrogram of the cluster analysis. Twelve major clusters emerge at the bottom of the dendrogram (first level clusters). The first level clusters possess small ESS (mean = 0.0024; *SD* = 0.0071) suggesting that items within these clusters are very close. Moreover, high AU *p*-values indicate that the 12 clusters are well supported by the data. Each of the 12 clusters refers to images of one of the following social interactions (names are as in the naming task): acrobatics, hockey, soccer, arguing, fight, waving, shaking hands, rescue, begging, dance/dancing, kiss, and hugs.

The next higher level of the cluster hierarchy (second level clusters) is formed from two first level clusters. All of the second level clusters are well supported by the data as indicated by large AU *p*-values ($p \geq .95$). These clusters are more variable with respect to how similar items are perceived within each cluster (indicated by ESS) compared to the first level clusters. Specifically, items within two second level clusters (clusters summarizing hockey and soccer images, and arguing and fight images) are perceived as relatively close (ESS values of 0.53 and 0.66, respectively). Three other clusters each summarizing two of the first level clusters are well supported by the data. Items within these clusters are perceived as more distant (waving-shaking hands cluster: ESS = 1.96, kiss-hug cluster: ESS = 3.51, rescue-begging cluster: ESS = 9.41).

We observed other higher levels (see Figure 5). They combine first level clusters with other second level clusters or combine several higher order clusters. These clusters are overall more heterogeneous than second level clusters as indicated by higher ESS values and AU *p*-values smaller than .99 and are not further discussed here.

Group naming. To determine the names corresponding to these clusters we counted the frequency of names of groups that corresponded to these second level clusters. Figure 5 shows the two most frequent group names in grey. The most frequent group names corresponding to the hockey-soccer cluster were

TABLE 2
Most frequent social interaction descriptions listed for each social interaction image separately

Image ID	*Description*	*Percent*	*Image ID*	*Description*	*Percent*	*Image ID*	*Description*	*Percent*
1	Hug	46.67	41	Arguing	26.67	81	Begging	66.67
2	Hug	46.67	42	Arguing	26.67	82	Begging	66.67
3	Hug	40.00	43	Arguing	26.67	83	Begging	66.67
4	Hug	40.00	44	Arguing	26.67	84	Begging	66.67
5	Hug	40.00	45	Arguing	26.67	85	Begging	66.67
6	Hug	46.67	46	Arguing	20.00	86	Begging	66.67
7	Hug	40.00	47	Arguing	26.67	87	Begging	66.67
8	Hug	46.67	48	Arguing	26.67	88	Begging	66.67
9	Hug	46.67	49	Arguing	26.67	89	Begging	60.00
10	Hug	46.67	50	Arguing	26.67	90	Begging	66.67
11	Kiss	53.33	51	Fight	33.33	91	Rescue	35.71
12	Kiss	66.67	52	Fight	33.33	92	Rescue	20.00
13	Kiss	66.67	53	Fight	33.33	93	Rescue	33.33
14	Kiss	66.67	54	Fight	33.33	94	Rescue	33.33
15	Kiss	60.00	55	Fight	33.33	95	Rescue	33.33
16	Kiss	66.67	56	Fight	40.00	96	Rescue	33.33
17	Kiss	66.67	57	Fight	33.33	97	Rescue	33.33
18	Kiss	66.67	58	Fight	33.33	98	Rescue	33.33
19	Kiss	60.00	59	Fight	33.33	99	Rescue	33.33
20	Kiss	60.00	60	Fight	33.33	100	Rescue	42.86
21	Acrobatics	42.86	61	Shaking hands	33.33	101	Hockey	40.00
22	Acrobatics	46.67	62	Shaking hands	33.33	102	Hockey	40.00
23	Acrobatics	40.00	63	Shaking hands	33.33	103	Hockey	40.00
24	Acrobatics	40.00	64	Shaking hands	33.33	104	Hockey	40.00
25	Acrobatics	40.00	65	Shaking hands	33.33	105	Hockey	40.00
26	Acrobatics	40.00	66	Shaking hands	33.33	106	Hockey	40.00
27	Acrobatics	46.67	67	Shaking hands	33.33	107	Hockey	40.00
28	Acrobatics	40.00	68	Shaking hands	33.33	108	Hockey	40.00
29	Acrobatics	46.67	69	Shaking hands	33.33	109	Hockey	40.00
30	Acrobatics	40.00	70	Shaking hands	33.33	110	Hockey	40.00
31	Dance/ dancing	66.66	71	Waving	57.14	111	Soccer	66.67
32	Dance	42.86	72	Waving	53.33	112	Soccer	53.33
33	Dance	33.33	73	Waving	46.67	113	Soccer	53.33
34	Dance	33.33	74	Waving	53.33	114	Soccer	66.67
35	Dance/ dancing	80.00	75	Waving	50.00	115	Soccer	66.67
36	Dance/ dancing	80.00	76	Waving	60.00	116	Soccer	66.67
37	Dance/ dancing	66.66	77	Waving	53.33	117	Soccer	66.67
38	Dance/ dancing	66.66	78	Waving	53.33	118	Soccer	53.33
39	Dance	40.00	79	Waving	60.00	119	Soccer	66.67
40	Dance	40.00	80	Waving	60.00	120	Soccer	66.67

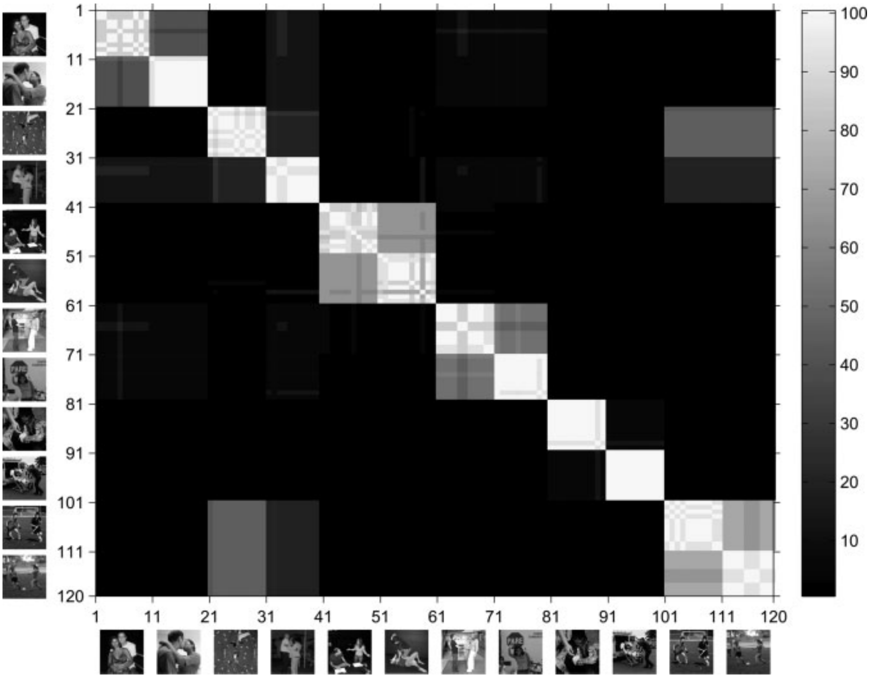

Figure 4. Confusion matrix of the social interaction grouping task. The confusion matrix shows the relative frequency with which participants grouped different social interaction images. Each index of vertical and horizontal axis refers to a different social interaction image (for index-image assignment see Table 2). Images of the same social interaction type are placed adjacent to each other. For sake of clarity one example image of each social interaction type is shown along the two axes. Relative frequency is grey-scale coded. Higher grey values indicate higher relative frequency.

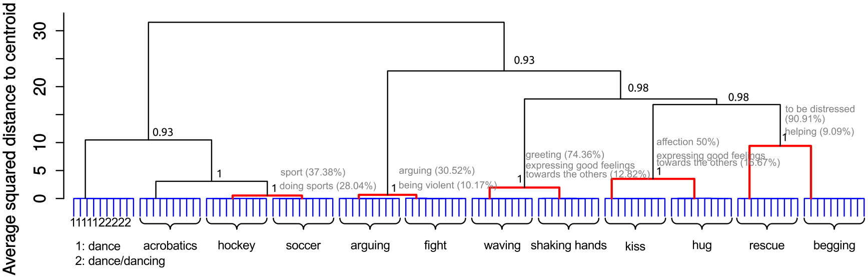

Figure 5. Dendrogram of the social interaction cluster analysis. For sake of clarity we collapsed social interaction images with the same image name along the x-axis. The y-axis shows the ESS. The black numbers indicate the Approximately Unbiased (AU) probability. First level clusters are marked in blue and second level cluster are marked in red. The light grey text to the right of the vertical line gives the two most frequent group names that corresponded to the cluster along with its relative frequency (in%).

"sports" and "doing sports". The names "arguing" and "being violent" were often used for groups that corresponded to the arguing-fight cluster. The most frequent group names corresponding to the waving-handshake cluster were "greeting" and "expressing good feelings towards others." For the kiss-hug cluster the most frequent group names were "affection" and "expressing good feeling towards others." Finally the most frequent group names for the rescue-beg cluster was "to be distressed" and "helping."

Taken together, participants grouped social interaction images. Specifically, the dendrogram shows two prominent levels of clusters. The first level of cluster consists of social interactions of the same type (e.g., handshakes) and interactions are very similar in terms of their ESS. The second level clusters are formed from two first level clusters and therefore combine different social interactions.

Discussion

Our results show that participants categorize both object images and social interaction images. In particular, two levels of clusters emerge for both stimulus types. The first cluster level consists of images which participants always describe in the same way. The second cluster level summarizes images that are associated with different names. The comparison of the two object cluster levels with previously reported object abstraction levels suggests that the first cluster level corresponds to the sub-ordinate level and the second cluster level to basic level. Because a similar pattern is observed for social interaction categorization, the results support the idea that different abstraction levels exist for social interaction categorization. In particular, images of the sport and the arguing clusters are perceived as very similar and are possible candidates for basic level social interaction clusters.

In this experiment, we measured the perceived similarity between higher order objects and social interactions clusters in a between-subject fashion: Similarities across participants' grouping behaviour for objects and social interactions were the constituent component for the similarity measures of higher order clusters. The replication of previously reported higher order object clusters (e.g., basic level like clusters) with this method gives some confidence that the between-subject similarities are indicative of higher order cognitive categories. However, we wanted to ensure that similar patterns are found when image similarity for higher order clusters is measured within the same participant. We employed a multilevel grouping paradigm, in which participants first grouped images into folders and then grouped these folders into higher-level folders. The grouping of folders into higher level folders was done until participants felt satisfied with the result.

EXPERIMENT 2

Experiment 2 replicated Experiment 1 using a multilevel grouping paradigm. Because the results of object categorization are well known, we only probed social interaction categorization in Experiment 2. Due to a technical error participants did not see the acrobatics images of Experiment 1. While acrobatics images formed a subordinate cluster, they did not form a basic level like cluster. Hence, this error should have little effect on the critical aspect of Experiment 2, namely, to investigate basic level like clusters in social interaction recognition.

Methods

The methods were identical to Experiment 1 except for the following.

Participants

Seventeen new participants (8 males, Age: mean = 26.5 years, *SD* = 4.77) recruited from the local community of Tübingen participated in the experiment.

Stimuli

We presented five images of each predefined social interaction category. The acrobatics category was not presented due to a technical error.

Procedure

We employed a multilevel grouping paradigm. Participants were instructed to group images into folders and then subsequently group these folders into other newly created folders according to their perceived similarity. Participants did not name the images.

Results

Figure 6 shows the results of the cluster analysis. This plot shows a striking similarity with the first two cluster levels of Figure 5. Specifically, we found that field hockey, soccer, dancing, kiss, hug, handshake, waving, fight, arguing, rescue, and begging images constitute separate first level clusters. From these first level clusters the following five second level clusters emerged: field hockey and soccer, kiss and hug, handshake and waving, fight and arguing, and rescue and begging. The findings are identical to the social interaction grouping results of Experiment 1 (see Figure 5). We statistically compared the cluster results of Experiment 1 and Experiment 2 using those social interactions that were common to both experiments. We measured the cluster results similarities with two frequently used indices for cluster result comparisons, namely the adjusted Rand index (ARI) and the Jaccard index. Both indices vary between 0 and 1,

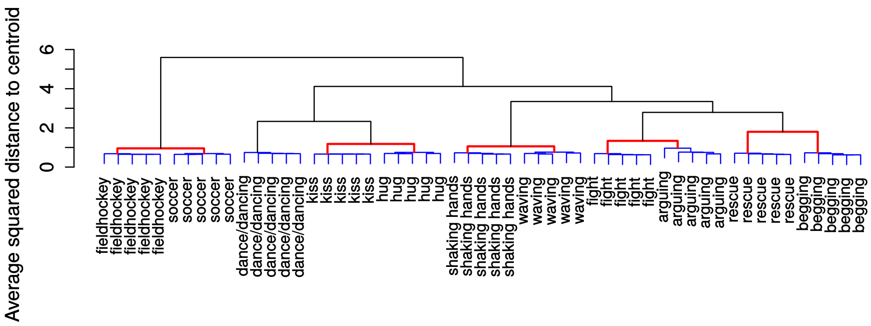

Figure 6. Dendrogram of the social interaction cluster analysis of Experiment 2. First level clusters are marked in blue and second level cluster are marked in red.

where higher values indicate larger similarity between the cluster results. We obtained a value of 0.85 for the adjusted Rand index and a value of 0.76 for the Jaccard index when comparing the first two clusters that were common to Experiments 1 and 2. These values indicate a very good agreement between the first two cluster levels of Experiments 1 and 2. Hence, we find very similar first and second level categories using a multilevel grouping approach compared to the between subjects approach (Experiment 1).

What cues guide the categorization of social interaction images? Likely candidates are motor and visual cues that have been identified as being important for object categorization (Rosch, 1978). In particular, objects at the basic and sub-ordinate level are associated with similar motor movements. Consequently, we expected that object images of the first and second level of the cluster hierarchy are associated with similar motor movements. In Experiment 3, we examined whether a similar finding holds for social interaction categorization.

EXPERIMENT 3

The purpose of Experiment 3 was to examine whether motor movements associated with objects and social interactions are similar for first and second level cluster levels. We used a rating task to determine motor movement similarity across different levels of categorization. We reasoned that higher similarity ratings of two different limb movements suggest that the two movements are more similar and consequently share more movement cues, which in turn indicates higher cue validity. We compared the similarity of motor movements associated with objects and social interactions between first and second cluster levels. We decided to collect movement similarity ratings based on words rather than images in order to minimize the effect that other visual cues (e.g., overall similarity of the displayed scene) might have on similarity ratings.

The experiment consisted of two parts. In the first part participants saw images (showing either objects or social interactions) and described a typical motor movement associated with each image. For objects, participants were instructed to report a typical movement when one would interact with the displayed object. For social interactions, participants were instructed to report a typical movement when one would engage in the displayed social interaction. In the second part of the experiment, participants were presented with pairs of motor movement descriptions and rated the physical similarity of each pair. Specifically, to assess whether movements are similar for first and second cluster levels, we composed pairs of motor movement descriptions in two different ways. We sampled motor descriptions from images of the same first level cluster to measure first level cluster motor movement similarity. Then, we sampled images from two different first level clusters (e.g., soccer and hockey) that formed a second level cluster to measure second level cluster motor movement similarity.

We compared first and second cluster movement similarity ratings to determine whether motor movements of second level clusters are significantly different from the motor movements of first level clusters. As for objects, we expected to replicate previous results, namely that first and second level cluster movement similarity ratings are similar (Rosch et al., 1976). As for social interactions, we were interested in whether a similar pattern emerges for social interactions.

Methods

Participants

Sixteen participants participated in the experiment (eight females, mean age: 28.31, *SD* = 6.32). All participants had normal or corrected-to-normal vision and gave their written informed consent prior to the experiment. Participants received €8/hour as compensation for their participation in the experiment. The experiment was approved by the ethics committee of the University of Tübingen and was conducted in accordance with the Declaration of Helsinki.

Apparatus and stimuli

We used the same computer and stimulus set as described in Experiment 1. The assessment of first level cluster movement similarity was based on movement descriptions for images that were sampled from the same first level cluster. We, therefore, randomly drew pairs of images from all first level clusters that formed second level clusters. For objects there were 12 first level clusters (see Figure 3; 2 × 12 = 24 images) and for social interactions there were 10 first level clusters that formed second level clusters (see Figure 5; 2 × 10 = 20 images). The analysis of second level cluster movement similarity was based on movement descriptions for images that were sampled from different first level clusters but were part of the same second level cluster. Hence, we sampled one image from each of the two first level clusters that formed a second level cluster to form pairs of images (see Figures 3 and 5). For example, we randomly sampled one image from the soccer cluster and one image from the field hockey cluster for the assessment of the between cluster similarity for sports. Because there were six object clusters (12 images) and five social interaction (10 images) second level clusters, 22 images were sampled for the assessment of between cluster similarity. Hence, participants saw a total of 66 images (36 object images and 30 social interaction images). These images were presented to participants using an Internet browser (custom written html and php code), which also served to record the movement descriptions (using text boxes) of the images and the movement similarity ratings (seven point scale using radio buttons).

Procedure

The experiment consisted of two parts. In brief, participants saw 66 images and described a typical movement that was associated with the displayed image content in the first part. In the second part, participants rated pairs of physical movement descriptions in terms of their physical similarity. More specifically, in the first part of the experiment the 36 object images and 30 social interaction images were displayed in random order right below each other (separated by horizontal grey bars). A text box was located to the right of each image. Participants were instructed to look at the images one at a time and (for each image) to name a typical physical movement when one would interact with the image content. In particular, as for objects participants were asked to name a typical movement when one would interact with the displayed object. As for social interactions, participants were asked to describe a typical physical movement when being involved in executing the displayed social interaction. Participants wrote down the typical physical movement associated with the displayed image content in the text box right next to the image. Participants had to fill out the text boxes of all 66 images to continue to the second part of the experiment by means of a button press. At the beginning of the second part, participants called the experimenter to receive the instructions for the second part. In the second part of the experiment (a newly loaded page in the web browser), participants rated pairs of their own movement descriptions (i.e., not that of others) in terms of their physical similarity. Participants were not told that they were to rate their own movement descriptions. To do so, the question “How similar are the following two movements in terms of their physical similarity?” was presented with a seven point rating (Likert) scale below the descriptions. The left end of the scale was labelled “not at all” and the right end was labelled “completely.” The two physical movement descriptions were presented to the left and right of the rating scale. Participants made 33 movement similarity ratings that were presented right below each other and were separated by a grey horizontal bar. 12 of 33 ratings contrasted motor descriptions within first level object clusters (within cluster object similarity ratings), 10 ratings contrasted motor descriptions within first level social interaction clusters (within cluster social interaction similarity ratings). Eleven ratings contrasted motor descriptions from two different first level clusters that formed a second level cluster (six object and five social interaction second level clusters). The latter ratings were used to assess second level cluster movement similarity. Participants were verbally informed only about the task but not about the logic of the experiment by the experimenter prior to each part of the experiment. The entire experiment took approximately 30–45 minutes to complete.

Results and discussion

To compare the cue validity of motor cues for the categorization of objects and social interactions on the second level, we compared first and second level cluster motor movement similarity ratings for objects and social interactions. The mean ratings for first and second cluster levels movements are shown in Figure 7 for objects and social interactions separately. The first and second level cluster movements' similarity for objects seems comparable. In contrast, social interaction movements appear less similar on the second level compared to the first level. To test the interaction between cluster level and stimulus type, we calculated the difference scores between first and second level clusters for objects and social interactions separately. We then compared these two sets of

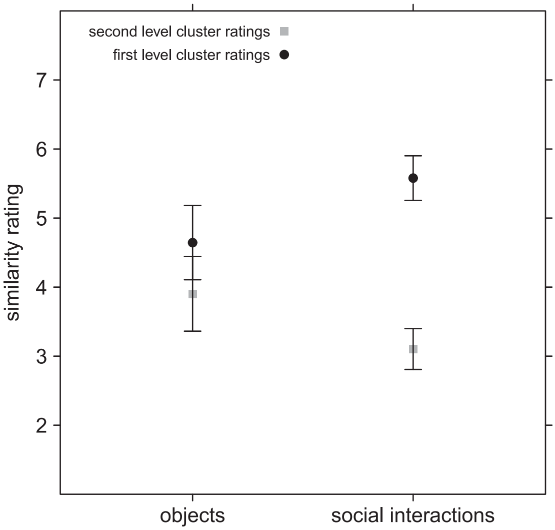

Figure 7. Movement similarity ratings of Experiment 3. Average movement similarity ratings of Experiment 2 shown for objects and social interactions (along the x-axis), and for first (circles) and second (squares) level clusters separately. Higher similarity ratings indicate higher perceived movement similarities. Bars indicate one standard error from the mean.

difference scores using a dependent two samples test. Because the data deviated from the normal distribution, we used a non-parametric Wilcoxon matched-pairs test to examine the statistical significance of the observed differences. The interaction of stimulus type (objects vs. social interactions) and cluster level (first vs. second level) on similarity ratings was significant, $T = 29$, $p = .044$. For objects, we found no significant difference between first and second level cluster movement similarity ratings, $T = 88.5$, $p = .110$. These results replicate previous reports about movements associated with the interaction of an object being similar for objects of the same basic level (Rosch et al., 1976). In contrast, first and second level cluster ratings were significantly different for social interactions, $T = 136$, $p < .001$, suggesting that second level social interaction clusters consists of social interactions with significantly different physical motor movements.

In sum, we replicate previous findings and show that second level object clusters consist of objects that evoke similar motor movements. This result is in line with the idea that motor movements are reliable cues for object categorization. In contrast, second level social interaction clusters constitute social interactions that are associated with significantly different movements. Hence, participants group social interactions that are associated with significantly different motor movements at the second level of the cluster hierarchy.

Visual similarity has been considered another important cue in the emergence of visual object categories. In Experiment 4, we compared visual similarity of objects and social interactions between the first and second cluster levels to assess the usefulness of this cue for object and social interaction recognition.

EXPERIMENT 4

Methods

Participants

Fourteen participants (4 females, mean age: 34) from the Max Planck Institute for Biological Cybernetics, Tübingen, Germany, volunteered for this experiment. All participants were naïve with regards to the research question. All participants had normal or corrected-to-normal vision and gave their consent prior to the experiment.

Apparatus and stimuli

The stimuli were the same stimuli as in Experiment 2. We used a laptop computer with a 15 inch screen and resolution of 1366 × 768 pixels for displaying the images. Stimuli were displayed on an Ubuntu 14.10 system using the nautilus browser with the thumbnail preview turned on, which resulted in displaying all images in a folder. Stimulus pairs (see Procedure) were put into separate folders and for each trial the contents of another folder were displayed. Stimuli had an approximate size of 7 × 7 cm on the screen.

Procedure

Participants sat in front of the computer screen and were told that they were to see an image pair on each trial displaying either two objects or two social interactions. They received the instruction to judge the similarity of these two images with respect to their visual appearance (e.g., actions' shape similarity) using a 10 point rating scale, where 0 was "not at all similar" and 10 was "completely similar." Participants gave their response verbally and the experimenter recorded their rating. The answer interval was not time restricted. Then the experimenter selected a new folder in order to display a new stimulus pair.

Similar to Experiment 3, we measured first level visual similarities for the 22 first cluster levels that participants used to form a second level cluster level in Experiment 1 (10 social interactions and 12 objects). An image pair was constructed for probing first level cluster visual similarities by randomly sampling two images from the same first level category (22 two image pairs in total). In order to measure second level visual similarities, images were randomly drawn from each of the two first level categories that formed a second level category. This resulted in an additional five image pairs probing each of the five second level social interaction clusters and six image pairs probing each of the six second level object clusters. The images for the 33 pairs were randomly drawn for each participant separately. The presentation order of the 33 image pairs was random for every participant.

Results

The mean similarity ratings for each stimulus type and cluster level are shown in Figure 8. Similarity ratings seemed to be higher for objects than for social interactions and higher for first level clusters than for second level clusters. We used a completely crossed ANOVA with stimulus type (objects, social interactions) and cluster level (first, second cluster level) as within subject factors. The main effect of stimulus type, $F(1,13) = 19.42$, $p < .001$, and cluster level, $F(1,13) = 127.8$, $p < .001$, was significant. The interaction between stimulus type and cluster level was not significant, $F(1,13) = 0.001$, $p = .979$. These results show that objects are perceived as visually more similar than social interactions both for first and second cluster levels. Because social interactions are generally perceived as less similar than objects, our results suggest that participants grouped social interaction images that were perceived as visually more distinct than object images. Hence, visual similarity seems to be a less efficient heuristic for the categorization of social interactions than for objects.

It is possible that participants might have used the semantic interpretation of the stimulus rather than its physical visual appearance in their visual similarity

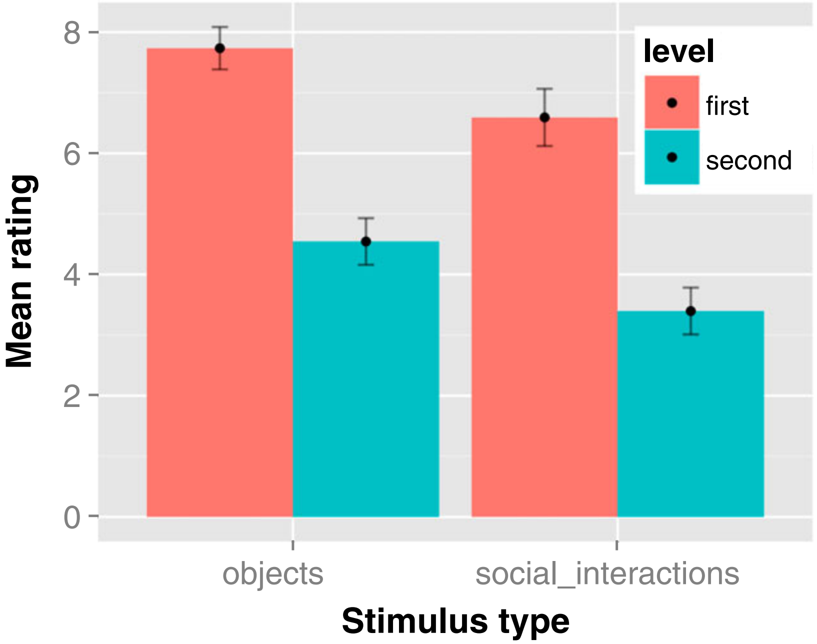

Figure 8. Mean visual similarity ratings of Experiment 4. These ratings are shown for each stimulus type and cluster level separately. Higher similarity ratings indicate higher perceived visual similarity. Bars indicate one standard error from the mean.

judgments. To rule out this confound, we conducted an additional experiment in which we measured the visual similarity of first and second level object and social interaction clusters using statistical measures typically employed in computer vision.

EXPERIMENT 5

Experiment 5 investigated visual similarity of category examples across different levels of the cluster hierarchy using computer vision. The question addressed here is whether visual categorization methods, which proved useful for objects and scenes, can also provide an effective basis for the visual categorization of social interactions. Within computer vision, a number of effective methods for objects and scene categorization have been developed and applied successfully to deal with diverse and variable categories. At present, it is still an open question whether representations that are useful for objects, object configurations, actions and scenes, are also effective for dealing with the visual categorization of social interactions, or whether the categorization of social interactions, which humans perform naturally, requires additional and perhaps different categorization cues compared with object categorization.

We used three well-known methods for comparing image similarity, which are widely used in the context of image categorization: Bag-of-Words (BoW) (Fei-Fei & Perona, 2005), Gist (Oliva & Torralba, 2001), and the Histogram-of-Gradients (HoG) (Dalal & Triggs, 2005) descriptors. In brief, the commonality among the three approaches is that they all describe an image in terms of a set of local visual features and select representative features from this feature set.

The Gist approach uses Gabor filters at different locations, orientations and spatial scales and extracts representative features using regression or discriminant analysis. It was shown to give a good overall similarity between scenes (but is less useful for comparing specific objects). The BoW approach uses local image patches from a set of representative images as image descriptors, and k-means clustering for the identification of representative patches, which are called "visual words." An image is then represented in terms of the distribution of the visual words in the image, and image similarity is measured by comparing the distributions. This measure is a visual analogue to comparing text documents by using their word distributions. The essence of the HoG approach is to describe an image region in terms of histograms of local orientation features. Images can then be compared based on local image gradients, while allowing local image distortions. Technically, a region is divided into a number of cells, an orientation histogram is computed at each cell, and the histograms are concatenated to create a description vector. Classification is then obtained e.g., by using a support vector machine (SVM) applied to the description vectors. The HoG representation is widely used for describing objects, object parts and object configurations for the purpose of recognition. For more information about each approach consult the cited literature (above).

For each of the three approaches (Gist, BoW, HoG), the following evaluations were performed. Using the images from Experiment 1, the pair-wise similarities between all the images were calculated. We then compared the similarities between images of both objects and social interactions, on both the first and second level clusters, to assess the usefulness of visual similarity as a cue for visual categorization for different categories and at different levels. Similarity scores were expressed as the ratio of within-category similarity to between-category similarity. A similarity score of one indicates that within-category similarity is identical to between-category similarity and, therefore, not a unique or category-specific cue (low cue validity).

A second evaluation was also performed, in which we directly assessed the usefulness of the three image similarity measures for the classification of object and social interaction images, in the following manner. For each object and social interaction category, we took one image as a representative of the category, and used it to classify all other images. The images from the category of the selected image constituted the positive set, and the rest of the images the negative set. Next, the distance of each image to the reference image was computed and used as its score. These scores are then used for classification: for any criterion level, images that are closer to the reference image are classified as class-examples and the more distant images as non-class. By varying the criterion, a full precision-recall curve is obtained. (Precision-recall curves are typically used in classification evaluation.) The classification performance is then measured by the average precision (AP) of the curve. If the variability of image similarities within a category is low compared with the distance to the other categories, it will give rise to better classification and a higher AP. This was done for each image in every category, resulting in several AP results for each category. Finally, the average AP was calculated for each category. This evaluation was performed at both cluster levels.

Methods

Stimuli and apparatus

We used the image set of Experiment 1. The image similarity was calculated based on the first and second level clusters of Experiment 1. The analyses were done in MATLAB.

Analysis

For the first evaluation we calculated the within- and between-category similarity for each of the three approaches separately. As described, we used the ratio of within- and between-group similarity as a similarity measure. A similarity ratio of one indicates that the within category similarity is as large as the between category similarity. Hence, a ratio of one indicates that shape similarity is not category-specific and therefore not a unique cue for clustering. The second evaluation used the image similarity as a classification measure. An Average Performance (AP) in each category was calculated—a value of one indicates that the category is fully separable from the other categories. A low AP value indicates that shape similarity is not category-specific and therefore not an effective cue for categorization.

Implementation details

For the HoG approach, the cell size was 8 pixels, as in Dalal and Triggs (2005), derived from all the images in Experiment 1. When describing an image using BoW, the image is usually divided into smaller neighbourhoods; the implementation divided the image into 2 × 2 regions, and histograms were compared within each region. For the Gist approach, the standard code (Oliva & Torralba, 2001) was used, at http://people.csail.mit.edu/torralba/code/spatialenvelope/.

Results and discussion

Figure 9 shows the results of the image analysis. Results of the first evaluation are shown in the top tow, using similarity values for each of the three measures. Values closer to 1 indicate that within-category similarity becomes increasingly analogous to between-category similarity and, therefore, indicates that image similarity is a less useful cue for categorization. For all three measures social interactions are associated with higher values than objects, indicating that shape similarity is less useful for social interaction categorization than for object categorization. Results of the second evaluation are shown on the bottom row, using the average precision (AP) of the classifier for each of the three image similarity approaches separately. Higher AP values indicate that image similarity in question is more useful for classification. Objects are associated with higher AP values than social interactions, suggesting that image similarity is more useful for the classification of object than social interaction images. Furthermore, the AP values have a tendency to decrease when going from the first to the second cluster level, suggesting that in the second cluster level image similarity is less useful for classifying both objects and social interaction.

Due to six different dependent measures, we analyzed the observed pattern for statistical significance using a between subjects MANOVA. We first assessed the correlations between the six types of similarity measures (BoW, HoG, Gist, AP_{BoW}, AP_{HoG}, AP_{Gist}). Table 3 shows that correlations between variables were mostly moderate ($0.2 > |r| > 0.6$) confirming the appropriateness of conducting a

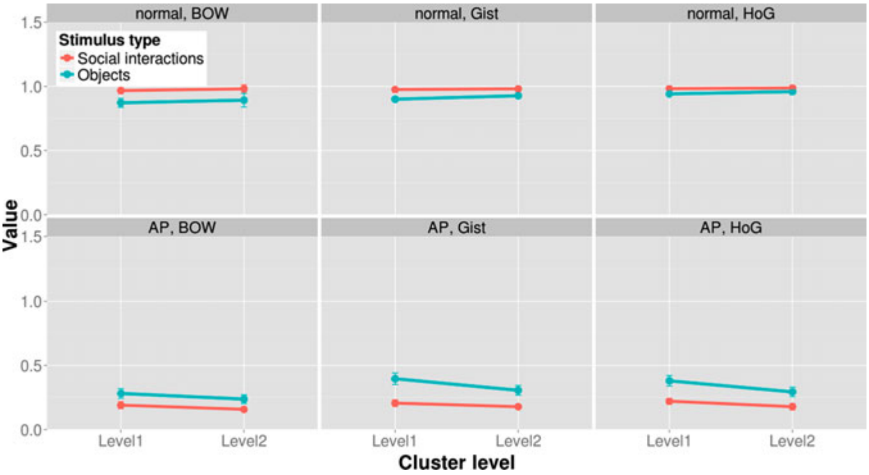

Figure 9. The image similarity measures shown for the BoW, Gist, and HoG approach separately. The top row shows the ratio of the within vs. between similarity measure. The bottom row shows the average precision (AP) of the classification for each of the three approaches.

TABLE 3
Correlations and Holm corrected *p* values (in brackets) between the different types of image similarity measures used in Experiment 5

	BoW AP	*BoW normal*	*Gist AP*	*Gist normal*	*HoG AP*	*HoG normal*
BoW AP	1 (0.00)	−0.76 (0.00)	0.51 (0.01)	−0.42 (0.07)	0.45 (0.05)	−0.39 (0.10)
BoW normal	−0.76 (0.00)	1 (0.00)	−0.25 (0.59)	0.19 (0.59)	−0.24 (0.59)	0.25 (0.59)
Gist AP	0.51 (0.00)	−0.25 (0.15)	1 (0.00)	−0.97 (0.00)	0.83 (0.00)	−0.78 (0.00)
Gist normal	−0.42 (0.01)	0.19 (0.27)	−0.97 (0.00)	1 (0.00)	−0.78 (0.00)	0.76 (0.00)
HoG AP	0.45 (0.01)	−0.24 (0.16)	0.83 (0.00)	−0.78 (0.00)	1 (0.00)	−0.97 (0.00)
HoG normal	−0.39 (0.02)	0.25 (0.15)	−0.78 (0.00)	0.76 (0.00)	−0.97 (0.00)	1 (0.00)

MANOVA. We compared the influence of cluster level and stimulus type on the six types of similarity measures using a between subjects MANOVA. The MANOVA showed a significant effect of stimulus type, $F(1,31) = 4.454$, $p = .003$, and no significant main effect for cluster level, $F(1,31) = 0.902$, $p < .509$, and no significant interaction between stimulus type and cluster level, $F(1,31) = 0.182$; $p = .979$. These results indicate that across the six similarity measures social interactions exhibit less visual similarity than objects. This pattern is the same for the first and second cluster level. Importantly, the average precision of the classifier measure suggests that image similarity is a less useful cue for the classification of social interactions than for objects at the first cluster level.

Additionally, we were interested in the comparison of the dendrograms resulting from the physical similarity measures with those of the behavioural data (Experiment 1). We therefore calculated pairwise similarities (HoG, BoW, and Gist similarities) for the object and social interaction images separately. The six similarity matrices (three physical measures times two stimulus types) derived from these pairwise similarities were used in cluster analyses as described in Experiment 1. To compare the similarity of cluster analysis results of the physical measure with the one obtained from participants (in Experiments 1 and 2), we used the above mentioned ARI and the Jaccard Index. For both indices, higher values indicate larger agreement between the cluster analysis results. The ARI and Jaccard index are shown in Figure 10. Both indices were higher for objects than for social interactions indicating that the cluster results based on physical measures were more similar to the behavioural results for objects than for social interaction grouping. These results show that physical measures were able to explain more of the behavioural object categorization than the social interaction categorization.

In summary, six prominent measures of image similarity indicate that image similarity is overall a less useful cue for the categorization of social interactions than for the categorization of objects. Hence, using image similarity as a

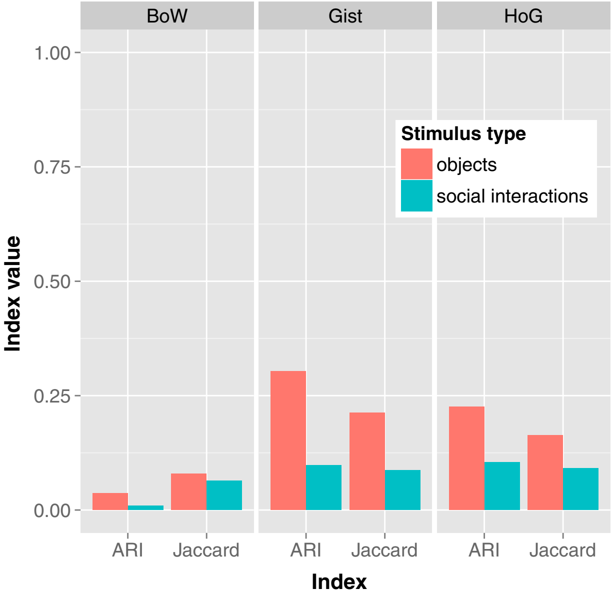

Figure 10. Comparison of clusters derived from physical similarity judgement (Experiment 5) and the behavioural grouping experiment (Experiment 1). The clusters were compared using the Adjusted Rand Index (ARI) and the Jaccard Index (along x-axis). Higher index values indicate higher similarity of the cluster results. The comparison results are shown for each stimulus type (indicated by different colours) of the three measures used in Experiment 5 (across panels).

heuristic for visual categorization would be less efficient for the categorization of social interaction than for the categorization of object images.

Previous research has shown that the significance of the basic level in visual object recognition (akin to the second cluster level in the present study) rests in the superior recognition performance of objects at the basic level (Grill-Spector & Kanwisher, 2005; Mack, Gauthier, Sadr, & Palmeri, 2008; Rosch et al., 1976). Brief glimpses of object images (e.g., de la Rosa et al., 2011) allow better detection than basic level categorization and better basic level categorization than sub-ordinate categorization performance. Experiment 6 examined whether recognition advantages are associated with some of the social interaction cluster levels observed in Experiment 2. As a control, we also probed object recognition in these three recognition tasks.

In Experiment 6, we briefly flashed an object or social interaction image and a noise image (two interval forced choice paradigm) and asked participants to name the interval that contained the object or social interaction (detection), the second level cluster, and the first level cluster on every experimental trial. For brief presentations of object images, we expected participants to correctly name the second level object cluster significantly more often than first level object cluster if objects were recognized faster on the basic than on the sub-ordinate level. As for social interactions, we wanted to explore the recognition performance associated with the different cluster levels.

EXPERIMENT 6

Method

Participants

Ten participants from the community of Tübingen (6 males; age range: 20–32 years) gave their informed consent prior to the experiment and received €8/hour for their participation. All participants were naïve with regards to the research question. The study was conducted in line with the Declaration of Helsinki.

Apparatus and stimuli

Stimuli were presented on a Sony (Tokyo, Japan) Monitor (CPD-G500) using the Psychtoolbox 3 (Brainard, 1997; Kleiner, Brainard, & Pelli, 2007; Pelli, 1997). The gamma-corrected monitor had a refresh rate of 140 Hz.

We sampled new images for each first and second cluster level of Experiment 1 to obtain a larger variety of individual actions and objects for each category and to increase external validity. A total of 600 object and 600 social interaction images with 100 images for each category was used. Each category consisted of 50 exemplar images (objects: pigeon, sailboat, VW Beetle, German Sheppard, barn and rose; social interactions: kiss, fight, begging, soccer, handshake, dance) and 50 non-exemplar images (other greetings were waving, high fives; other sports were ice hockey, volleyball, basketball; other types of help were carrying a table together, carrying an injured person, carrying a stretcher; other types of affection were hugging, cuddling; other means of fights were shouting at each other). The non-exemplar stimuli were taken from the same basic level category because if these stimuli were sampled across different basic level categories participants could use basic level classification to solve the sub-ordinate classification task. A new patch of Gaussian visual noise was used as a non-object image on every trial. The mask that followed the image presentation was a scrambled version of the object image by chopping up an image of the respective stimulus type into 10 by 10 pixels tiles and randomly rearranging the tiles. All stimuli were presented centrally, had a size of 5.89° visual angle, luminance of 127 RGB pixel value, and a pixel contrast of 20 RMS RGB.

Procedure

We used an experimental procedure identical to Experiment 1 reported in the study of de la Rosa and colleagues (de la Rosa et al., 2011). This procedure probes the degree of recognition of a briefly glimpsed stimulus. In brief, participants were informed with example images about each first-level cluster and second-level cluster at the beginning of the experiment (these were not used during testing). A trial consisted of two temporally backward-masked image presentations of which one presented an object or social interaction image (depending on the condition) and the other a patch of visual noise. The order of the image and noise presentation interval was counterbalanced across all experimental trials of a participant. The presentation time for both presentation intervals was the same and randomly chosen for each trial. After the presentation of the two intervals, participants' recognition performance for three recognition tasks was probed: detection, first cluster level, and second cluster level recognition. Specifically, participants selected the presentation interval of the object or social interaction image (detection task), the first-level cluster name, and the second-level cluster name from a list of given answer labels presented on an answer screen. The answer interval was not time restricted and the answer screen stayed on until participants had given all responses. The answer labels were "1" and "2" for the detection task referring to the presentation interval that contained the object/social interaction. The second level cluster names were "bird," "car," "building," "flower," "dog," "ship" in the object recognition task and "affection," "arguing," "greeting," "distressed," "sports" in the social interaction recognition task. The answer labels in the sub-ordinate object recognition task were "poodle," "German Sheppard," "parrot," "pigeon," "rose," "sunflower," "barn," "house," "Smart car," "Beetle," "sailboat," "cruise ship." Finally, the labels "hockey," "soccer," "arguing," "fight," "waving," "shaking hands," "rescue," "begging," "kiss," "hug" were the answer options in the sub-ordinate social interaction recognition task. We have previously shown probing visual recognition in this way provides similar results to one-interval-forced choice tasks, which are run for each recognition task separately (de la Rosa et al., 2011).

Each presentation time (7, 21, 28, 36, 57, 78, 121 ms) was probed six times during an experimental run amounting to a total of 42 trials per run. Seven runs constituted an experiment (total of 294 trials). In the object recognition tasks, the probability of correctly guessing the presentation interval (detection task), the second cluster level, and the first cluster level was p = .5, p = 1/6, and p = 1/7, respectively. As for social interactions, the probability of correctly guessing the presentation interval (detection task), the second cluster level, and the first cluster level was p = .5, p = 1/5, and p = 1/6, respectively (note that there was one less social interaction second level cluster than object cluster (see Experiment 1)). The testing order of stimulus type (social interactions, objects) was counterbalanced across participants.

Results and discussion

Recognition performance was corrected-for-guessing according to Macmillan and Creelman (Macmillan & Creelman, 2005, p. 252). Figure 11 shows the corrected-for-guessing accuracy as a function of presentation time for each recognition task and stimulus type separately. For both objects and social interactions, recognition performance was best for the detection task, second best for the second cluster level recognition task, and worst for the first cluster level recognition task. However, the difference between first and second cluster level recognition was much more pronounced for object than for social interaction recognition.

We analyzed the effect of task and presentation time on object and social interaction recognition in separate ANOVAs because the primary interest of Experiment 6 was to examine whether first cluster level differed significantly from second cluster level recognition in object and social interaction recognition.

In the following, we report Greenhouse-Geisser corrected *p*-values where assumptions of sphericity had not been met.

Objects

We examined the effect of presentation time and task on accuracy using a complete within-subject ANOVA. The ANOVA showed a significant main effect of presentation time, *F*(6,54) = 308.62, $p < .001$, a significant main effect of task, *F*(2,18) = 164.32, $p < .001$, and a significant interaction between task and presentation time, *F*(12,108) = 26.27, $p < .001$. Differences between the three

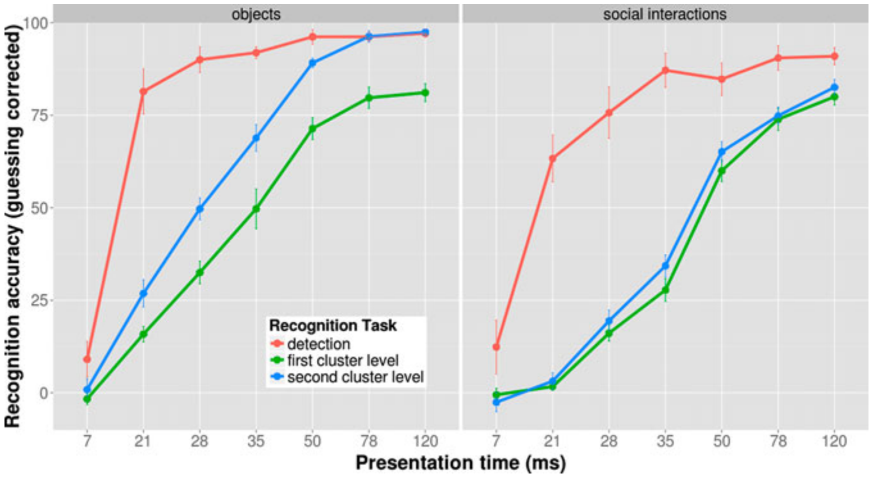

Figure 11. Average recognition accuracy (corrected for guessing) in Experiment 6 shown for each stimulus type, presentation time, and task separately. Bars indicate one standard error from the mean.

recognition tasks, therefore, depend on presentation time. We examined this interaction in the following way. Because previous research showed that detection performance is better than basic-level categorization, and basic-level recognition is better than sub-ordinate recognition, we compared detection and second cluster level recognition, and then first and second cluster level recognition.

Detection vs. recognition at the second cluster level. We investigated differences between detection and second cluster level recognition in a two-way ANOVA with presentation time and recognition task (detection, second cluster level recognition) as within subject factors. The main effect of presentation time was significant, $F(6,54) = 259.31$, $p < .001$, and the main effect of task was significant, $F(1,9) = 101.30$ $p < .001$. The interaction of presentation time and task was also significant, $F(6,54) = 34.25$, $p < .001$, suggesting that detection and second cluster level recognition performance differences varied with presentation time.

First vs. second level cluster recognition. A two-way ANOVA with presentation time and recognition task (first and second cluster level recognition) as within subject factors showed a significant main effect of presentation time, $F(6,54) = 241.72$, $p < .001$, and, more importantly, a significant main effect of task, $F(1,9) = 70.12$, $p < .001$. The interaction between presentation time and recognition task was also significant, $F(6,54) = 5.54$, $p = .004$. These results suggest that recognition of objects at the second cluster level is faster than at the first level and that this difference depends on presentation time. Because the first and second level clusters in the present study are akin to previously basic and sub-ordinate objects levels, the results replicate previous findings suggesting that object recognition is faster at the basic level than at the sub-ordinate level (e.g., de la Rosa et al., 2011).

Social interactions

We investigated the effect of presentation time and task on accuracy using a completely crossed within-subject ANOVA. The ANOVA showed a significant main effect of presentation time, $F(6,54) = 257.30$, $p < .001$, a significant main effect of task, $F(2,18) = 74.95$, $p < .001$, and a significant interaction between task and presentation time, $F(12,108) = 24.41$, $p < .001$. These results indicate that differences in recognition performance across the three tasks are dependent on presentation time. We dissected this interaction by comparing detection and second cluster level recognition, and first and second cluster level recognition in two separate ANOVAs.

Detection vs. recognition at the second cluster level. A two-way ANOVA with presentation time and recognition task (detection, second cluster level recognition) as within subject factors found a significant main effect of presentation time, *F*(6,54) = 165.41, $p < .001$, and a significant main effect of recognition task, *F*(1,9) = 62.99, $p < .001$. The interaction between recognition task and presentation time was also significant, *F*(6,54) = 25.04, $p < .001$. The analysis suggests that detection and second cluster level recognition are associated with varying performance differences across presentation times.

First vs. second level cluster recognition. An ANOVA with presentation time and recognition task (first and second cluster level recognition) as within subject factors revealed a significant main effect of presentation time, *F*(6,54) = 265.79, $p < .001$, and a significant main effect of recognition task, *F*(1,9) = 6.07, $p < .036$. The interaction between presentation time and recognition task was non-significant, *F*(6,54) = 1.72, $p = .132$. The significant main effect of task indicated that social interactions are better recognized at the second cluster level (41.17% accuracy) than at the first cluster level (38.33% accuracy).

Comparing the first and second cluster level recognition advantage across stimulus types

Figure 11 suggests that the difference between second and first cluster level recognition is more pronounced for object than for social interaction recognition. We compared this difference directly in a two way ANOVA with task (first and second cluster level only), stimulus type, and presentation time as within subject factors. The main effect of presentation time, *F*(6,54) = 385.19, $p < .001$, the main effect of task, *F*(1, 9) = 43.82, $p < .001$, and the main effect of stimulus type, *F*(1,9) = 193.99, were significant. The interaction between presentation time and task, *F*(6,54) = 6.16, $p < .001$, and the interaction between presentation time and stimulus type, *F*(6,54) = 10.97, $p < .001$, was significant. More importantly, the interaction between task and stimulus type, *F*(1,9) = 74.14, $p < .001$, was also significant. Hence, the difference between first and second cluster level recognition is smaller for social interaction than for object recognition (see Figure 11). The three way interaction between presentation time, stimulus type, and task was non-significant, *F*(6,54) = 1.51, $p = .193$.

Discussion

We examined recognition performance of objects and social interactions at the first and second cluster level and in a detection task. We found a recognition advantage at the second cluster level compared to the first cluster level for both stimulus types. This recognition advantage, however, was significantly smaller for social interaction than for object recognition. Note that the difference between first and second cluster level recognition for social interaction recognition was small also in terms of its absolute value. In summary, social interactions are recognized slightly better at the second cluster level.

The results bear implications for the interpretation of the cognitive categories measured in Experiments 1 and 2. The improved visual recognition performance at the second cluster level compared to the first cluster level has previously been taken as evidence for the special status of the second cluster level for visual recognition of objects (therefore also termed basic level or entry level; Jolicoeur et al., 1984). The results of this experiment replicate these results and go beyond this knowledge by showing that they do not apply straightforwardly to social interactions. Specifically, while we replicate the well known basic/entry-level advantage in object recognition, we provide evidence for only a small performance advantage in the recognition of social interactions.

GENERAL DISCUSSION

We examined the visual categorization of social interaction images. For a better interpretation of the social interaction categorization results, participants also conducted all tasks with object images. In Experiment 1 we asked participants to group social interaction and object images into self created groups so that items within a group were more similar to each other than to the items of contrasting groups. Similarity ratings were based on the between subject variability of the grouping behaviour. A cluster analysis on the grouping data revealed a cluster hierarchy in which images of the same actions (e.g., handshake) or images of items (e.g., VW Beetle) were at the lowest level (i.e., first level) of the hierarchy. Often two first level clusters formed the next higher cluster level (second level).

Experiment 2 replicated the grouping task of Experiment 1 using a multi-level grouping approach in which similarity ratings were based on within subject similarity judgments. There was a striking similarity between the results of Experiments 1 and 2 suggesting that the results were independent of whether similarity was measured in a between (Experiment 1) or within subject fashion (Experiment 2).

Experiment 3 compared motor movements similarity ratings between first and second object and social interaction cluster levels. We found that similarity ratings of motor movements associated with an object interaction were similar for first and second level clusters. This finding is in line with previous reports that motor movements are reliable cues for object categorization (Rosch, 1978). In contrast, the motor similarity ratings for social interactions were significantly different between the second level cluster and their constituent first level clusters. This result supports the idea that participants tally social interactions into the same category whose motor patterns were rated statistically significantly different.

Experiment 4 examined the visual similarity of social interaction and object images at the first and second level of the cluster hierarchy by means of participants' ratings. The results showed that objects were overall perceived as more similar than social interactions. The results suggest that visual similarity is a less effective heuristic for the categorization of social interactions compared to objects. To ensure that visual similarity ratings were not biased by the semantic interpretation of the object or social interaction, we borrowed methods from computer vision to assess visual similarities in Experiment 5. Specifically, we used three different methods commonly used to compare visual similarities of object and social interaction images at the first and second cluster level. We found object images of the same category to be more similar than social interaction images. Moreover, the AP values suggest that image similarity is a less useful cue for social interaction classification than for object classification. Hence, grouping social interaction images based on visual similarity is a less efficient heuristic than for object images.

Finally, in Experiment 6 we assessed the visual recognition performance associated with the detection and the recognition of objects and social interactions on the first and second cluster level. Although we find for both stimulus types that second level cluster recognition is significantly better than first level cluster recognition, this difference was very small for social interactions recognition.

Taken together the results promote the idea of different abstraction levels in the recognition of social interactions from images (Experiments 1 and 2). The existence of more than one abstraction level in the recognition of social interactions has broad theoretical implications. Current influential theoretical accounts of action recognition assume that actions are recognized in only one way (i.e., at only one abstraction level) (Fleischer et al., 2013; Giese & Poggio, 2003; Rizzolatti et al., 2001). The observation of at least two abstraction levels in social interaction recognition demonstrates a more complex mapping of visual action information onto semantic knowledge than previously assumed. In the future, models of action recognition might want to provide more detailed information about how sensory action information is filled with semantic information. Moreover, the existence of two abstraction levels in social interaction recognition points to the importance of controlling for the abstraction level in behavioural and imaging studies. In particular, one cannot exclude the possibility that the recognition of actions at different abstraction levels might involve different neural populations. To warrant a correct interpretation of neural correlates underlying action recognition, one needs to control for the abstraction level in recognition tasks. Our investigation might be also interesting for clinical research (e.g., Autism Spectrum Disorder) as they raise the question whether the mapping of visual information onto semantic knowledge occurs in a similar fashion in these patient populations.

These abstraction levels underlying social interaction recognition seem to be somewhat qualitatively different from object recognition abstraction levels: Physical similarity (as measured by visual appearance and motor movements) was a less efficient heuristic for explaining the emergence of social interaction categories than object categories (Experiments 3 to 5). Moreover, second level social interaction recognition leads to a much smaller recognition advantage than second level object recognition (which is akin to basic level object recognition). Finally, we observed reliable cluster levels beyond the second level for objects only. These higher levels have resemblance with previously reported superordinate object levels. Overall, our findings suggest that the category structure differs between objects and social interactions.

The object grouping patterns of Experiments 1 and 2 resemble previously reported object abstraction levels: First level clusters have strong resemblance to sub-ordinate object levels because they were composed of only one object type (e.g., Smart car) and the naming of the elements in these clusters was consistent with sub-ordinate object level names. Moreover, images within each cluster were perceived as highly similar as suggested by very small ESS values. The second level clusters seems to correspond to basic object levels as their composition and naming is in line with previous reports of basic object levels (Mack et al., 2008; Mack & Palmeri, 2011; de la Rosa et al., 2011). Moreover, second level object recognition is associated with a clear recognition advantage over first level cluster recognition. In summary, first and second object cluster levels seem to correspond well with sub-ordinate and basic abstraction levels, respectively.

We observed different abstraction levels also for social interactions. In contrast to object abstraction levels, the emergence of these abstraction levels was not explained to the same degree by motor and visual similarities between within-category items. Which possible cues drive the categorization of social interactions? Of relevance in this regard is a recent study investigating the tuning of processes underlying action categorization using an adaptation paradigm (de la Rosa, Streuber, Giese, Bülthoff, & Curio, 2014). The study found that action categorization adaptation after-effects were modulated by social context that preceded the adaptors. Because the social context manipulation caused participants to change their interpretation of the action displayed by the adaptors, the modulation of the adaptation after-effects is indicative of the sensitivity of action categorization processes to the semantic content of the displayed action. One can therefore speculate that the visual processes underlying the categorization tasks in the present study are also tuned to some degree to action semantics. Action categorization and, therefore, the emergence of action abstraction levels might be based to some degree on the semantic meaning of an action.

The basic level (second level) social interaction recognition is associated with only a small recognition advantage compared to object recognition. While the exact reasons for this difference remain subject to future research, visual expertise might account for some of this difference. Tanaka and colleagues observed that the basic level in object recognition depends on visual expertise. For example, dog and bird experts recognize dogs and birds as fast on the subordinate level as on the basic level (Tanaka & Taylor, 1991). It is likely that humans are visual experts in perceiving social interactions because social interactions occur frequently in everyday life. In this regard, social interactions might be similar to faces for which it has been argued that specialized neural processes have evolved for their processing due to their frequent occurrence (Kanwisher, McDermott, & Chun, 1997; Tarr & Gauthier, 2000). This expertise might account for the diminished recognition difference between basic level and subordinate levels in social interaction compared to object recognition. In the future, it would be interesting to examine the plausibility of this explanation for recognition difference between basic and subordinate level recognition of objects and social interactions.

How far are these results applicable to dynamic social interactions? At this stage, it is difficult to say with certainty whether the reported findings directly apply to dynamic social interactions. It is likely that similar abstraction levels also exist for the recognition of dynamic social interactions. This suggestion is mainly motivated by anecdotal evidence. For example, there is little doubt that seeing two persons shaking hands can be interpreted as shaking hands or as a greeting. However, it is less clear whether the recognition advantage can also be found with dynamic action stimuli. An interesting future question would be the examination of whether a recognition advantage is also found in the recognition of dynamic social interactions (see also de la Rosa et al., 2013 for the recognition of dynamic social interactions).

How far do the reported results depend on the specific stimulus set? Previous studies suggest that object categorization performance depends on the composition of the stimulus set (Macé et al., 2009). Specifically, as the stimulus set becomes more diverse the primacy of categorization performance at the basic level decreases or even vanishes (Macé et al., 2009). While these results might seem to go against the idea of fixed abstraction levels, the results of this research could be well explained by fixed abstraction levels in which the highest abstraction level is activated first (Macé et al., 2009). More compelling evidence for the plasticity of abstraction level comes from studies showing that categorization performance depends on long term visual expertise (Tanaka & Taylor, 1991). Based on this research we cannot exclude the possibility that categorization is different for different stimuli sets. However, we find it more likely that long term visual experience will alter visual categorization.

In summary, we investigated visual categorization of objects and social interactions and found evidence for different abstraction levels in the recognition of static social interactions and objects. Basic-level like abstraction levels showed a clear recognition advantage for objects and a small recognition advantage for social interactions. Moreover, shape and movement cues proved a less efficient heuristic for social interaction categorization than for object categorization. Social interaction categorization appears to be different from object categorization. Our results suggest that action information is associated with semantic meaning at different abstraction levels. Our study therefore points to a novel class of cognitive processes in action recognition that have received little attention so far.

REFERENCES

Baldwin, D., Andersson, A., Saffran, J., & Meyer, M. (2008). Segmenting dynamic human action via statistical structure. *Cognition*, *106*, 1382–1407. doi:10.1016/j.cognition.2007.07.005

Blake, R., & Shiffrar, M. (2007). Perception of human motion. *Annual Review of Psychology*, *58*(1), 47–73. doi:10.1146/annurev.psych.57.102904.190152

Barraclough, N. E., Ingham, J., & Page, S. A. (2012). Dynamics of walking adaptation aftereffects induced in static images of walking actors. *Vision Research*, *59*, 1–8. doi:10.1016/j.visres.2012.02.011

Boucart, M., Moroni, C., Thibaut, M., Szaffarczyk, S., & Greene, M. (2013). Scene categorization at large visual eccentricities. *Vision Research*, *86*, 35–42. doi:10.1016/j.visres.2013.04.006

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, *10*, 433–436. doi:10.1163/156856897X00357

Corter, J. E., & Gluck, M. A. (1992). Explaining basic categories: Feature predictability and information. *Psychological Bulletin*, *111*, 291–303. doi:10.1037/0033-2909.111.2.291

Dalal, N., & Triggs, B. (2005). Histograms of oriented gradients for human detection. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, *1*, 886–893. doi:10.1109/CVPR.2005.177

Dittrich, W. H. (1993). Action categories and the perception of biological motion. *Perception*, *22*(1), 15–22. doi:10.1068/p220015

Fabre-Thorpe, M., Delorme, A., Marlot, C., & Thorpe, S. (2001). A limit to the speed of processing in ultra-rapid visual categorization of novel natural scenes. *Journal of Cognitive Neuroscience*, *13*, 171–180.

Fei-Fei, L., & Perona, P. (2005). A Bayesian hierarchical model for learning natural scene categories. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, *2*, 524–531. doi:10.1109/CVPR.2005.16

Fleischer, F., Caggiano, V., Thier, P., & Giese, M. A. (2013). Physiologically inspired model for the visual recognition of transitive hand actions. *The Journal of Neuroscience*, *33*, 6563–6580. doi:10.1523/JNEUROSCI.4129-12.2013

Freedman, D. J., & Miller, E. K. (2008). Neural mechanisms of visual categorization: Insights from neurophysiology. *Neuroscience and Biobehavioral Reviews*, *32*, 311–329. doi:10.1016/j.neubiorev.2007.07.011

Giese, M. A., & Poggio, T. (2003). Neural mechanisms for the recognition of biological movements. *Nature Reviews Neuroscience*, *4*, 179–192. doi:10.1038/nrn1057

Grill-Spector, K., & Kanwisher, N. (2005). Visual recognition: As soon as you know it is there, you know what it is. *Psychological Science*, *16*, 152–160. doi:10.1111/j.0956-7976.2005.00796.x

Jolicoeur, P., Gluck, M. A., & Kosslyn, S. M. (1984). Pictures and names: Making the connection. *Cognitive Psychology*, *16*, 243–275. doi:10.1016/0010-0285(84)90009-4

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *The Journal of Neuroscience?: The Official Journal of the Society for Neuroscience*, *17*, 4302–4311.

Kleiner, M., Brainard, D., & Pelli, D. G. (2007). What's new in psychtoolbox-3? *Perception*, *36*, 14.

Loula, F., Prasad, S., Harber, K., & Shiffrar, M. (2005). Recognizing people from their movement. *Journal of Experimental Psychology: Human Perception and Performance*, *31*, 210–220. doi:10.1037/0096-1523.31.1.210

Macé, M. J.-M., Joubert, O. R., Nespoulous, J.-L., & Fabre-Thorpe, M. (2009). The time-course of visual categorizations: You spot the animal faster than the bird. *PloS One*, *4*, e5927. doi:10.1371/journal.pone.0005927

Mack, M. L., Gauthier, I., Sadr, J., & Palmeri, T. J. (2008). Object detection and basic-level categorization: Sometimes you know it is there before you know what it is. *Psychonomic Bulletin & Review*, *15*(1), 28–35. doi:10.3758/PBR.15.1.28

Mack, M. L., & Palmeri, T. J. (2011). The timing of visual object categorization. *Frontiers in Perception Science*, *2*, 165. doi:10.3389/fpsyg.2011.00165

Macmillan, N. A., & Creelman, C. D. (2005). *Detection Theory: A User's Guide*. Lawrence Erlbaum Associates.

Oliva, A., & Torralba, A. (2001). Modeling the shape of the scene : A holistic representation of the spatial envelope. *International Journal of Computer Vision*, *42*, 145–175. doi:10.1023/A:1011139631724

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*, 437–442. doi:10.1163/156856897X00366

Rizzolatti, G., Fogassi, L., & Gallese, V. (2001). Neurophysiological mechanisms underlying the understanding and imitation of action. *Nature Reviews Neuroscience*, *2*, 661–670.

Roether, C. L., Omlor, L., Christensen, A., & Giese, M. A. (2009). Critical features for the perception of emotion from gait. *Journal of Vision*, *9*, 1–32. doi:10.1167/9.6.15

de la Rosa, S., Choudhery, R. N., & Chatziastros, A. (2011). Visual object detection, categorization, and identification tasks are associated with different time courses and sensitivities. *Journal of Experimental Psychology. Human Perception and Performance*, *37*(1), 38–47. doi:10.1037/a0020553

de la Rosa, S., Mieskes, S., Bülthoff, H. H., & Curio, C. (2013). View dependencies in the visual recognition of social interactions. *Frontiers in Psychology*, *4*, 752. doi:10.3389/fpsyg.2013.00752

de la Rosa, S., Streuber, S., Giese, M., Bülthoff, H. H., & Curio, C. (2014). Putting actions in context: Visual action adaptation aftereffects are modulated by social contexts. *PloS One*, *9*, e86502. doi:10.1371/journal.pone.0086502.g006

Rosch, E. (1978). Principles of categorization. In E. Rosch & B. Lloyd (Eds.), *Cognition and categorization* (pp. 27–48). Hillsdale, NJ: Lawrence Erlbaum.

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, *8*, 382–439. doi:10.1016/0010-0285(76)90013-X

Shimodaira, H. (2002). An approximately unbiased test of phylogenetic tree selection. *Systematic Biology*, *51*, 492–508. doi:10.1080/10635150290069913

Tanaka, J. W., & Taylor, M. (1991). Object categories and expertise: Is the basic level in the eye of the beholder? *Cognitive Psychology*, *23*, 457–482. doi:10.1016/0010-0285(91)90016-H

Tarr, M. J., & Gauthier, I. (2000). FFA: A flexible fusiform area for subordinate-level visual processing automatized by expertise *Nature Neuroscience*, *3*, 764–769. doi:10.1038/77666

Thorpe, S. J., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature*, *381*, 520–522. doi:10.1038/381520a0

Thorpe, S. J., Gegenfurtner, K. R., Fabre-Thorpe, M., & Bülthoff, H. H. (2001). Detection of animals in natural images using far peripheral vision. *The European Journal of Neuroscience*, *14*, 869–876. doi:10.1046/j.0953-816x.2001.01717.x

Troje, N. F., Westhoff, C., & Lavrov, M. (2005). Person identification from biological motion: Effects of structural and kinematic cues. *Perception & Psychophysics*, *67*, 667–675. doi:10.3758/BF03193523

VanRullen, R., & Thorpe, S. J. (2001). Is it a bird? Is it a plane? Ultra-rapid visual categorisation of natural and artifactual objects. *Perception*, *30*, 655–668. doi:10.1068/p3029

Ward, J. (1963). Hierarchical grouping to optimize an objective function. *Journal of ASA*, *58*, 236–244.