The Grounded Colour Naming Game

Joris Bleys Martin Loetzsch
Artifcial Intelligence Laboratory
Vrije Universiteit Brussel, Belgium
jorisb@arti.vub.ac.be

Michael Spranger Luc Steels
Sony Computer Science Laboratory
Paris, France
steels@arti.vub.ac.be

***Abstract*— Colour naming games are idealised communicative interactions within a population of artificial agents in which a speaker uses a single colour term to draw the attention of a hearer to a particular object in a shared context. Through a series of such games, a colour lexicon can be developed that is sufficiently shared to allow for successful communication, even when the agents start out without any predefined categories. In previous models of colour naming games, the shared context was typically artificially generated from a set of colour stimuli and both agents in the interaction perceive this environment in an identical way. In this paper, we investigate the dynamics of the colour naming game in a robotic setup in which humanoid robots perceive a set of colourful objects from their own perspective. We compare the resulting colour ontologies to those found in human languages and show how these ontologies reflect the environment in which they were developed.**

I. THE (GROUNDED) COLOUR NAMING GAME

Within the general framework of artificial language evolution (for reviews see [8], [23], [17]), *language games* [25], [16] have been used to study various aspects of language, such as the marking of perspective reversal [19]. A language game is a routinised linguistic interaction in which agents try to achieve communicative goals using their personal conceptual and linguistic knowledge.

In the domain of colour, *colour naming games* [18] have been used to study how a population of agents can develop their own colour lexicon through language, which consists of both a set of colour categories and terms to refer to these categories. Each interaction involves two agents, a speaker and a hearer that are randomly chosen from a population. Both agents jointly perceive a shared context consisting of a set of colour stimuli. The goal of the speaker is to draw the attention of the hearer to a randomly selected colour stimulus using one single non-compound colour term.

To achieve this goal, the speaker first categorises the topic using his own conceptual inventory and utters the term that is associated with this category. If no discriminating category was found, an invention strategy is applied. The hearer finds the category associated to this term and points to the stimulus that fits this category best. The speaker compares the object pointed at by the hearer to his own intended object. When these are the same, he signals communicative success (non-linguistically). If the hearer doesn't know the colour term yet or points to the wrong stimulus, the speaker will signal a communicative failure and point to the topic. If the word is unknown to the hearer, the hearer will use an adoption strategy. Based on the outcome of the game both agents modify their colour lexicon slightly to be able to communicate better in future interactions using an alignment strategy.

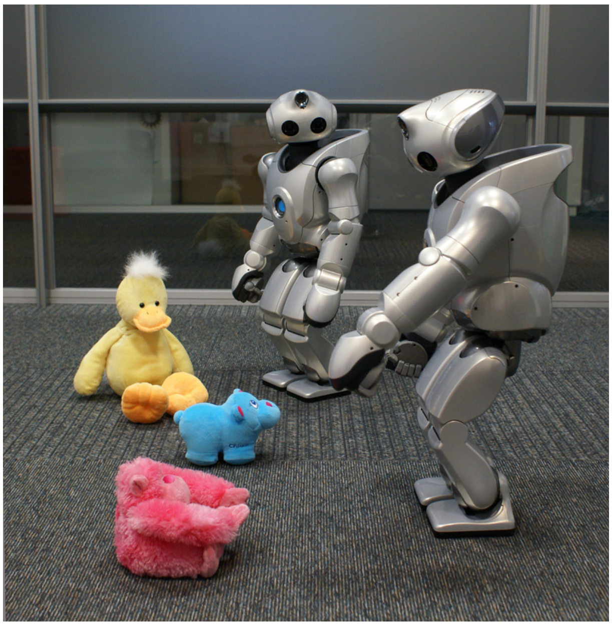


Fig. 1. The grounded colour naming game. The speaker (furthest) needs to draw the attention of the hearer (nearest) to one of the objects in a shared context using one term that is associated to a single colour category. The game is a success when the hearer is able to identify this object.

Over the course of a series of such interactions, agents develop a colour lexicon that is sufficiently shared to allow for successful communication in such language games [18]. The dynamics of the colour naming game have been studied [12] and the resulting colour ontologies have been compared to anthropological findings [1].

Most of these studies have assumed that no difference exists in how both agents perceive the colour stimuli in the context. This is not very realistic for embodied interactions in which embodied agents perceive colourful objects around them using their own vision system. Embodied agents will never share the same position in the world and hence will perceive the objects from their own perspective. The colours of the objects perceived by the agents will never be identical, due to for example differences in lighting conditions or appearances when perceived from different sides.

To investigate the influence of this *perceptual deviation*, we introduce the *grounded colour naming game* in which agents are embodied in humanoid robots (Fig. 1) and individually perceive scenes through their cameras. The setup is similar to other language game experiments with robots (e.g. Lego robots [22], pan-tilt cameras [15], Sony Aibo robots [19], [10] and the same robots as used in this paper [24]). This is the first study on the influence of embodiment on the dynamics of the colour naming game.

We will discuss the difference in colour distributions in the stimuli sets of a simulated and an embodied experiment and quantify the perceptual deviations of embodied agents. We will show how these different aspects affect the dynamics of the colour naming game and compare the resulting ontologies to those found in human languages.

II. ROBOTIC SETUP AND VISUAL PERCEPTION

The Sony humanoid robots [4] used in this experiment are about 60 cm high, weigh approximately 7 kg and have 38 degrees of freedom. The robots are placed in a closed office environment in which a set of coloured objects are placed. Before each interaction, the experimenter modifies the current scene by adding or removing an object or by changing the position or orientation of an object in the scene. Each scene contains between two and four coloured objects from a set of 20 objects (Fig. 2). The main sensor used for perception is one of the three CCD cameras in the head of the robot.

The main goal of the robot's vision system is to construct persistent internal representations of the objects in the robot's environment. This system involves three sub-systems. First, low-level vision routines process raw camera images to yield basic *percepts* (Figs. 3(a)-3(d)). Percepts are connected regions that differ from the background of the environment. The statistics of the environment's background are acquired in a calibration phase.

Second, these foreground regions are tracked in subsequent camera images despite changing positions and appearances of the objects. In order to do so, the vision system needs to establish a correspondence between the internal *object model* and the image regions that refer to the same physical object, a process known in robotics as *anchoring* [3]. Colour histograms of already established object models are used to classify image regions with respect to their similarity to object models (Fig. 3(d)). Kalman Filters [6] are used to associate classified regions to object models based on colour similarity and position in the image (Fig. 3(e)).

Third, when needed in communicative interactions, the vision system encodes a set of visual properties about each object model. These properties are colour, position, height and width, but in this experiment only the colour information of each object is used. The camera of the robot delivers up to 30 images per second with a resolution of 176×144 pixels in the YCrCb[1] colour space. The colour of the object is the average colour of all pixels that make up the object. This average colour is transformed into the perceptually equidistant colour space CIE $L^*a^*b^*$[2], which has been especially designed to represent the difference in colour sensations experienced by human subjects [2]. Furthermore, in order to be able to point to objects, the position of objects is computed in a robot egocentric reference system (Fig. 3(f)).

The robotic setup, including the vision system and mechanisms to establish *joint attention* [21], is described in more detail in [14] and [9].

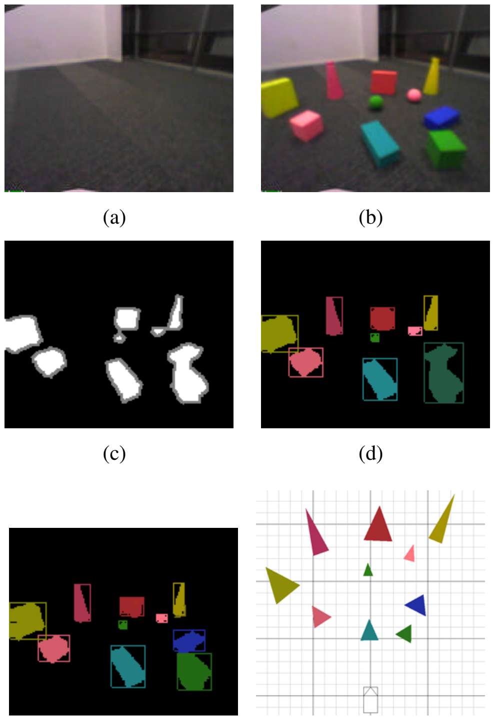


Fig. 3. The object vision system. (a): a raw camera image taken during the calibration phase. (b): a camera image of a scene containing objects. (c): the result of noise-reduced foreground/ background classification. (d): the segmented foreground regions drawn in their average colour and with bounding boxes. Note that the partially overlapping blue and green blocks in the right bottom of the original image are segmented into the same foreground region. (e): classification of foreground pixels using existing colour models. Pixels are drawn in the average colour of the most similar object model. (f): computation of colour, position and size in a robot-egocentric reference system. The width and height of objects is indicated by the width and height of the triangles.

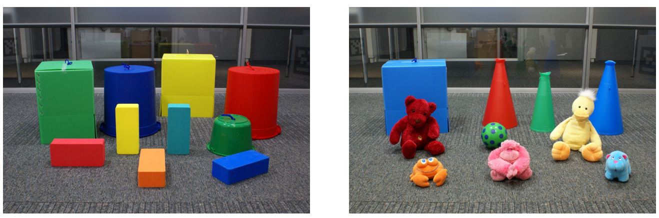


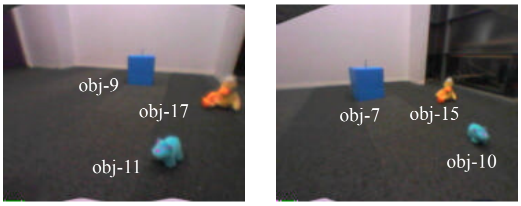
Fig. 2. Objects that were presented to the robots. Left: ten geometric objects (carton boxes, buckets, foam bricks). Right: ten toy-like objects (cones, a ball, animals).

[1]Y is the luma or brightness channel, Cb and Cr are the blue-difference and red-difference chroma components.

[2]L^* is the lightness channel, a^* is the green-red opponent channel and b^* is the yellow-blue opponent channel.

III. PERCEPTUAL DEVIATION AND STRUCTURE IN EMBODIED DATA

When moving from simulated to embodied experiments, the colour stimuli differ in two main ways. The first difference is that in embodied experiments, it is very unlikely that both speaker and hearer experience the colours of a physical object in an identical way as lighting conditions and appearances of objects may vary from the different perspectives of the robots. This is what we call *perceptual deviation*, which is illustrated in Fig. 4. The average difference in perceptual experiences for the same objects in the world across all scenes used in our experiment, is shown in Fig. 5.



	obj-9	obj-11	obj-17	obj-7	obj-10	obj-15
L^*	35.5	51.2	50.5	35.6	52.8	62.2
a^*	7.7	-17.1	26.7	7.2	-20.1	27.9
b^*	-40.7	-14.0	39.6	-39.0	-11.3	52.5

Fig. 4. Comparison between the colour perceptions of two robots for an example scene. The robots see the yellow duck (obj-17 for the left robot and obj-15 for the right robot) from different sides and distances and thus perceive very different a^* and b^* values for the same object.

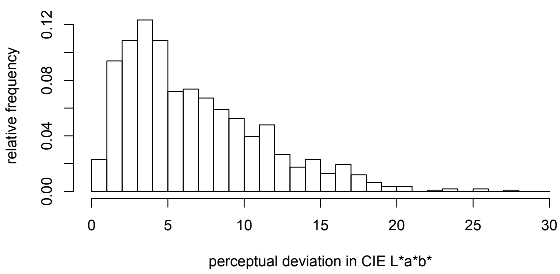


Fig. 5. A histogram of the perceptual deviation between the speaker and the hearer for the grounded data set used for the experiments reported in this paper (mean = 6.721; st. dev. = 4.575). This distribution is skewed towards lower deviations.

The second main difference between simulated stimuli and embodied data is that the colour stimuli in embodied experiments contain a higher level of *structure*. Because the number of used objects is typically limited in embodied experiments due to practical constraints (in our experiment to 20 objects), some colours do not occur at all while other colours will appear more often than others (Fig. 6). Using a one nearest-neighbour classification algorithm, we determine the relative frequency of the English colour categories [20] in the grounded data: red (.28), green (.20), purple (.12), black (.11), blue (.10), brown (.09), orange (.04) and yellow (.04). The colour categories pink, grey and white are not represented in the grounded data.

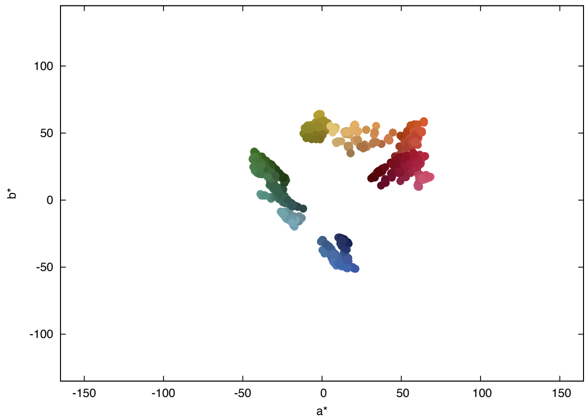


Fig. 6. The colours of all the objects in all contexts in the data set used in our experiment projected on the hue plane of the CIE $L^*a^*b^*$ colour space. The colour data in the embodied experiment is clearly structured, with more colours appearing around the actual colour of the objects used.

In contrast, artificially generated contexts usually consist of a (constrained) subset of a larger set of stimuli. This set possibly reflects the colour distributions of real-world environments based on a series of photographs, such as an urban or natural environment [1]. If no such distribution is reflected, each stimulus will as likely be represented in a shared context such as for example the set of Munsell chips which were originally used in anthropological research [7], [11] (Fig. 7). Although stimuli sets that do reflect real-world colour distributions contain more structure than those that do not, the embodied data set is far more structured as it reflects the different appearances of a limited number of objects.

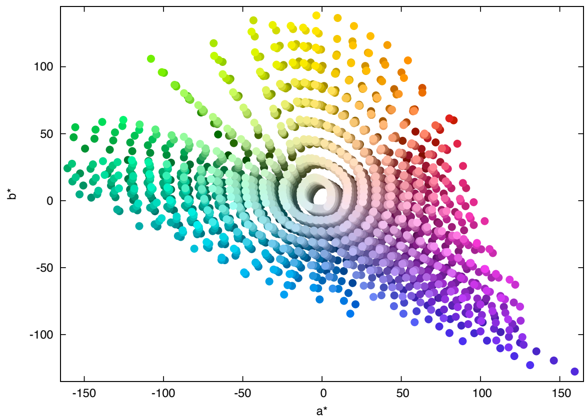


Fig. 7. A plot of all the Munsell chips that can be used to generate artificial contexts. These chips are evenly distributed in the colour space and provide a better coverage of the colour space than the embodied data.

IV. OUR MODEL

Cognitive psychologists have found prototype theory [13] to be an adequate model of human categorisation for colours. Our model is based on this theory. Each category is represented by its prototype which is a single point in colour space and represents its most representative colour. A standard one-nearest neighbour algorithm is used to classify an object as belonging to the category of which the prototype is closest to the stimulus of the object.

In colour literature, the distinction is made between the focus (the colour sample which is named fastest) and the centroid (the colour central to all colours that belong to the category) of a colour category. Interestingly, a discrepancy exists between the locations of these two (e.g. [20] for English). We use the location of the centroids when we compare our results to those reported in literature, as these resonate better with the one-nearest neighbour classification algorithm that we use.

The invention strategy is called upon when the speaker can not discriminate the topic from the context, which in this model means that another stimulus is more similar to the prototype of the category of the topic, than the topic itself. This strategy involves expanding the conceptual inventory with a new colour category that is centred on the colour of the topic, and inventing a new term to express this category in language. To allow colour categories to become established in the population, invention happens at a low rate. In all the reported experiments this rate is 0.005.

The adoption strategy is used whenever the hearer encounters an unknown word and also entails expanding the ontology with a new category that is centred on the current topic, but instead of inventing a new term the hearer will associate it with the unknown term. Note that this strategy implies that no synonyms will be present in the resulting colour lexicons.

The agents maintain a score for each category which represents its success in previous interactions. The alignment strategy at the end of a language game depends on the outcome of the game but is the same for the speaker and hearer. In the case of a successful game, the score of the used category is increased and the prototype of the used category is slightly shifted in the direction of the current topic. In the case of a failed game, the score of the category is decreased. Categories with a low score are removed from the colour lexicon.

V. DISCERNING THE IMPACT OF EMBODIMENT

We compare three environmental conditions to discern the influence of the two main differences when moving from simulated to real-world perception. In the first condition (shared simulated perception), agents will perceive artificial contexts which are sets of randomly chosen Munsell chips. In the second (shared grounded perception), both agents artificially share the same grounded perception coming from one robot body. In the third (individual grounded perception), both agents perceive the environment through their individual robot bodies.

In order to measure the influence of the structure in a grounded world, we compare the performance of conditions I and II. Basic characteristics of the scenes in condition II are carefully controlled in condition I. These characteristics are based on the set of all embodied scenes used and entail the distribution of context sizes, the total number of colour stimuli and the minimal and maximal distance between different colours within one scene/context. The better these characteristics are controlled, the better we can discern the impact of the structure in the embodied data.

To quantify the impact of the perceptual deviation between speaker and hearer, we compare conditions II and III.

VI. RESULTING DYNAMICS

The three environmental conditions are compared across four different experiment types. In the first experiment type, the baseline experiment, two agents that share a predefined colour lexicon (based the colour categories English [20]) play colour naming games without any of the three strategies (invention, adoption and alignment) activated. This experiment gives an idea how two English human subjects would perform in the three different environmental conditions. The resulting communicative success is shown in Fig. 8, which is roughly around 80%. The interactions in which the two agents fail, are those in which the topic could not be discriminated.[3] The presence of structure in the world has a positive impact on communicative success and the additional problem posed by perceptual deviation seems to have negative impact. This negative impact is rather limited, which resonates with a previous analysis of the same grounded data which indicated that colour is the least variable when different perspectives are used, unlike for example the spatial positions of the objects [24].

In the acquisition experiment, a learner needs to master a colour lexicon used by a speaker which is equipped with the same predefined colour lexicon as in the baseline experiment starting from an empty ontology. To achieve this goal, the adoption and alignment strategies are activated in the hearer. As shown in Fig. 8, this strategy results in a level of communicative success that is almost as high as in the baseline experiment after 2k games using the 8 colour terms for the colours categories that are represented in the grounded data (see Section III). This indicates that the used strategies are adequate to acquire an ontology from a teacher.

In a third experiment, the formation experiment, we were also interested in how a population can invent and coordinate its own colour lexicon. For this purpose, we enabled all three strategies, including the invention strategy, in all agents. These three strategies allow a population to develop an ontology of colour categories that is sufficiently shared to play the colour naming game successfully. Figure 8 shows the communicative success in a population of 10 agents after 10k games per agent. These agents are more successful than

[3] When all non-discriminable scenes were removed the communicative success reached 100%. The removal of these scenes reduced the impact of perceptual deviation on communicative success, as the minimal distance between the stimuli in one context increased.

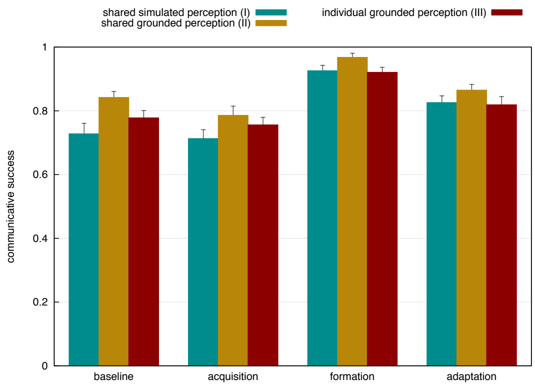


Fig. 8. Resulting communicative success of four different experiment types for the three different conditions, grouped per experiment type. In the baseline experiment, two agents with predefined and fixed colour lexicon based on the English colour categories interact. In the acquisition experiment, one learning agent needs to learn an ontology from a teaching agent which has the same predefined ontology as in the baseline experiment (2k games). In the formation experiment, a population of 10 agents invent and co-ordinate an ontology from scratch (10k games/agent). In the adaptation experiment, a population of 10 agents start out with the same predefined ontology as in the baseline experiment, but are now allowed to adapt their categories to their functional needs (10k games/agent). The results are averaged over 10 runs.

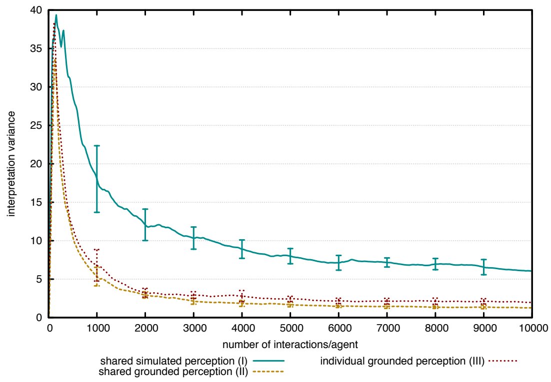


Fig. 9. The interpretation variance in formation experiments is significantly lower for the grounded conditions than for shared simulated perception. In the grounded condition it is lower when perception is shared between both interacting agents.

the ones in the baseline and acquisition experiment, mainly due to the higher number of colour categories (around 20 for conditions II and III and 25 for condition I). Fig. 9 shows the interpretation variance (the average distance between the prototypes of the colour categories associated with the same form within the population) of the population over time. It is significantly lower for conditions II and III because the structure in the environment restricts the possible location of the colour categories used by the agents. It is also slightly lower in the condition in which both agents share their perception.

Finally, we have studied the impact of having adaptive colour categories, independent of the ontology size, in the adaptation experiment. In this experiment a population of 10 agents starts out with a colour lexicon based on the English colour categories but is allowed to change these categories to its functional needs using the alignment strategy. The invention and adoption strategies are disabled. The communicative success after 10k games per agent is shown in Fig. 8. Compared to the success of the baseline experiment, an overall improvement is observed in all three conditions.

VII. COMPARISON TO HUMAN CATEGORIES

We compare the resulting ontologies to the colour categories of English [20] using two different methods: the direct comparison method and the use of naming benchmark. In the direct comparison method, we compute the distance between the two ontologies in the CIE $L^*a^*b^*$ colour space. The lower this distance, the more similar the two ontologies are. The naming benchmark consists of naming the colour chips that were consistently named by English subjects [20]. The higher the performance on this benchmark, the more similar the performance of the agents to human performance is. Both methods require a matching procedure in which each category of the resulting ontology is paired to a category of the English ontology in such a way that the pair-wise distance in the colour space is minimal.

Using these two methods, we compare the ontologies resulting from two different experiment types: the acquisition experiment and the formation experiment. To rule out the impact of ontology size on the comparison, we control the maximum number of colour categories in the formation experiment to be the number of categories that are learned in the acquisition experiment. As in the acquisition experiment the teacher only uses the colour terms that are present in the embodied dataset, which are listed in Section III, this maximum number of colour categories is set to 8. For each experiment type we compare the three environmental conditions as described in the previous section.

The results of the direct comparison method and the naming benchmark are shown in Figs. 10 and 11 respectively. Both methods show that in general, the acquisition experiments lead to ontologies that are more similar to the colour categories of English than the formation experiment. The main reason for this is that in the acquisition experiment the predefined colour lexicon of the speaker is identical to the one we compare to and hence guides the learner to categories that are similar to the English colour lexicon, whereas in the formation experiment no such guidance is present.

In the acquisition experiment, condition I seems to yield ontologies that are more similar to human colour categories than in conditions II and III, as shown in the results of the direct comparison method. The prototypes of the categories acquired by the learner are situated on the centre of all stimuli that the speaker has named using the term that is associated with that category. In conditions II and III some

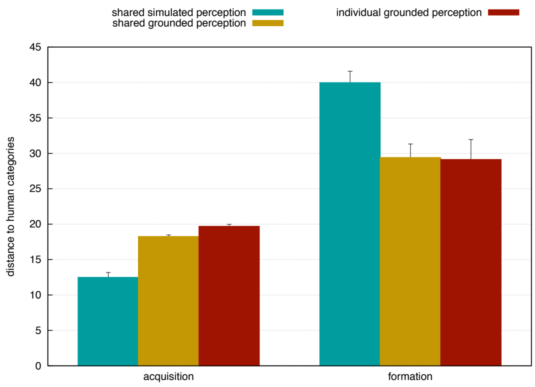


Fig. 10. Results of the direct comparison method. Two experiment types are compared in three environmental conditions, grouped per experiment type. In the acquisition experiment (2k games), one agent needs to learn the ontology of a teaching agent. In the formation experiment, a population of 10 agents invents and co-ordinates a colour ontology from scratch. The results are averaged over 10 runs.

	RE	GN	PU	BK	BL	BR	OR	YL	total
	4	22	14	3	25	4	6	8	86
	4	19	14	3	22	4	6	8	80
I	4	15.7	10.3	3	21.8	3.7	6	8	72.5
II	4	17.5	9.6	3	21.1	4	6	8	73.2
III	4	16.5	10.2	3	21.7	4	6	8	73.4
I	1.6	3	6.7	3	6.2	0	2.8	7.9	31.2
II	1.9	14.3	10.5	1.1	10.3	0.1	3.5	7.6	49.3
III	1.7	13.4	7.7	2	9.7	0.3	3.6	7.4	45.8

Fig. 11. Results of the naming benchmark, broken down by category: red (re), green (gn), purple (pu), black (bk), blue (bl), brown (br), orange (or) and yellow (yl). The top part shows the baseline performance using the centroids of the English colour ontology. The middle and bottom part show the performance of the acquisition experiment, respectively formation experiment, for the three environmental conditions.

colour categories are only partially represented, and hence the location of the prototypes acquired by the learner do not fully correspond to the locations of the prototypes of the teacher. In condition I however, all categories are fully represented, leading to a smaller difference with the ontology of the teacher. The results of the naming benchmark show no clear distinction between the three environmental conditions for this experiment.

In the formation experiment, both comparison methods suggest the same conclusion: conditions II and III seem to produce ontologies that are more similar to English colour categories than condition I. Although in this experiment no guiding teacher is present, the structure in the grounded data partially takes over the guiding role of the teacher. The colours of the objects presented to the robots (Fig. 2) are better examples of the basic colour categories for English than the stimuli in the simulated world in which no such structure is present.

VIII. Conclusion

We have shown that our model for the colour naming game is robust to overcome the main difficulties arising from embodiment in an experiment using humanoid robots. In embodied experiments, speaker and hearer perceive the world from a different perspective and hence experience the colours of the objects around them differently. In our experimental setup, the impact of this perceptual deviation is rather limited as the lighting conditions are constant in the office environment. In future experiments this influence is expected to increase when objects are used that have different colours when observed from a different perspective or when local lighting conditions are used.

We have attested the positive impact of adapting categories to the functional needs of the agents on the resulting communicative success, even when compared to an experiment in which static categories are completely shared within a population. This finding resonates with previous studies in experimental psychology [5] in which it is shown that humans align their ontologies when interacting with each other, even in the course of a single dialogue.

As the resulting ontologies reflect the structure of the environment in which they are developed, these ontologies will bear more resemblance to the colour categories of English when the environment consists of objects that are good examples of these categories than in an environment in which the colours are uniformly distributed over the colour spectrum.

IX. Acknowledgments

This work is funded through a fellowship of the Institute of the Promotion of Innovation through Science and Technology in Flanders (IWT-Vlaanderen), with additional funding from the EU FP7 ALEAR project and Sony Computer Science Laboratory in Paris.

References

[1] T. Belpaeme and J. Bleys, “Explaining universal colour categories through a constrained acquisition process,” *Adaptive Behavior*, vol. 13, no. 4, pp. 293–310, 2005.

[2] D. H. Brainard, “Color appearance and color difference specification,” in *The Science of Color*, S. Shevell, Ed. Optical Society of America, 2003, pp. 191–216.

[3] S. Coradeschi and A. Saffiotti, “An introduction to the anchoring problem,” *Robotics and Autonomous Systems*, vol. 43, no. 2-3, pp. 85–96, 2003.

[4] M. Fujita, Y. Kuroki, T. Ishida, and T. T. Doi, “Autonomous behavior control architecture of entertainment humanoid robot SDR-4X,” in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS ’03)*, Las Vegas, Nevada, 2003, pp. 960–967, vol. 1.

[5] S. Garrod and G. Doherty, “Conversation, co-ordination and convention: An empirical investigation of how groups establish linguistic conventions,” *Cognition*, vol. 53, pp. 181–251, 1994.

[6] R. E. Kalman, “A new approach to linear filtering and prediction problems,” *Transactions of the ASME-Journal of Basic Engineering*, vol. 82, no. 1, pp. 35–45, 1960.

[7] P. Kay, B. Berlin, L. Maffi, and W. R. Merrifield, *The World Color Survey*. Stanford: Center for the Study of Language and Information, 2003.

[8] S. Kirby, “Natural language from artificial life,” *Artificial Life*, vol. 8, no. 2, pp. 185–215, 2002.

[9] M. Loetzsch, M. Spranger, and L. Steels, “The grounding of object naming and identity,” 2010, submitted.

[10] M. Loetzsch, R. van Trijp, and L. Steels, "Typological and computational investigations of spatial perspective," in *Modeling Communication with Robots and Virtual Humans*, ser. Lecture Notes in Computer Science, I. Wachsmuth and G. Knoblich, Eds. Berlin: Springer, 2008, vol. 4930, pp. 125–142.

[11] R. E. MacLaury, *Color and Cognition in Mesoamerica*. Austin: University of Texas Press, 1997.

[12] A. Puglisi, A. Baronchelli, and V. Loreto, "Cultural route to the emergence of linguistic categories," *Proceedings of the National Academy of Sciences*, vol. 105, no. 23, pp. 7936–7940, 2008.

[13] E. Rosch, "Natural categories," *Cognitive Psychology*, vol. 4, pp. 328–350, 1973.

[14] M. Spranger, "World models for grounded language games," Diploma thesis. Humboldt-Universitt zu Berlin, 2008.

[15] L. Steels, "The origins of syntax in visually grounded robotic agents," *Artificial Intelligence*, vol. 103, no. 1-2, pp. 133–156, 1998.

[16] ——, "Language games for autonomous robots," *IEEE Intelligent Systems*, vol. 16, no. 5, pp. 16–22, 2001.

[17] ——, "Semiotic dynamics for embodied agents," *IEEE Intelligent Systems*, vol. 21, no. 3, pp. 32–38, 2006.

[18] L. Steels and T. Belpaeme, "Coordinating perceptually grounded categories through language. A case study for colour," *Behavioral and Brain Sciences*, vol. 24, no. 8, pp. 469–529, 2005.

[19] L. Steels and M. Loetzsch, "Perspective alignment in spatial language," in *Spatial Language and Dialogue*, K. R. Coventry, T. Tenbrink, and J. A. Bateman, Eds. Oxford University Press, 2009, to appear.

[20] J. Sturges and T. A. Whitfield, "Locating basic colours in the Munsell space," *COLOR Research and Application*, vol. 20, no. 6, pp. 364–376, 1995.

[21] M. Tomasello, "Joint attention as social cognition," in *Joint Attention: Its Origins and Role in Development*, C. Moore and P. J. Dunham, Eds. Hillsdale, NJ: Lawrence Erlbaum Associates, 1995.

[22] P. Vogt, "Anchoring of semiotic symbols," *Robotics and Autonomous Systems*, vol. 43, no. 2-3, pp. 109–120, 2003.

[23] K. Wagner, J. A. Reggia, J. Uriagereka, and G. S. Wilkinson, "Progress in the simulation of emergent communication and language," *Adaptive Behavior*, vol. 11, no. 1, pp. 37–69, 2003.

[24] P. Wellens, M. Loetzsch, and L. Steels, "Flexible word meaning in embodied agents," *Connection Science*, vol. 20, no. 2, pp. 173–191, 2008.

[25] L. Wittgenstein, *Philosophische Untersuchungen*. Frankfurt: Suhrkamp, 1967.