A computational model of social-learning mechanisms

Manuel Lopes[†], Francisco S. Melo[†♭], Ben Kenward[‡], José Santos-Victor[†]

[†] *Institute for Systems and Robotics, Instituto Superior Técnico, Lisboa, Portugal*

[‡] *Department of Psychology, Uppsala University, Uppsala, Sweden*

[♭] *School of Computer Science, Carnegie Mellon University, Pittsburgh, USA*

Abstract

In this paper we propose a computational model that describes how observed behaviour can influence an observer's own behaviour, including the acquisition of new task descriptions. The sources of influence on our model's behaviour are: beliefs about the world's possible states and actions causing transitions between them; baseline preferences for certain actions; a variable tendency to infer and share goals in observed behaviour; and a variable tendency to act efficiently to reach rewarding states. Acting on these premises, our model is able to replicate key empirical studies of social learning in children and chimpanzees. We demonstrate how a simple artificial system can account for a variety of biological social transfer phenomena, such as goal-inference and over-imitation, by taking into account action constraints and incomplete knowledge about the world dynamics.

Keywords: social learning; imitation; emulation; computational model

1 INTRODUCTION

The behaviour of other individuals is a crucial source of information for social animals and particularly for young children. On a physical level, there are two sources of information available when observing an individual act – the motor patterns the individual performs, and the outcome of the actions. Another information source is the intention behind the behaviour, which may be inferred from the actor's choices among possible actions. Knowledge of how the world works and expectations about others' normal behaviour are usually necessary for extracting useful information from such observations. Different social learning processes exploit these different information sources to different degrees (Call & Carpenter, 2002).

Two broad categories of social learning, focusing on different kinds of information, are emulation and imitation (Call & Carpenter, 2002; Whiten et al., 2004). In *emulation*, the observer learns about results and changes that can be accomplished in the environment, and sets about to replicate such states and changes, not necessarily paying heed to the specific observed motor patterns. In *imitation*, the observer copies the specific motor patterns and consequent results that are jointly inferred to have been part of the behaviour intention. Because imitation,

unlike mimicry, is defined as goal-directed, not every part of an action sequence is necessarily copied – for example, one would not generally copy a cough when repeating a spoken sentence.

Young children and apes are able to both imitate and emulate, but utilise the different strategies to different extents (Whiten et al., 2004; Want & Harris, 2002; Tennie et al., 2006). Tasks where more than one action can achieve the same effect can be used to confirm that subjects do imitate specific motor patterns. For example, chimpanzees and two-year-old children copy a demonstrator's choice of a push or twist action to remove a bolt to open a box (Whiten et al., 1996). Children can be selective about which actions should be imitated (Gergely et al., 2002; Williamson & Markman, 2006), but in general they are rather prone to imitate even parts of action sequences that are not obviously necessary to achieve the goal – a phenomenon known as *over-imitation* (Horner & Whiten, 2005). Over-imitation can be diminished by reducing social cues (McGuigan et al., 2007; Brugger et al., 2007) or by increasing the urgency of task completion (Lyons et al., 2005), and it has been argued that it may occur for a variety of social reasons (Nielsen, 2006), or because the observers encode the demonstrator's actions as causally meaningful (Lyons et al., 2005).

Other species, such as dogs, have also been shown to switch strategies after having observed a demonstration (Range et al., 2007). The aforementioned studies have identified distinct behavioural patterns of social learning, but little is known about the conditions that prompt these behaviours, the underlying neural mechanisms that explain them or even how the switching between them is controlled.[1]

The goal of this study is to provide a simple computational model that may allow biologists and psychologists to plan new experiments leading to a deeper understanding of these mechanisms. To this purpose, we provide a computational framework for the different behaviour models suggested in the literature that accounts for salient aspects of social influence, replicating key empirical results. We argue that the ability of our model to replicate different classes of behaviour by making simple trade-offs between the different "sources of information" available to the learning agent provides a significant contribution toward a parsimonious interpretation for these classes of behaviours. We also discuss several predictions from our model that may suggest interesting new experimental paradigms.

We note that there are other mechanisms of social learning, such as stimulus enhancement, that are cognitively simpler and therefore of less interest to cognitive psychology, but that can also confer evolutionary advantages (Noble & Franks, 2002). Melo et al. (2007) model some of these simpler social learning mechanisms using a somewhat similar formalism.

2 MODEL

We begin by giving a summarised description of our model of an individual (human or otherwise) observing and performing behaviour (see Fig. 1). We provide only a sketch of the algorithm and refer to Appendix A for further technical details.

[1] For a study of the brain regions involved in action understanding in typical and atypical situations, we refer to the work of Brass et al. (2007).

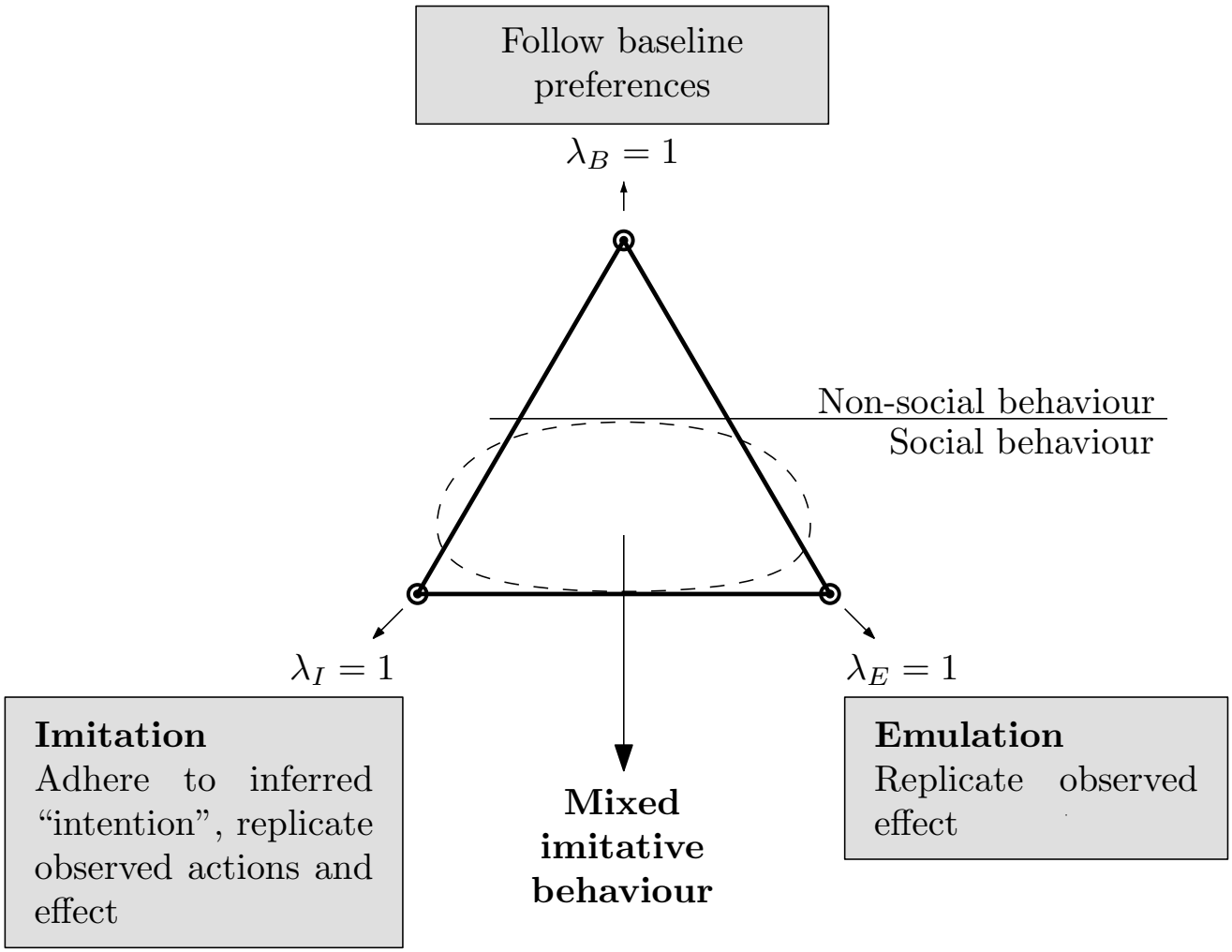


Figure 1: "Strategy weighting triangle", representing the combination of several simpler behaviours: Non-social behaviour, emulation and imitation. The line separates behaviour that appears to be socially influenced from behaviour that does not, but does not necessarily correspond to the agent's reasoning.

The demonstrator and observer generally act within the same world (for exceptions see below), which can be in a number of possible states, but only in one state at any one time. Transitions between states are caused by actions that the demonstrator takes during demonstration and the observer takes after exposure to the demonstration. These possible states and transitions are predetermined and constant during the demonstrator's or observer's actions. The observer has knowledge of all the possible states and transitions of the world it acts in. Incomplete world knowledge is simulated by certain possible real world states or transitions being absent from the world model that the observer acts in (see below).

Our model was kept as simple as possible while being capable of reproducing key biological results. It takes into account the agent's baseline preferences for different actions and the information available from the demonstration. Specifically, it considers the end-effect of the demonstrated actions as well as the possible intentions of the demonstrator (inferred from the demonstrated actions). The model also takes into account the learning agent's knowledge (even if imperfect or incomplete) about world dynamics and possible state transitions. We do not consider explicitly the way this knowledge can be acquired, but rather assume that knowledge about the world, be it incomplete or imprecise, is acquired prior to or as a consequence of observing the demonstration.

Each of the aforementioned sources of information (baseline preferences, end-effect, and inferred intention) is processed by the observer in a specific "module". For any given world state, each module computes a preference score for each possible course of action. The list of preference scores is called a utility function and generally denoted using the symbol Q. For example, the end-effect replicating module would rank actions in the following descending order of preference: an action leading directly to the final effect; an action that can lead indirectly to the final effect; an action that makes the final effect unreachable. The modules process information as follows:

- The module addressing the baseline preferences of the agent evaluates actions in terms of energy consumption, which it prefers to minimise. So, for example, this module always prefers to perform "no action" than any other action. The utility function Q_B associated with this module therefore ranks possible action sequences according to their overall energy consumption.

- The end-effect replicating module computes a utility function Q_E that evaluates the actions in terms of their probability of reproducing the observed result/effect. In our simulations, this effect is always taken as the final state observed in the demonstration, and this module will select the sequence of actions minimizing the number of steps until this final state is reached. In particular, it need not select the same actions observed in the demonstration.

- Finally, for the intention replicating module, the utility function Q_I is more complex to compute, since it involves inferring the demonstrator's intended goal. We therefore describe this module in more detail. It infers the intention behind the demonstration using a teleological argument, by assuming the demonstrator is *goal-oriented* and is thus trying to fulfil some particular goal. The demonstrator's goal is assumed to be one or more desired states and/or transitions between them. Notice that this is a rather broad definition of goal, that may also encompass actions (*i.e.,* state transitions) for their own sake, independently of the states they reach. This broad definition of goal allows the possibility that our model imitates actions without understanding the deeper purpose behind them, in cases in which it infers only actions themselves to be the intended part of the demonstration.

  The module operates by "enumerating" all the possible goals in the current system, calculating for each one the relative probability that it would give rise to the demonstrated behaviour, and choosing the one that maximises this probability. The module's calculated utility function therefore ranks the actions with respect to the most likely goal, given the demonstration.

To illustrate the interaction between the different elements in our learning model, consider the simple example depicted in Fig. 2. In this example, the system consists of only two states, X and Y, the transition between which is triggered by any of the actions of the agent. Supposing that the demonstration consists of action A, let us analyse the output of each of the modules in our model in both scenarios in Fig. 2.

The module addressing the baseline preferences would simply output a ranking of the two actions. For example, if the baseline preference stated that the agent preferred action B to action A, then we could have $Q_B(\cdot, A) = 0$ and $Q_B(\cdot, B) = 1$.

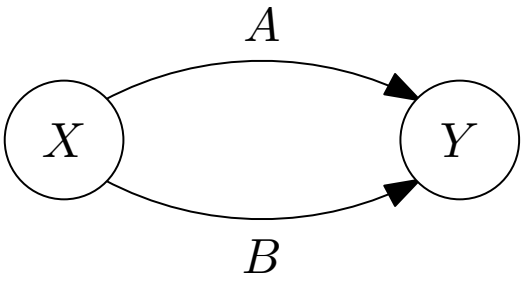


Figure 2: Simple example scenario in which the world can be in one of two states, X and Y, and the agent can trigger the transition between these by choosing either of the available actions.

The end-effect replication module, in this case, merely states that the agent should reach the final state (Y). If both actions are equally successful in achieving that, then this means that $Q_E(\cdot, A) = Q_E(\cdot, B) = 1$. In this case, the effect-replication module does not bias the action choice toward any of the two possible actions. It is interesting to note that the same would not hold if, for example, action A only succeeds in achieving the transition with 0.9 probability. In this case, the end-effect replication module would output $Q_E(\cdot, A) = 0.9$ and $Q_E(\cdot, B) = 1$.

Finally, the intention replication module would, in this case, output $Q_I(\cdot, A) = 1$ and $Q_I(\cdot, B) = 0$, translating the fact that the agent finds the goal "reach state Y using action A" to be more likely than merely "reach state Y", because action B could have been used, but was not.

Note that intentions as inferred by this module may best correspond to either motor intentions or prior intentions (Carpenter et al., 2002; Searle, 1983) or a combination of the two, depending on the specific case. When only an action is inferred to be the intention, it corresponds best to the concept of motor intention, but when reaching a particular state is inferred to be part of the intention, then it can be seen as modelling a prior intention. This aspect of model interpretation is complicated by the fact that prior intentions exist on different levels, for example, the prior intention behind the motor intention to push a switch may be simply to move the switch from one position to another, or it may also be to turn on a light. We take this into account when discussing our results.

We refer to Appendix A and to the supplementary material for further details on how each module computes the corresponding utility function.

The three sets of behaviour preferences (*i.e.,* the three utility functions) are combined to yield a final utility function, Q_{out}, defined as

$$Q_{\text{out}} = \lambda_B Q_B + \lambda_E Q_E + \lambda_I Q_I,$$

where λ_B, λ_E and λ_I are three positive weights verifying $\lambda_B + \lambda_E + \lambda_I = 1$. The behaviour the agent actually performs is simply the preferred behaviour as defined by the utility function Q_{out}.

Figure 1 provides a pictorial description of the proposed model. Each vertex of the "strategy weighting triangle" corresponds to the behaviour computed by one of the modules described above. The value of the three parameters λ_B, λ_E and λ_I can be chosen in order to differently weight the contribution of the corresponding behaviours' to the final one. The three extreme behaviours are:

- Following baseline preferences, thus ignoring the demonstration (non-social behaviour);

- Emulation, where the agent replicates the end-effect of the observed actions; and
- Imitation, where the agent replicates the inferred goal/intention of the demonstrator.

The inference algorithm used in the intention replication module samples the space of possible goals in the world system, computing their likelihoods given the observed demonstration. However, in many situations there may be several different goals that are equally likely to produce the observed demonstration. In such cases, goals with tied probability are ranked randomly, which, when combined with the utility functions from the other modules (that have no random component), leads to stochasticity in the final performed behaviour, as will be apparent in the next section. To compensate for this stochasticity we perform $1,000$ simulation runs for each condition and for each modelled experiment.

Finally, we note that different choices of parameters will lead to different combinations of the resulting behavioural preferences computed by each of the three modules and, thus, to different final behaviours. We do not propose a method for choosing these weights, but observe that, in general, their choice for each particular individual will depend on its social, environmental and internal context.

3 SIMULATIONS

In this section we model three well-known social learning experimental paradigms to assess how well our model can replicate the corresponding results. We also perform a simulation that does not correspond to any existing experimental paradigm.

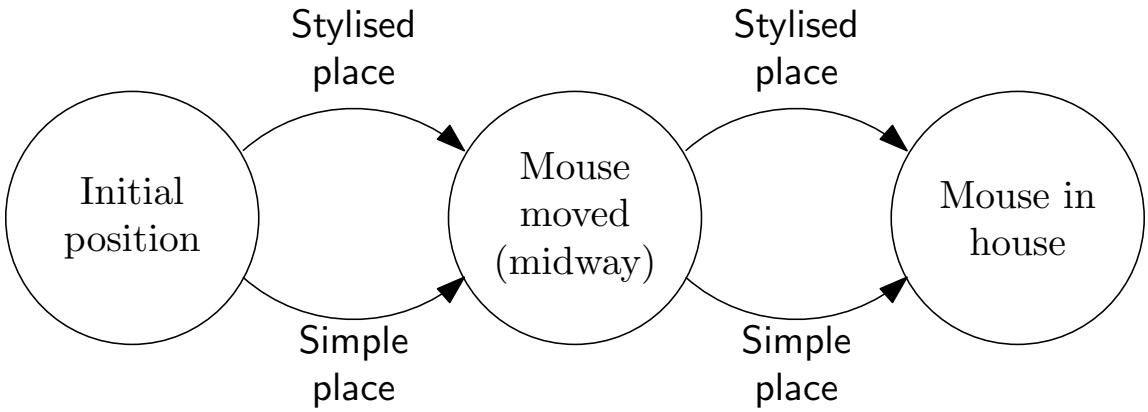
3.1 Imitation of the inferred intentions of observed behaviour

We begin by demonstrating that our model can replicate the tendency of primates to interpret and reproduce observed actions in a teleological manner – that is, in terms of the inferred goals of the action (Csibra & Gergely, 2007). For example, Bekkering et al. (2000) allowed 3- to 6-year-old children to observe a demonstrator reaching across her body to touch a dot painted on a table to one side of her, using the hand on her other side. Children tended to copy the dot touching action, but not the use of the contra-lateral hand. However, when the same action was performed without a dot, the children's tendency was to imitate the use of the contra-lateral hand. In the first case, the children interpreted dot touching as the intention, and therefore chose their own easier way to touch the dot. In the second case, as there was no clear target of the action, the action itself is interpreted as the intention and is therefore imitated more faithfully.

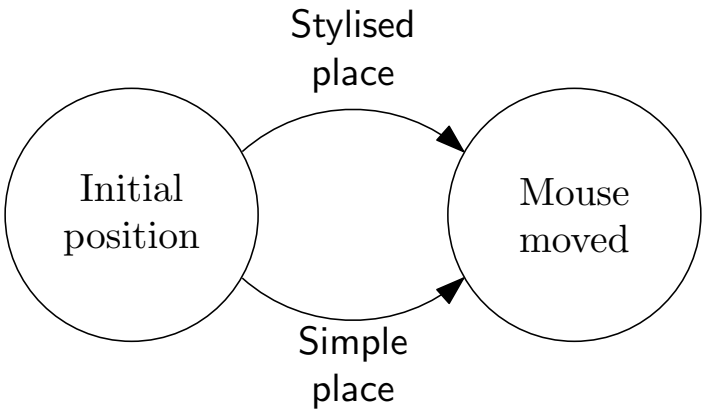
Carpenter et al. (2005) designed an experiment with the same logic but adapted for infants. A demonstrator moved a toy mouse across a table from one point to another, using either a "simple" action or a "stylised" action (i.e. placing the mouse at a particular location by hopping or sliding). In one condition, the final point of the move was inside a little house, and in the other condition, no house was present. Similar to the older children in the study of Bekkering et al. (2000), and presumably for similar reasons, the 14- and 18-month-old tested in this

experiment showed a much greater tendency to replicate the specific mouse moving action observed when there was no house to move the mouse into. We present our simulation in terms of the Carpenter et al.'s (2005) study, but the results are generalisable.

Figure 3 represents the world dynamics for this problem. We assume the mouse to be in an initial, resting state, from which it can transition into the "moved" state using either a stylised or a simple action. In the house condition there is also an "in house" state that can be reached from the "moved" state. For the no-house condition there is no "in house" state, only the "initial" and "moved" states. In each simulation, the observer is exposed



(a) Condition with the toy house.



(b) Condition without the toy house.

Figure 3: World model for the first set of simulations: (a) Condition with the toy house and (b) condition without the toy house. Circles represent world states and arrows represent the transition between them. The arrows are labelled according to the action inducing the transition. We omitted the "No action" possibility from the diagram, as it does not induce any state transition.

to one demonstration (of the stylised action) and then is allowed to act. The observer has a baseline preference (derived from energetic considerations) for using a "simple" over a "stylised" place. It can also choose to do nothing

("No action") and has a baseline preference for the latter option over the former two. In the house situation the end-effect module uses the simple action to reach the final state, but, for the non-house situation, the end-effect module is irrelevant.

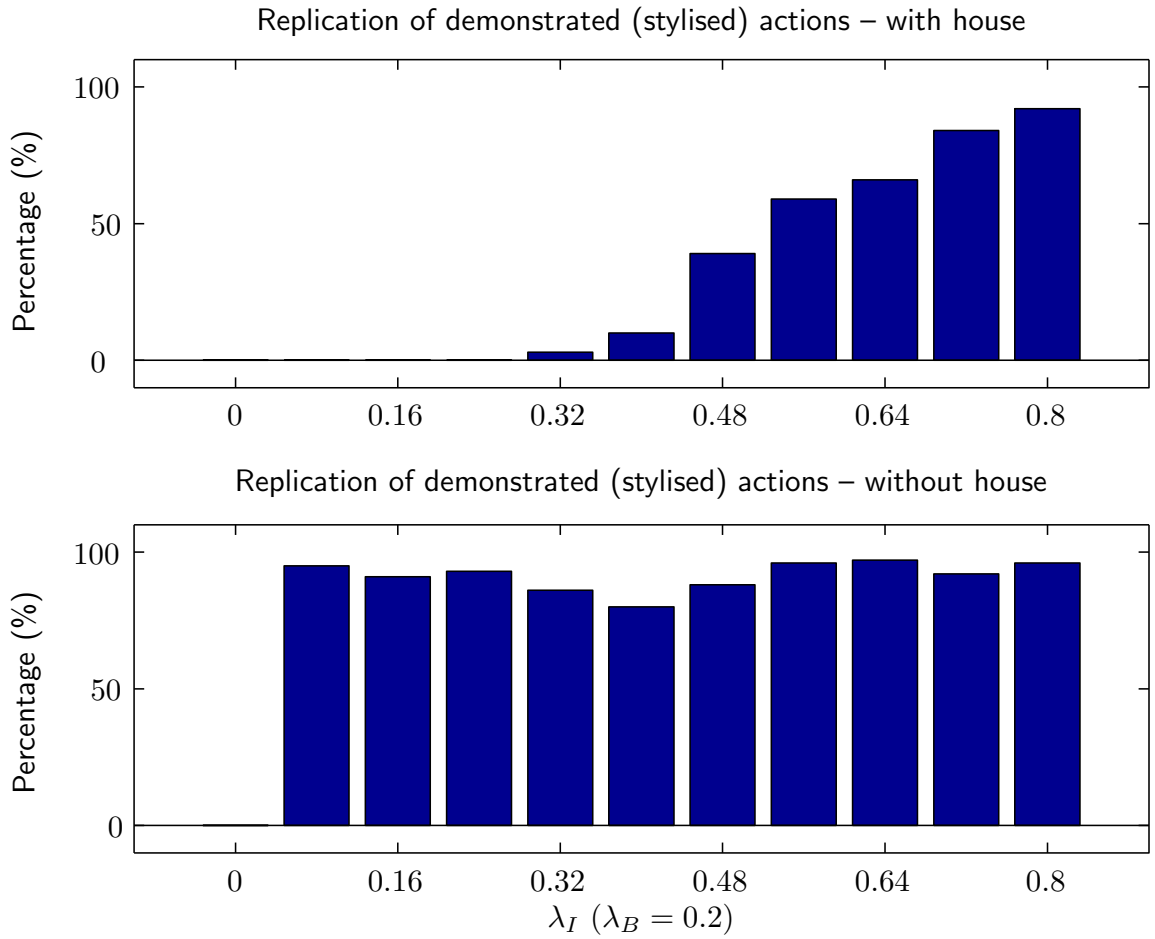


Figure 4: Percentage of simulation runs in which the modelled agent replicates the demonstrated stylised action as λ_I (the weight of the intention replicating module) is increased, in the house and no-house conditions. Whenever the stylised action was not replicated, the simple action was performed. The weight of the baseline preference module, λ_B, is kept constant with a value of 0.2. Recalling that $\lambda_I + \lambda_B + \lambda_E = 1$, λ_E (the weight of the end-effect replicating module) decreases to 0 as λ_I increases to 0.8.

The results can be found in Figure 4. In all results shown in this work we perform a variation of the parameters and evaluate the resulting behaviour. In this case we see what happens when increasing the tendency to follow the inferred intention of the demonstration (λ_I) while reducing the tendency to replicate the end-effect (λ_E). In the "with house" condition the probability of choosing the demonstrated action increases with λ_I. In the "no house" condition the resulting behaviour is usually to faithfully imitate the demonstration. The only parameter values at which the empirical results are not replicated are when λ_E is close to zero – in other words when the agent gives no weight to the final observed effect. This results reproduces the findings of Carpenter et al. (2005) and therefore confirms the logic of the standard interpretation of this experiment – the results can be explained by the assumption that the infant infers what the demonstrator's intention was, adopts the same intention, and imitates only as much as is necessary to achieve it.

3.2 Sensitivity to action constraints in goal-directed imitation

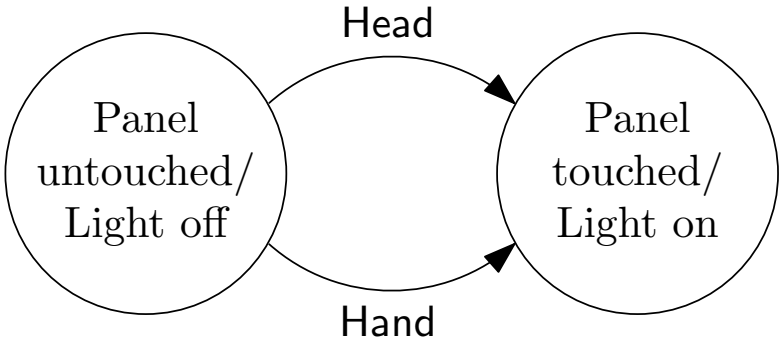
In an experiment originally designed to test infants' memories of novel actions, Meltzoff (1988) exposed 14-month-olds to a demonstrator who performed unusual actions on objects, and found that the infants reproduced the actions when presented with the objects a week later. One of the objects was a box with a panel that lit up when the demonstrator touched it with his forehead, and most infants copied the use of the forehead rather than using their hand. Gergely et al. (2002) extended this experiment by including a condition in which the demonstrator was restricted and could not use her hands because she was holding a blanket wrapped around herself. In this case, only 21% of the infants copied the use of the forehead, whereas in a control condition replicating the study of Meltzoff (1988) without a held blanket, 69% of the infants copied the use of the forehead. Possibly, in this latter case, infants detect no constraints upon the demonstrator's action and therefore encode the use of the forehead as a specific part of the intention, whereas in the restricted case, they detect the constraint as a non-task related reason for the use of the forehead and as such do not encode the specific action as part of the intention.

We simulate this experiment, using world models that reflect the different possible transitions in the constrained and unconstrained conditions (Fig. 5). There is a baseline preference for using the hand over using the head to contact the panel. Again, the observer can also choose to do nothing ("No action") and has a baseline preference for the latter option over the former two. In this experiment, the final effect module does not distinguish between the two possible ways of activating the switch, while the imitation module prefers the head in the unrestricted condition and is indifferent between the two in the restricted condition. This is due to the fact that the constraints of the demonstrator were taken into account when inferring the intention.

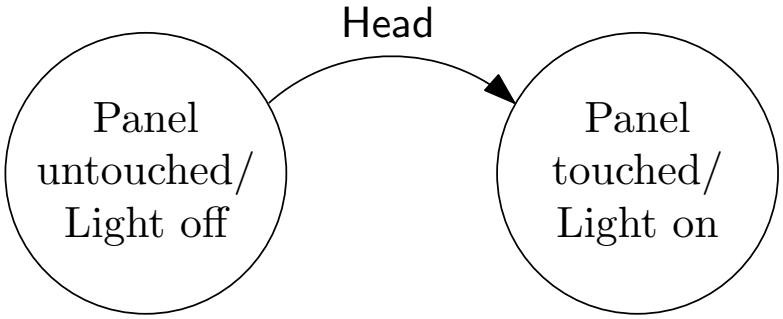
Again, the simulation results of Figure 6 closely replicate those from the empirical study. Unless λ_I (the tendency to replicate the inferred intention) is zero, the use of the head is more likely to be copied in the hands-free condition. The reason is that, in the hands-free condition, head use tends to be classed as part of the intention because it was chosen over a possible alternative, whereas in the restricted condition there was no alternative.

Our simulation therefore confirms the logic of part of Gergely et al.'s (2002) analysis of their empirical results, both in terms of imitation of the inferred intention and of sensitivity to the constraints on the demonstrator. Note, however, that Gergely et al. (2002) also go a step further in their interpretation – they suggest that in the unrestricted condition the infants "may have inferred that the head action must offer some advantage in turning on the light". There are two ways in which this goes beyond the simplest logic necessary to explain the results, as demonstrated by our model. Firstly, it is not necessary to assume that the infants believed that the method used offered an advantage. Our model replicates the results by inferring the intention to act in a certain way, but it does not infer anything about the underlying motive for the demonstrator's choice of intended action. It is therefore possible that in cases such as this, infants may imitate intended actions without necessarily making any inferences about why those actions may or may not be effective. Note however that in cases where causality is more transparent, infants may make such inferences (Brugger et al., 2007).

Secondly, it is not even necessary to assume that the inferred and adopted intention was to turn on the light. The inferred intention may have been the prior intention of turning on the lights, but a simpler and sufficient



(a) Unrestricted condition.



(b) Restricted condition.

Figure 5: The world model representing the experiment of Gergely et al. (2002). In the restricted condition, the action "Use hand" is not available, representing the fact that the agent is assumed to appreciate that hand use is not possible in this situation. The observer does not operate under constraint even after having observed the constrained condition. As before, we omitted the "No action" possibility from the diagram, as it does not induce any state transition.

interpretation of both our model and the empirical result is that the intention was simply the motor intention of contacting the panel (for recent demonstrations of how infants have difficulty motivating behaviour by knowledge of such arbitrary contingencies, see Klossek et al., 2008; Kenward et al., 2009).

3.3 Sensitivity to imperfect knowledge

The simulations in Subsections 3.1 and 3.2 replicated experiments in which, in the right conditions, children copy faithfully a demonstrated action, even if it is not necessary to achieve the desired end state. These results were replicated most accurately at intermediate values of λ_I (the tendency to replicate the inferred intentions) – when this parameter is close to 1, the observed action sequence was almost always copied faithfully even when it is plausible that specific action choices were not an integral part of the intention.

To investigate what happens when the learner does not have complete knowledge of the world dynamics, we now model a type of experiment that has been designed to further investigate the imitation/emulation balance in different circumstances and ages, and also comparatively with chimpanzees.

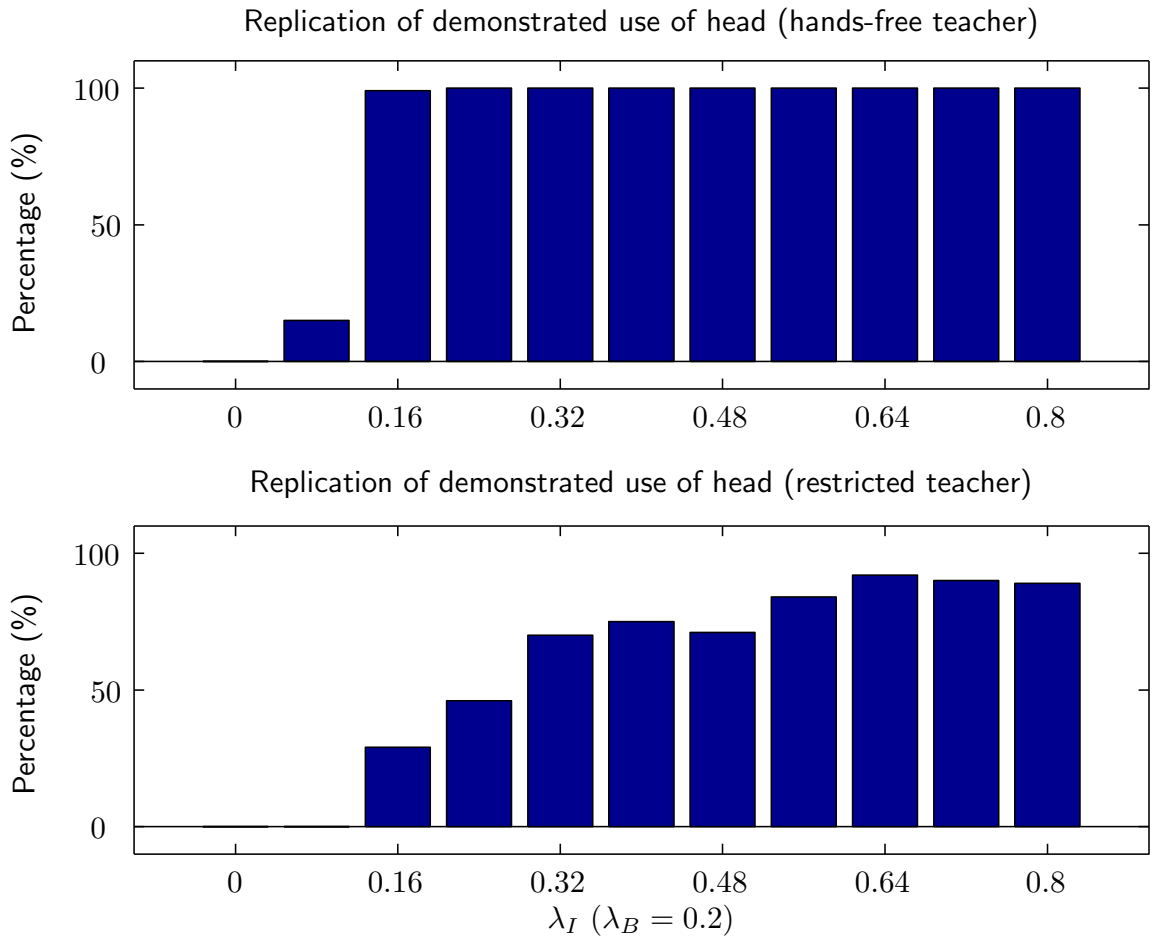
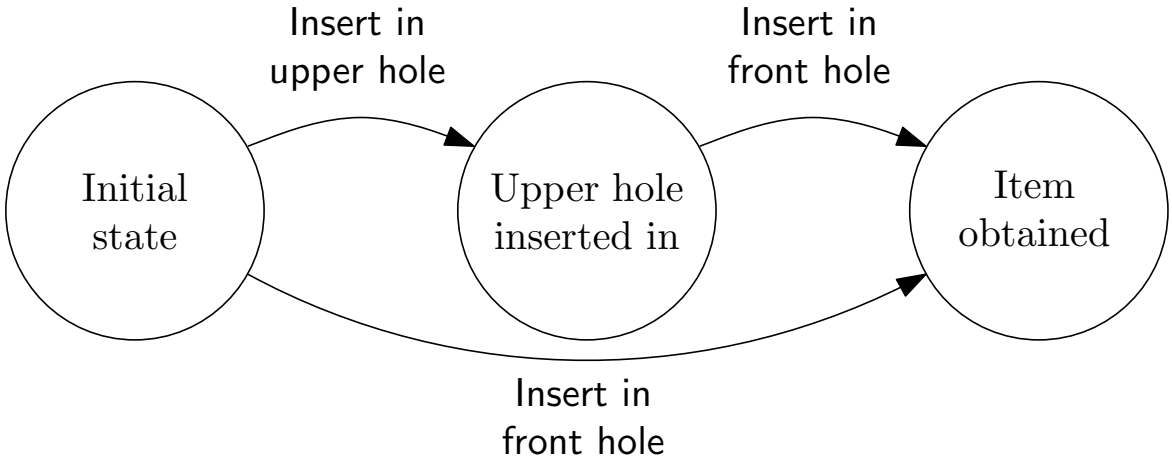


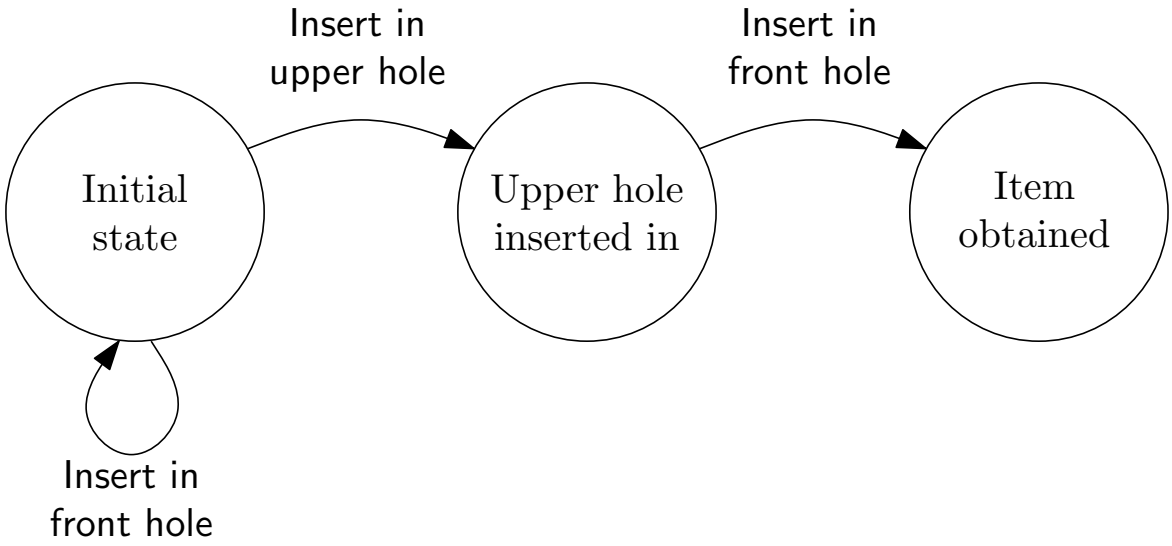
Figure 6: Percentage of runs in which the modelled agent replicates the demonstrated use of head as λ_I (the weight of the intention replicating module) is increased. Whenever the action was not performed with the head, it was performed with the hand. The weight of the baseline preference module, λ_B, is kept constant with a value of 0.2 (so λ_E, the weight of the end-effect replicating module, decreases to 0).

The archetypical such experiment includes the demonstration of a sequence of actions, not all of which are actually necessary to achieve the outcome. Horner and Whiten (2005) presented preschoolers and chimpanzees with two identical boxes, one opaque and one transparent. The demonstration consisted of inserting a stick into a hole on the top of a box and then into a hole on the front of the box, with the latter step causing the retrieval of a reward. The insertion of the stick into the top hole was unnecessary in order to obtain the reward, but the causal physical relations were only visible with the transparent box. The results showed that 3 and 4-year-old children tended to imitate both actions no matter whether they had observed and were tested on the transparent or opaque box. On the contrary, chimpanzees were able to switch between emulation and imitation if causal information was available; after having observed demonstrations with a transparent box, the chimpanzees had a greatly reduced tendency to insert the stick into the upper ineffective hole.

We simulated this experiment, using the model depicted in Figure 7. In the experiment in Subsection 3.2 the learning agent considered that only the demonstrator had a restriction, *i.e.,* it could not use the hand, but that the learning agent could use the hand. In this case the learner does not know the real dynamics of the world and



(a) Transparent box.



(b) Opaque box.

Figure 7: The world model for each of the two conditions in Horner and Whiten (2005). Notice that the difference in the world models represents the different *knowledge* of the learner about the world in the two conditions, rather than differences in the causal system. We again omitted the "No action" possibility from the diagram.

so has to rely on the demonstration to infer them. In the "transparent" condition the learner knows that it is possible to directly open the front lock and get the reward. In the opaque condition the learner does not know that this is possible.

In each simulation (in both conditions), the observer is exposed to one demonstration of the action "Insert in upper hole" followed by the action "Insert in front hole", and is then allowed to act. The baseline preference module makes no distinction between the two actions, *i.e.,* both actions are equally preferable. The observer can also choose to do nothing and has a baseline preference for the latter option over the former two.

The simulation results greatly depend upon the particular condition considered (see Fig. 8). In the opaque condition the learner is faced with a lack of world knowledge and so, both the intention and end-effect replicating

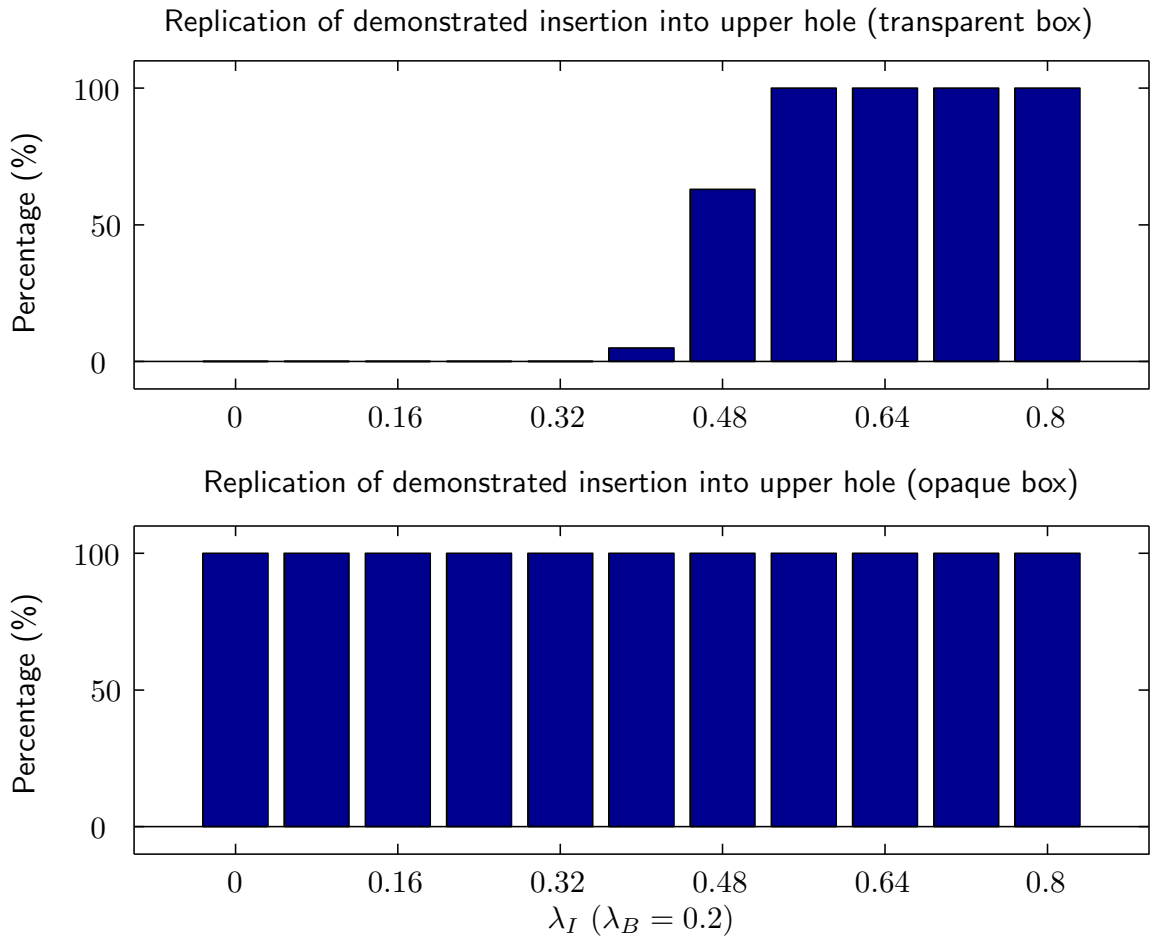


Figure 8: Percentage of runs in which the modelled agent replicates the demonstrated insertion into the upper hole, as λ_I (the weight of the intention replicating module) is increased. Whenever the upper hole was not inserted into, the front hole only is inserted into. The weight of the baseline preference module, λ_B, is kept constant with a value of 0.2 (so λ_E, the weight of the end-effect replicating module, decreases to 0).

modules can only choose to open both locks to obtain the item. In the transparent condition the end-effect replicating module chooses the most efficient method, while the intention replicating module infers that the more complex alternative was intended, because it was chosen over a simpler alternative, and so copies both actions. With the transparent box, the tendency to insert the stick in the upper hole, which has no visible effect, increases with λ_I. This shows that, as expected, unless emphasis is placed upon the imitation of inferred intentions rather than the tendency to simply obtain the reward, the model tends to emulate with the transparent box. With the opaque box, it is not clear what the effect of inserting in the upper hole is, and it is therefore not possible to know that the reward may be obtained without first inserting in the upper hole as is demonstrated. The agent therefore always inserts in both holes, independently of the value of λ_I.

Our simulation results suitably replicate the results from both children and chimpanzees, with a higher value of λ_I for children. Horner and Whiten (2005) suggest that the difference occurs because chimpanzees are primarily motivated to select the most efficient method they know to achieve the end effect, whereas children are more motivated to copy the inferred intentions of the demonstration (see also Tomasello et al. (2005)). Imitation in

cases such as the transparent box has been termed *over-imitation* because actions are imitated despite the fact that they serve no visible purpose (Lyons et al., 2005; Horner & Whiten, 2005).

Our model confirms the logic of the interpretation of the phenomenon of over-imitation in terms of the inferring and sharing of intended goals, without necessarily understanding the higher level prior intention. Note, however, that our model does not include an explanation for why children should be motivated to imitate the actions that do not appear to have an effect – the λ_I parameter is simply set high to enable this motivation.

The model does demonstrate that a complex motivation is not necessary to explain the results of the experiment modelled here – even a simple automatic tendency to imitate (Dijksterhuis & Bargh, 2001) would suffice. However, it is also possible, for example, that children make inferences about the opaque causal structure of actions with no visible consequence – in other words, that individuals imitate actions because they have encoded them as causing useful outcomes, even though they don't know how (see Section 4 and Lyons et al., 2005).

3.4 Intermediate behaviors

We now present again the simulation of Subsection 3.2 but now we evaluate the outcome of increasing λ_I (the tendency to replicate the inferred intention) *while* λ_E *is set to* 0. This corresponds to completely ignoring the behaviour coming from the end-effect replicating module and slowly "shifting" the interest of the agent from its baseline preferences toward the replication/imitation of the observed demonstration. In this new situation it is important to recall that we always allow the agent the possibility of performing no action. In terms of baseline preferences, we consider that the agent prefers to do nothing over using the hand and prefers to use the hand over using the head.

The obtained results are depicted in Fig. 9. The result shows that the agent starts by performing no action, then *replicates the observed effect*, choosing the most effective action, and only for higher values of λ_I does the replication of the observed action appear. In previous simulations, the agent never chose to perform no action, because λ_B (the weight of the baseline preference module) was very small. The existence of an "intermediate" behaviour (more obvious in the restricted condition) in the absence of the end-effect replicating module, is a prediction of the model that could be very interesting to observe in animals.

Our interpretation of this behaviour is the following. For $\lambda_I = 0$ the agent is focused on "energy conservation", opting to do nothing. Increasing the interest for replicating the observed demonstration leads the agent to compromise, replicating only "part" of the demonstration (touching the panel/turning the light on) while maintaining the energy consumption to a minimum (using the hand). This corresponds to the intermediate emulative behaviour. By further increasing the importance of replicating the observed demonstration while reducing the energy concerns, the agent finally adopts the imitative behaviour, as observed in our results.

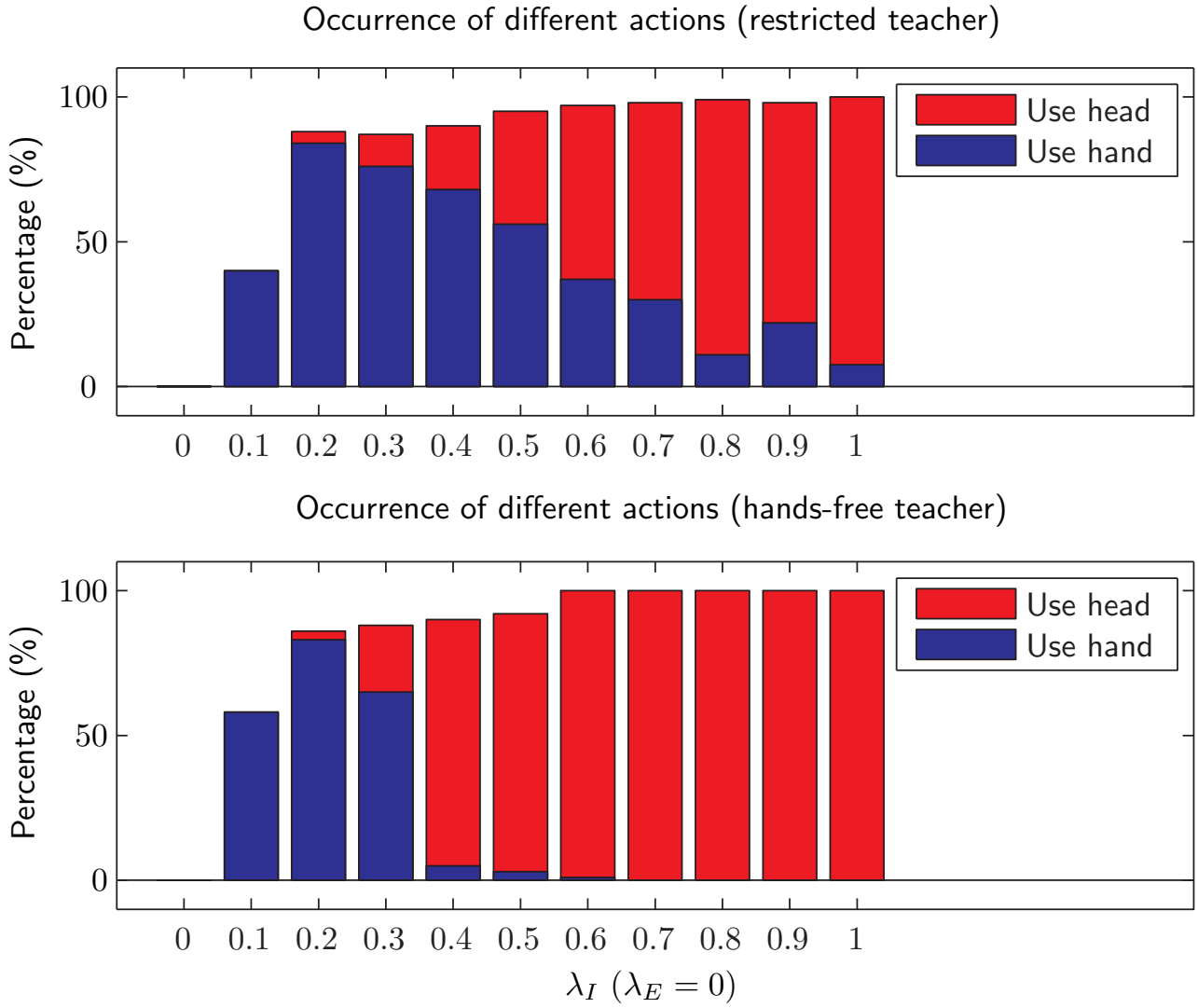


Figure 9: Rates of occurrence of the different actions as λ_I (the weight of the intention replicating module) is increased. When none of the two indicated actions is performed, no action is performed. The weight of the end-effect replicating module, λ_E, is kept constant at 0, hence λ_B (the weight of the baseline preference module) starts at 0.8 and decreases to zero. Note that the agent starts by performing no action, then emulates (although this emulation does not arise from the end-effect replicating module) and then finally imitates.

4 GENERAL DISCUSSION

The motivation for our study stemmed from the fact that, while many experiments have been conducted investigating the conditions under which children and apes use different strategies for incorporating observed behaviour into their own repertoire, there is still no definitive theory about the mechanisms which enable switching between strategies.

We started from the taxonomy proposed by Call and Carpenter (2002) to build a unifying mathematical model of types of social influence on behaviour, mainly imitation and emulation. Notwithstanding, we believe that the separation of socially acquired behaviours into different categories might not necessarily correspond to independently operating cognitive mechanisms, but to different ways of integrating the mechanisms.

It is worth noting that attention plays an important part in determining the goal of an action or understanding

the relevant part of a demonstration, something that is not explicit in Call and Carpenter's model. Although we did not explicitly model such a mechanism, we implicitly included attentional information when designing the world models we used. This effect can be seen in the work of McGuigan et al. (2007), in which the experiment described in Section 3.3 was replicated, additionally including a condition in which the demonstration was presented on a video screen with only the demonstrator's hands and the apparatus visible. They found that this degradation of the demonstration's social context caused 3-year-olds to adopt an emulative rather than imitative approach.

Important evidence of how young children represent and imitate others' actions in terms of intended goals comes from their ability to socially learn complete actions which they have only seen partially demonstrated, due to mistakes or inability of the demonstrator (Meltzoff, 1995; Johnson et al., 2001). In another study of our model, we demonstrated that the learning agent is also capable of handling such accidental or incomplete actions, by correctly interpreting the task even when there are errors in the demonstration. The inference module is robust to mistakes in the demonstrated action sequence if these are, in a sense, incompatible with the general goal that can be inferred from the demonstration (Lopes et al., 2007).

The sources of information that shape the behaviour of our model are three-fold: (i) beliefs about the world's possible states and actions transitioning between them, and baseline preferences among these actions; (ii) a variable tendency to infer and adopt intentions of observed behaviour; and (iii) a variable tendency to attempt to achieve observed results. Acting on these premises, our model was able to replicate the results from three archetypical empirical experiments from important methodological paradigms in infant, preschooler, and chimpanzee social learning (equivalent situations are presented in the works of McGuigan et al., 2007; Lyons et al., 2005; Brugger et al., 2007; Schwier et al., 2006). We have thus demonstrated that a rather parsimonious artificial system, using a single computational formalism and only two variable parameters, can account for a variety of phenomena observed in empirical social learning experiments, such as goal-inference taking into account action constraints and incomplete knowledge, over-imitation and flexible constraint-sensitive imitation.

On the basis of the obtained results and established facts from social psychology, we now discuss the interpretation of our model together with possible reasons for some of the observed behaviours, and make several testable predictions.

A switch between imitation and emulation might be triggered by changing the value (to the learner) of the social interaction or of the effect. Our model produces different behaviours with different weights on the different modules, representing the influence of the importance of each element to different experimental participants in different circumstances, a subject widely studied in behavioural psychology. These mathematical values correspond to psychological characteristics such as: urgency, motivation and desire. Lyons et al. (2005) found that increasing the urgency to solve a task reduced the tendency to over-imitate in 3- to 5-year-olds.

The greater utilization of imitation by children might be explained by a stronger focus on others' intentions, mediated by social cues. Social cues have been observed to be important in promoting imitation. Infants were observed to imitate intended results, even when the demonstrator makes a mistake and fails to obtain the result

(Meltzoff, 1995). Brugger et al. (2007) also found that 15-month-olds were not very prone to over-imitate but that social cues could increase the tendency. The arguments of Tomasello et al. (2005) support our contention that the differences between children and chimps in the experiment simulated in Section 3.3 can be accounted for by simply varying the parameter controlling weighting between intentions and end effects.

It is also interesting to note that similarly to the way in which children's motivation to imitate can be manipulated, chimpanzees may also show different tendencies to imitate, depending on background factors. For example, Tomasello et al. (1993) argue that enculturated chimpanzees are better imitators than wild chimpanzees. This may be because exposure to a complex human environment equips them with different motivations (or abilities to process to different types of information, see in the workd of Lyons et al., 2006). It can also be speculated that even in humans, different backgrounds in terms of exposure to complex action sequences might similarly affect tendency to imitate, via effects on motivation or ability.

The development of this unifying model allows us to reason not only about possible interpretations but also to predict the behaviour in novel or more complex situations. For instance:

"Pure imitation" vs. "pure emulation" behaviours will become more indistinguishable as the complexity of the task increases. If the mechanism of social learning is, as we suggested, a combination of several strategies, then the resulting behaviour will be different from that which would be produced by any of the strategies operating individually. In the experiments considered herein, where the agent has to perform only one or two actions, this effect is not visible. However, we expect this aspect to become visible if there is a longer sequence of optional actions.

One good example is that proposed by Williamson and Markman (2006) who present one of the few experiments with a sequence of actions, where the action pattern observed did indeed correspond to a mixed imitative behaviour (see also the work of Flynn & Whiten, 2008). In experiments with a robot we also observed such a phenomenon, where the resulting behaviour was neither pure emulation nor pure imitation (Lopes et al., 2007).

A continuous change in the value of the social interaction (the tendency to imitate) vs. the tendency to conserve energy may lead to several intermediate behaviours. This phenomenon was observed in the simulation shown in Section 3.4. It reinforces our previously made arguments that more complex situations, involving more alternative action possibilities, will result in more complex arrays of behaviour. We predict that such a phenomenon might be observable experimentally, for example in an imitation game with children in which the motivation to imitate the goals of the demonstrator is manipulated. A task which is in itself inherently rather boring might allow effective manipulation of motivation to imitate, by varying how engaging the demonstrator is. Our model predicts that in such a situation, as engagement is increased, first behaviours will appear that appear emulative, and then behaviours will appear which appear more faithfully imitative.

These behaviours which appear emulative may occur (at intermediate levels of motivation to imitate) even when there is no motivation to emulate (for example if the end-effect is inherently unrewarding). This is because, as observed in our model, a partial imitation may appear emulative although it is not in fact motivated by the

achievement of and end-effect for its own sake.

$\diamondsuit$

We argue that all animals that are able to imitate and emulate (such as children and chimpanzees) need to have, at least, the mechanisms considered in our model. Given that young children and chimpanzees are both known to be able to imitate and emulate (Whiten et al., 2004; Want & Harris, 2002; Tennie et al., 2006) depending upon circumstances, we suggest that our computational framework can be used as an adequate model for both these species, with a generally higher value of λ_I for children than chimpanzees. This is to say that, when faced with prioritising either faithful imitation or achieving the results as fast as possible, different species weight differently the different motivations and sources of information.

The components of our model thus seem sufficient to explain much of what is known about tendencies to imitate or emulate in children and chimpanzees. We are unable to conceive of a simpler model to replicate these results and as such we believe that our computational model provides a parsimonious explanation for the observed behaviours. And, although in some situations similar behaviours could be obtained with simpler mechanisms such as mimicry, stimulus enhancement, response facilitation and contextual facilitation (Melo et al., 2007; Byrne, 2002; Noble & Franks, 2002), such mechanisms cannot account for all the phenomena reviewed in this work.

A Technical details

Now we proceed with the details about the underlying model[2]. At each time instant, the learner must choose an action from its repertoire of action primitives $\mathcal{A}$, depending on the state of the environment. We represent the state of the environment at time t by X_t and let $\mathcal{X}$ be the (finite) set of possible environment states. This state evolves according to the transition probabilities

$$\mathbb{P}\left[X_{t+1} = y \mid X_t = x, A_t = a\right] = \mathsf{P}_a(x, y), \tag{1}$$

where A_t denotes the learner's action primitive at time t. The action-dependent transition matrix P thus describes the dynamic behaviour of the process $\{X_t\}$.

We consider that the demonstration consists of a sequence $\mathcal{H}$ of state-action pairs

$$\mathcal{H} = \{(x_1, a_1), (x_2, a_2), \ldots, (x_n, a_n)\}.$$

Each pair (x_i, a_i) exemplifies to the learner the expected action (a_i) in each of the states visited during the demonstration (x_i). From this demonstration, the learning agent is expected to perceive what the demonstrated task is and, eventually by experimentation, learn how to perform it optimally. A decision-rule determining the action of the learner in each state is called *a policy* and is denoted as a map $\pi : \mathcal{X} \longrightarrow \mathcal{A}$.

[2]An extended version can be found at `http://users.isr.ist.utl.pt/~macl/myrefs/SL08app.pdf`

In our adopted formalism, a task can be defined using a function $u : \mathcal{X} \longrightarrow \mathbb{R}$ describing the "immediate desirability" of each particular state $x \in \mathcal{X}$ in terms of the task. Once u is known, the learner should choose its actions to maximize the functional

$$J(x, \{A_t\}) = \mathbb{E}\left[\sum_{t=1}^{\infty} \gamma^t u(X_t) \mid X_0 = x\right],$$

where γ is a discount factor between 0 and 1 that assigns greater importance to the immediate future than to the distant future.[3]

The relation between the function u describing the task and the optimal behavior rule can be evidenced by means of the function V_u given by

$$V_u(x) = \max_{a \in \mathcal{A}} \left[u(x) + \gamma \sum_{y \in \mathcal{X}} \mathsf{P}_a(x, y) V_u(y)\right]$$

The value $V_u(x)$ represents the expected (discounted) utility of a path of the process $\{X_t\}$ starting at state x, when the optimal behavior rule is followed. Letting

$$Q_u(x, a) = u(x) + \gamma \sum_{y \in \mathcal{X}} \mathsf{P}_a(x, y) V_u(y), \tag{2}$$

it holds that

$$V_u(x) = \max_{a \in \mathcal{A}} Q_u(x, a)$$

and the optimal policy associated with the function u is given by

$$\pi_u(x) = \arg\max_{a \in \mathcal{A}} Q_u(x, a).$$

The computation of π_u (or, equivalently, Q_u) given P and u is a standard problem and can be solved using any of several standard methods available in the literature (Bertsekas & Tsitsiklis, 1996).

$\diamondsuit$

Within the formalism just described, the fundamental imitation problem lies in the estimation of the function u from the observed demonstration $\mathcal{H}$. In the continuation, we discuss how this function u is computed by each of the modules in our model.

A.1 The proposed computational model

Our model takes into account the agent's baseline preferences, the effects of the demonstrated actions and the possible goals of the demonstrator. Each of these sources of information is processed in a specific "module", that

[3] The discount factor γ can be seen by the agent as a "probability of surviving" in the next time-step.

generates a representation of the corresponding behaviour. These behaviours are then combined by merging the corresponding representations using a standard convex combination.

As seen above, the function Q_u associated with a particular task can be used to compute the optimal policy π_u for that task. More generally, such a "Q-function" can be used to define a general policy and we will adopt this approach to represent the behaviours computed in each of the modules in our model.

The agent's baseline preferences: For each scenario, this component of the model simply outputs a previously defined function Q_B. This function encompasses the baseline preferences of the agent in that, if action a_1 is preferred over action a_2 in a particular state of the world x, then

$$Q_B(x, a_1) > Q_B(x, a_2).$$

This function can be seen as "part" of the definition of the agent: its values are set beforehand, independently of the demonstration.

Replicating the end-effect: Throughout the simulations in the paper, we considered the desired effect as the final state observed during the demonstration, hereby denoted as x_E. Replicating the effect thus consists in attaining x_E. The task of attaining x_E can be represented by means of a utility function u_E defined as

$$u_E(x) = \begin{cases} 1 & \text{if } x = x_E; \\ 0 & \text{otherwise.} \end{cases}$$

The function Q_E obtained from this utility represents a behaviour for reaching x_E as quickly as possible and can be easily computed using standard dynamic programming.

Inferring the goal of the demonstrator: We adopt the method by Melo et al. (2007), which is a basic variation of the *Bayesian inverse reinforcement learning* (BIRL) algorithm (Ramachandran & Amir, 2007).

For a given u-function, the *likelihood of a pair* (x, a) is defined as

$$L_u(x, a) = \mathbb{P}\left[(x, a) \mid u\right] = \frac{e^{\eta Q_u(x,a)}}{\sum_{b \in \mathcal{A}} e^{\eta Q_u(x,b)}}.$$

The parameter η is a user-defined *confidence parameter* that we describe further ahead. The value $L_u(x, a)$ translates the "plausibility" of the choice of action a in state x when the underlying task is described by u. Given a demonstration sequence

$$\mathcal{H} = \{(x_1, a_1), (x_2, a_2), \ldots, (x_n, a_n)\}.$$

the corresponding likelihood is

$$L_u(\mathcal{H}) = \prod_{i=1}^{n} L_u(x_i, a_i).$$

The method uses MCMC to estimate the distribution over the space of possible u-functions, given the demonstration (Ramachandran & Amir, 2007). It will then choose the maximum *a posteriori* u-function. Since we consider a uniform prior for the distribution, the selected utility is the one whose corresponding optimal policy "best matches" the demonstration. The confidence parameter η determines the "trustworthiness" of the method: it is a user-defined parameter that indicates how "close" the demonstrated policy is to the optimal policy (Ramachandran & Amir, 2007). Once the "best" u-function is chosen, standard dynamic programming is used to compute the corresponding Q-function, Q_I.

$\diamondsuit$

We conclude by discussing how the underlying structure in our formalism translates to biological terms. First of all, the assumed "world knowledge" consists of the set of possible states of the environment, $\mathcal{X}$, the repertoire of action primitives, $\mathcal{A}$, and the world dynamics, summarized by the transition probabilities P. Note, in particular, that the action repertoire $\mathcal{A}$ is fixed and known in advance. This means that our overall model addresses learning at the *task* level. The modelled agent does not learn new actions, but instead learns how to combine known actions in new ways. Formally, there is no reason why our model cannot be used at different levels of abstraction, but the biological correspondence may become less clear.

Secondly, we note that the goal-inference model is probabilistic and relies on a Bayesian formalism that can be exploited beyond what was described here. Its probabilistic nature implies that the goal-inference module is somewhat robust to some wrong ("accidental") actions if these are, in a sense, incompatible with the general goal that can be inferred from the demonstration. We refer to the work of Lopes et al. (2007) for further discussion on the robustness of the method to partially incorrect actions. On the other hand, the Bayesian formalism allows the inclusion of prior information in a straightforward manner. In other words, the Bayesian formalism easily accommodates prior information on possible utilities which, in our particular setting, would translate into prior information on the demonstrator's prior intentions.

Acknowledgments

This work was supported in part by the the EU projects RobotCub (IST-004370) and Contact (EU-FP6-NEST-5010) and also by FCT Programa Operacional Sociedade de Informação (POSC) in the frame of QCA III, the Carnegie Mellon-Portugal Program and the project PTDC /EEA-ACR/70174/2006.

References

Bekkering, H., Wohlschläger, A., & Gattis, M. (2000). Imitation of gestures in children is goal-directed. *Quarterly J. Experimental Psychology*, *53A*, 153-164.

Bertsekas, D. P., & Tsitsiklis, J. N. (1996). *Neuro-dynamic programming.* USA: Athena Scientific.

Brass, M., Schmitt, R. M., Spengler, S., & Gergely, G. (2007). Investigating action understanding: Inferential processes versus action simulation. *Current Biology*, *17*(24), 2117-2121.

Brugger, A., Lariviere, L. A., Mumme, D. L., & Bushnell, E. W. (2007). Doing the right thing: Infants' selection of actions to imitate from observed event sequences. *Child Development*, *78*(3), 806–824.

Byrne, R. W. (2002). Imitation of novel complex actions: What does the evidence from animals mean? *Advances in the Study of Behavior*, *31*, 77-105.

Call, J., & Carpenter, M. (2002). Three sources of information in social learning. In *Imitation in animals and artifacts.* Cambridge, MA, USA: MIT Press.

Carpenter, M., Call, J., & Tomasello, M. (2002). Understanding "prior intentions" enables two-year-olds to imitatively learn a complex task. *Child Development*, *73*(5), 1431–1441.

Carpenter, M., Call, J., & Tomasello, M. (2005). Twelve- and 18-month-olds copy actions in terms of goals. *Developmental Science*, *1*(8), F13–F20.

Csibra, G., & Gergely, G. (2007). "Obsessed with goals": Functions and mechanisms of teleological interpretation of actions in humans. *Acta Psychologica*, *124*, 60–78.

Dijksterhuis, A., & Bargh, J. A. (2001). The perception-behavior expressway: Automatic effects of social perception on social behavior. In *Advances in experimental social psychology* (Vol. 33, pp. 1–40). San Diego, USA: Academic Press.

Flynn, E., & Whiten, A. (2008). Imitation of hierarchical structure versus component details of complex actions by 3-and 5-year-olds. *Journal of Experimental Child Psychology*, *101*(4), 228–240.

Gergely, G., Bekkering, H., & Király, I. (2002). Rational imitation in preverbal infants. *Nature*, *415*, 755.

Horner, V., & Whiten, A. (2005). Causal knowledge and imitation/emulation switching in chimpanzees (Pan troglodytes) and children (Homo sapiens). *Animal Cognition*, *8*, 164–181.

Johnson, S., Booth, A., & O'Hearn, K. (2001). Inferring the goals of a nonhuman agent. *Cognitive Development*, *16*(1), 637-656.

Kenward, B., Folke, S., Holmberg, J., Johansson, A., & Gredebäck, G. (2009). Goal-directedness and decision making in infants. *Developmental Psychology.*

Klossek, U. M. H., Russell, J., & Dickinson, A. (2008). The control of instrumental action following outcome devaluation in young children aged between 1 and 4 years. *Journal of Experimental Psychology-General*, *137*(1), 39–51.

Lopes, M., Melo, F. S., & Montesano, L. (2007, Nov). Affordance-based imitation learning in robots. In *Ieee/rsj international conference on intelligent robots and systems* (p. 1015-1021). USA.

Lyons, D. E., Santos, L. R., & Keil, F. C. (2006). Reflections of other minds: how primate social cognition can inform the function of mirror neurons. *Current Opinion in Neurobiology*, *16*(2), 230–234.

Lyons, D. E., Young, A. G., & Keil, F. C. (2005). The hidden structure of overimitation. *Proceedings of the National Academy of Sciences*, *104*(50), 19751–19756.

McGuigan, N., Whiten, A., Flynn, E., & Horner, V. (2007). Imitation of causally opaque versus causally

transparent tool use by 3- and 5-year-old children. *Cognitive Development*, *22*, 353–364.

Melo, F., Lopes, M., Santos-Victor, J., & Ribeiro, M. I. (2007, April). A unified framework for imitation-like behaviors. In *4th international symposium on imitation in animals and artifacts.* Newcastle, UK.

Meltzoff, A. N. (1988). Infant imitation after a 1-week delay: Long-term memory for novel acts and multiple stimuli. *Developmental Psychology*, *24*(4), 470–476.

Meltzoff, A. N. (1995). Understanding the intentions of others: Re-enactment of intended acts by 18-month-old children. *Developmental Psychology*, *31*(5), 838–850.

Nielsen, M. (2006). Copying actions and copying outcomes: Social learning through the second year. *Developmental Psychology*, *42*(3), 555–565.

Noble, J., & Franks, D. W. (2002). Social learning mechanisms compared in a simple environment. In *Artificial life viii: Proceedings of the eighth international conference on the simulation and synthesis of living systems* (pp. 379–385). Cambridge, MA, USA: MIT Press.

Ramachandran, D., & Amir, E. (2007). Bayesian inverse reinforcement learning. In *20th int. joint conf. artificial intelligence.* India.

Range, F., Viranyi, Z., & Huber, L. (2007). Selective imitation in domestic dogs. *Current Biology*, *17*(10), 868–872.

Schwier, C., Maanen, C. van, Carpenter, M., & Tomasello, M. (2006). Rational imitation in 12-month-old infants. *Infancy*, *10*(3), 303–311.

Searle, J. R. (1983). *Intentionality: An essay in the philosophy of mind.* Cambridge, UK: Cambridge University Press.

Tennie, C., Call, J., & Tomasello, M. (2006). Push or pull: Imitation vs. emulation in great apes and human children. *Ethology*, *112*(12), 1159–1169.

Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and Brain Sciences*, *28*(5), 675–691.

Tomasello, M., Kruger, A. C., & Ratner, H. H. (1993). Cultural learning. *Behavioral and Brain Sciences*, *16*(3), 495–511.

Want, S. C., & Harris, P. L. (2002). How do children ape? Applying concepts from the study of non-human primates to the development study of "imitation" in children. *Developmental Science*, *5*(1), 1–13.

Whiten, A., Custance, D., Gomez, J.-C., Teixidor, P., & Bard, K. A. (1996). Imitative learning of artificial fruit processing in children (Homo sapiens) and chimpanzees (Pan troglodytes). *Journal of Comparative Psychology*, *110*, 3-14.

Whiten, A., Horner, V., Litchfield, C. A., & Marshall-Pescini, S. (2004). How do apes ape? *Learning & Behavior*, *32*(1), 36–52.

Williamson, R. A., & Markman, E. M. (2006). Precision of imitation as a function of preschoolers' understanding of the goal of the demonstration. *Developmental Psychology*, *42*(4), 723–731.