Scheduling parental behavior helps to acquire perspective transformations

N. M. Mayer, Tomoyasu Uragi, Yuichiro Yoshikawa, Takashi Minato, Hiroshi Ishiguro, JST ERATO Asada Project, Osaka University, 2-1 Yamadaoka, Suita, Osaka

We study the parent-child relationship by using a simple pattern of language, while the relative position of the caregiver and the caretaker change during the game. The aim is to model the joint learning of the language and the vision at the same time. In the following we consider a grid shaped world as sufficient for our problem. The following situation is assumed. The caregiver hints to an object by using the relation of his own body to the object, as left, right, far, near. Alternatively, the caregiver can use the caretakers coordinate system to express the relative position of the object. The visual input to the caretaker gives a hint to the position of the object and the position of the caretaker. The caretaker sits either on the same side as the caregiver or on the opposite side.

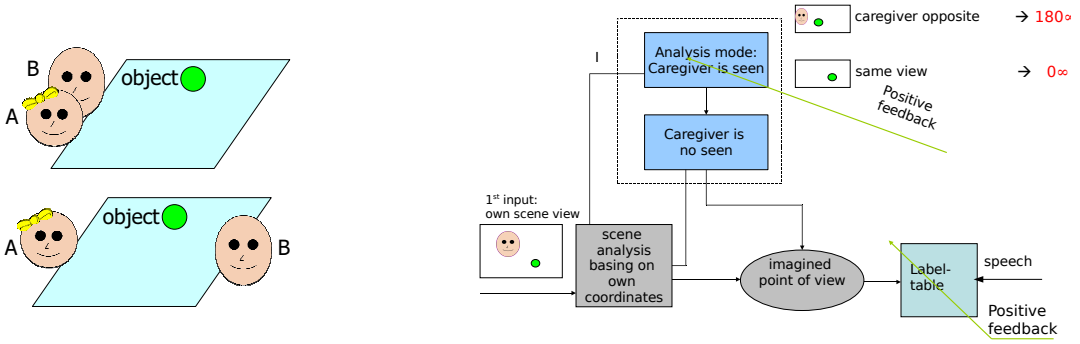
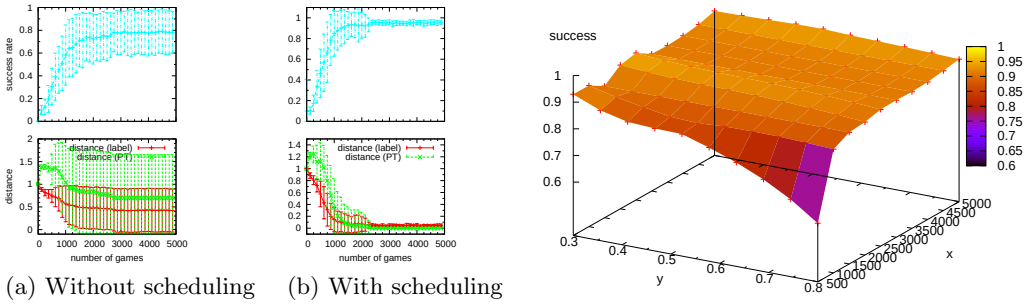


Figure 1: Left: Assumed experimental setting, right: Recognition system to classify right and left in the respective relation.

The language for the model is: The caregiver choses to say one of the following sentences (my or babies; left or right; near or far object: total $2^3 = 8$ different expressions). In addition, the caretaker has the visual input. It is assumed that the baby reacts in a way that indicates that he/she assumes a consistent relation between the language and the visual input. If the relation is correct the caretaker gives a positive feedback. The actual model learns from correlation learning. It is assumed that the actual crossmodel recognition and language learning of the object on one hand and the anticiption of the caretakers perspective on the other hand is performed in two distinct cortical areas. Since initially the point of view of the caregiver is not accessible (it has to be learned by the caretaker to account for the caregivers point of view) the language input is ambiguous in relation to the visual input. Thus, in order to construct the correct relations between language and visual input the anticipation of the caregivres point of view has to be learned in parallel. We derived a model that accounts for the parallel learning of both modules, of which we present the result in the present poster. In addition, a kind of sequential increase of complexity of the teachers policy is introduced (in the following "sequential learning").

The basic idea is here that the child initially filters his/her experience and takes only the experiences for learning that stem from collateral learning. In addition, scheduling is based on the idea that parental relationship and position on the use of perspective. It is now assumed that the learning that undertaken in the collateral setting as initially dominant. Later, more and more the contralateral setting is added. Fig. 2 shows the results for the performance of the caretaker as the percentage in which he/she interprets the situation correctly and gets positive feedback from the caregiver. It is clear to see that the scheduling improves the performance. In addition we investigated how different scenarios affect the performance. We assume the scheduling of the following shape $p = y/x \times t$ if $t < x$ and $p = y$ if $t > x$, where t is the time, p is the probability that the caregiver sits on the opposite side to the caretaker. The left side of Fig. 2 shows the results of the performance for different values of x and y. As one can see, only very extreme values of x and y lead to a significantly reduced performance, whereas for a wide range of values the performance is relatively stable independent of the exact choice of these parameters. **Related work:** Already earlier studies relate to this field of interest [1] but it seems difficult to prove it in experiments with real infants. In this context the Language Game by Steels and Loetzsch [2] has to be mentioned. In that approach two robots explain scenes to each other. In particular, they explain the movement and transfer of objects. The experimenters account for the change of the perspective, however in a straight forward engineering way, that does not account for learning processes. In addition, during the whole game the agents sit opposite to each other and thus a fixed meaning experessions as right or left is guaranteed. The meaning of words that indicate the relative position (like left and right) is already known. In addition, they do not consider the positional relationship between the body and face, which is a part of the baby's normal experience. Finally, different from the present approach both agents are learning (no caregivers).



(a) Without scheduling (b) With scheduling

Figure 2: Results: Left: Scheduling leads for the given setting better results than the non-scheduling scenario. Right: Assuming different scheduling scenarios the performance is plotted as a function of the two parameters x and y.

[1] Minoru Asada, Karl F. MacDorman, Hiroshi Ishiguro, and Yasuo Kuniyoshi. Cognitive developmental robotics as a new paradigm for the design of humanoid robots. *Robotics and Autonomous Systems*, 37(2–3):185–193, 2001.

[2] Luc Steels and Martin Loetzsch. Perspective alignment in spatial language. In Kenny R. Coventry, Thora Tenbrink, and John. A Bateman, editors, *Spatial Language and Dialogue*. Oxford University Press, 2007. to appear.