Rational imitation for robots:The Cost Difference Model

Dieter Vanderelst, Alan Winfield
Bristol Robotics Lab

March 7, 2017

Abstract

Infants imitate behavior flexibly. Depending on the circumstances, they copy both actions and their effects or only reproduce the demonstrator's intended goals. In view of this selective imitation, infants have been called rational imitators. The ability to selectively and adaptively imitate behavior would be a beneficial capacity for robots. Indeed, selecting what to imitate is one of the outstanding unsolved problems in the field of robotic imitation. In this paper, we first present a formalized model of rational imitation suited for robotic applications. Next, we test and demonstrate it using two humanoid robots.

1 Introduction

Imitation is a very important form of social learning in humans and has been suggested to underlie human cumulative culture (Legare and Nielsen, 2015; Tomasello, 2009). In keeping with its importance in human development, the ability to imitate emerges early in human infants. From their second year on, infants can imitate actions and their intended goals from demonstrators (e.g., Gariépy et al., 2014; Jones, 2009). Critically, infants imitate the demonstrated actions and their effects in a flexible way. Depending on the circumstances, they copy both actions and effects or only reproduce intended goals. In view of this selective imitation, infants have been called rational imitators (Gergely et al., 2002).

In a landmark paper, Meltzoff (1988) showed that 14-month-old children switch on a light by bending over and touching it with their head, if they have seen an experimenter do so. Later studies showed that if the experimenter's hands are occupied children tend to switch on the light using their hands (Gergely et al., 2002). The percentage of copied head-touch actions also declines when the demonstrator's hands are physically restrained (Zmyj et al., 2009; Gellén and Buttelmann, 2017). Apparently, when the experimenter's hand are occupied or restrained, the children deem the head touch to be irrelevant to the outcome. These results have been replicated by Beisert et al. (2012) and Paulus et al. (2011), albeit with a different interpretation.

Another aspect of rational imitation was demonstrated in a study by Carpenter et al. (2005). A demonstrator moved a toy mouse to a target position either using a sliding or hopping motion. If a toy house was present at the goal location, children were less likely to copy the motion than if no house was present. The authors assumed that the presence of the house induced the children to adopt the goal of placing the mouse in the house whilst disregarding the demonstrated motion. In the absence of the toy house, the children presumably perceived motions as being the goal, and therefore, as relevant.

In summary, young children (act as if they) are able to distinguish between relevant and irrelevant aspects of demonstrated behaviour. They seem to copy the actions more often if relevant for attaining the goal. In particular, they seem to (1) take into account the constraints of the demonstrator and (2) discount actions in favour of goals.

Since the advent of robotics, imitation has been suggested as a method for learning in robots. Billard et al. (2008) list two advantages of imitation learning. First, learning from a demonstrator greatly simplifies the search solutions to sensorimotor problems, which are typically hard. In addition, imitating robots would be programmable by lay-persons using the same methods they employ to teach other people. Robotic imitation faces a number of challenges (Dautenhahn and Nehaniv, 2002). One of the most fundamental issues is determining what to imitate (Carpenter and Call, 2006; Breazeal and Scassellati, 2002). Among other aspects, this involves determining the relevant parts of a demonstrated action and only copying those. Hence, the selective and rational imitation shown by children would be a beneficial capacity for robots (Gergely, 2003). Unfortunately, in spite of the considerable body of experimental data, the cognitive mechanisms underlying rational imitation remain elusive. In particular, no satisfactory and computationally explicit model of rational imitation in infants is available.

Initially, authors explained the results of experiments by assuming that infants reason teleologically about the goals and actions demonstrated (See Zmyj and Buttelmann, 2014, for references). Children are assumed to infer that (1) the demonstrator uses his or her head to switch on the lamp because his or her hands are constrained and (2), as such, the head touch is not necessary to successfully switch on the lamp. Therefore, when asked to switch on the lamp, the infant uses his or her hands. In contrast, when the demonstrator's hands are free, the infants are assumed to reason that the head touch is instrumental in obtaining the goal.

More recently, competing accounts have been advanced (See also Gellén and Buttelmann, 2017, for an overview). In particular, it has been proposed that many experimental results can be explained by differences in the difficulty for the infants to copy the demonstrator's actions (Zmyj and Buttelmann, 2014). According to this account, bending forward to touch a lamp with restrained hands is more difficult than doing so with free hands available to support the body. As such, an increased difficulty in exactly copying the demonstrated motion – termed a lack of 'motor resonance' (Paulus et al., 2011) – is assumed to reduce the extent to which infants copy a demonstrated action. Beisert et al. (2012) advanced yet another account of rational imitation in infants. These authors have claimed that attentional processes can fully explain selective imitation.

While it is undoubtedly (and unsurprisingly) true that both the feasibility of the demonstrated actions and attentional processes determine the fidelity of action copying, neither account fully accommodates the experimental findings (Zmyj and Buttelmann, 2014). For example, even in the absence of obvious differences in action difficulty, 12-month old infants copy a model with constrained hands less often (Zmyj et al., 2009). In addition, 12-month old – but not 9-month old – infants ignored the head touch action of a model with hands fixed to the table (Zmyj et al., 2009). It is difficult to see how infants would be susceptible to 'a lack of motor resonance' at 12 months but not at 9 months. Likewise, attentional mechanisms cannot explain effects across conditions that do not seem to recruit different levels of attention (Paulus et al., 2013; Kolling et al., 2014).

While the motor resonance and attention theories fall short in accommodating for some data, the reasoning hypothesis suffers mainly from being underspecified – although it can be noted that the idea of 'motor resonance' is less than fully specified either (Zmyj and Buttelmann, 2014). As a result, the reasoning account can be made to accommodate most findings *post facto*. For example, Paulus et al. (2011) conducted an experiment to distinguish between the reasoning account and the motor resonance model. They concluded that findings were more in line with the predictions of the motor resonance model. However, it is unclear whether the predictions these authors derive for the teleological reasoning account are the only interpretation possible (See Zmyj and Buttelmann (2014) for a similar remark).

In the absence of a complete and computationally explicit model, we propose a novel model for rational imitation, i.c. the Cost Difference Model (CDM). In particular, we aim for a model that supports rational imitation in robots. In contrast to the accounts discussed above – and in accord with our goal to exploit rational imitation to optimize the imitation behaviour in robots – we depart from a normative analysis of imitation learning. That is, we postulate the desirable properties of rational imitation and build a model satisfying these requirements.

2 The Cost Difference Model

2.1 Rationale

In agreement with current views on its adaptive value (e.g., Laland, 2004; Erbas et al., 2013), we propose that imitation is a method for acquiring better action policies (Argall et al., 2009). Action policies can be thought of as a series of subgoals that lead towards attaining the final goal. For example, an action policy for making spaghetti (final goal) are the steps (subgoals) as set out in the recipe.

Assuming that imitation is a learning strategy for adopting better action policies for satisfying goals, imitation has the possible advantage of being a cheaper (less risky) route to policy learning than individual, asocial learning. Nevertheless, indiscriminately copying behaviour is unlikely to result in better policies (Laland, 2004). Ideally, agents should only copy behaviour when an observed policy is better than the current existing action policy. Initially, we can assume better policies to be those requiring less energy. However, other optimization criteria could be imagined, including risk and time. In biological agents, better action policies are those ultimately resulting in increased fitness.

In this light, experimental findings on imitation in infants are somewhat puzzling. Infants copy demonstrated head touches in spite of clearly being able to switch on the light using their hands (which seems to be a better policy). Indeed, in control conditions, children spontaneously switch on the light using their hands. Moreover, even when infants eventually copy the head touch, most often they switch on the light using their hands first (Paulus et al., 2013, 2011; Gergely, 2003). So why do children copy the ineffective head touch policy given they have an alternative policy that seems more efficient?

In our view, this discrepancy can be explained by assuming that an agent observing a demonstrated action policy has only limited knowledge about its energetic cost. The agent might be able to estimate the energy requirement of the demonstrated policy, for example, using its own action planner (or internal simulation, Hesslow (2002, 2012)). However, this will yield an approximate estimate at best – especially when the demonstrated policy includes unfamiliar actions. In addition, the agent can estimate or retrieve the cost of its existing action policy and compare this to the estimated value of the demonstrated action policy. In agreement with this assumption, infants expect demonstrators to minimize the costs of actions (Liu and Spelke, 2017, and references therein). Moreover, actions that violate this assumptions recruit more attention from the infants.

Theoretically, the agent should reject the demonstrated policy whenever its cost is higher than that of the existing policy. However, the cost of the demonstrated policy is not directly accessible and is only an estimate. As such, seeing someone executing a costly action policy might indicate that the estimated cost is inaccurate. If so, it would be reasonable to actually execute the demonstrated policy and obtain a corrected estimate of its cost. Indeed, the potential long-term gain of chancing on an innovative policy would generally outweigh the cost of testing out the action.

In summary, we propose that the rational imitation observed in infants is the overt outcome of uncertainty about the cost of the demonstrated action policy. This is, when copying an action policy they are exploring its cost by physically executing it. This overt action will result in a better estimate of its real cost. Critically, our hypothesis predicts that explorative copying of actions should occur more often if the demonstrator is deemed trustworthy (Laland, 2004; Vanderelst et al., 2009). This is corroborated in experiments. Infants more often copy ineffective behaviour from trusted (Zmyj et al., 2010; Poulin-Dubois et al., 2011) or familiar (Beisert et al., 2012) demonstrators. In addition, the notion of imitation as a method for exploring an action's cost is supported by the finding (mentioned above) that, even when infants eventually copy head touches, most often they switch on the light using their hands first. Hence, when copying the head touches, they actually perform both actions most of the time (Paulus et al., 2013, 2011; Gergely, 2003). This would allow them to directly compare the cost of both action policies. Moreover, our account predicts that children should have a tendency to over-imitate irrelevant actions as they result in an unexpected high cost estimate triggering explorative imitation of the demonstrated actions. This has been confirmed in a series of experiments (Lyons et al., 2007; Keupp et al., 2013). In agreement with our thesis, infants seem to assume that demonstrators will minimize the costs of their actions. When demonstrators fail to do so, this recruits increased levels of attention (Liu and Spelke, 2017) which could the mechanism that leads to increased imitation (or over-imitation).

Finally, it should be pointed out that our functional description of rational imitation suggests similar adaptive advantages are to be gained by other species. As such, it is interesting that both chimpanzees (Buttelmann et al., 2007) and dogs (Range et al., 2007) have found to be selective imitators in much the same way as human infants.

Having outlined a functional account of rational imitation, we proceed to describe the computations we assume to underlie the selection of action policies for imitation. We propose this proceeds in three steps: (1) parsing the continuous stream of sensory input, (2) solving the correspondence problem, (3) comparing the costs of the existing and the demonstrated action policies.

2.2 Formalization

2.2.1 Parsing behaviour

Behaviour consists of dynamic and continuous motions, and their effects. Hence, the first challenge for an imitating agent is parsing this stream of sensory input into meaningful chunks of actions and resulting effects. Indeed, young infants have been shown to parse behaviour into goal oriented chunks (e.g., Baldwin et al., 2001). In principle, they might use a wealth of task-related knowledge to solve this problem. However, they could also exploit low-level sensory cues signalling the boundaries between behavioural units, especially in early developmental stages (Baldwin et al., 2001). Indeed, adults will often explicitly capture the child's attention before initiating a demonstration. Likewise, they use verbal cues to signal the action has been completed. Verbal cues are commonly used in experimental investigations of imitation to denote the start and ending of a demonstration (e.g., Paulus et al., 2011; Schwier et al., 2006; Zmyj et al., 2009). In addition, more basic sensory cues could be salient changes in visual and auditory input or object motion.

In our experiments, we assume the robot can use either task-related knowledge or low-level sensory cues to parse the behaviour of a demonstrator and do not model this step explicitly.

2.2.2 Solving the correspondence problem

The second computational step concerns solving the correspondence problem. That is, the module converts the observed behaviour into the coordinate system of the observer. The correspondence problem is far from trivial (Nehaniv and Dautenhahn, 2001), in particular when the body plan of the demonstrator and observer are different. Indeed, errors made in solving the correspondence problem are assumed to be an important bottleneck preventing successful infant imitation (Gattis et al., 2002). However, in the field of robotics, a substantial amount of research has resulted in a number of methods for solving this problem (e.g., Argall et al., 2009; Schaal et al., 2003; Nehaniv, 2007). Hence, in this paper, we assume the problem can possibly be solved using the methods proposed earlier. The output of this computational step, a sequence of states in the observer's coordinate system, will be denoted by as $\vec{o}_t$ with t indexing the time, with $t = [0, T]$.

2.2.3 Inferring the demonstrator's policy

In order to model imitation based on the assumptions introduced above, we need to propose a mechanism that allows agents to infer the demonstrated action policy from the observed sequence of states $\vec{o}_t$. This is, the imitator needs to infer from $\vec{o}_t$ which intermediate goals the demonstrator satisfies en route to the final goal. To the best of our knowledge, no account of the method used by infants to select relevant subgoals from observed actions is available. Hence, in what follows, we present an approach that is suitable for the current robotic experiments. It should be understood that this method is a first approach and could be refined in further work to suit other contexts.

In more formal terms, inferring the demonstrator's action policy can be thought of as selecting the *minimal* number of intermediate states from $\vec{o}_t$ required to explain the observed behaviour $\vec{o}_t$. This set of minimal required states, denoted as $\vec{o}_s$, are assumed to be the subgoals of the demonstrator. Below, we explain our current approach to selecting this minimal set of states $\vec{o}_s$.

We suggest the robot should select an iteratively expanding set of states $\vec{o}_s = \{o_0 \dots o_n \dots o_T\}$ from the observed states $\vec{o}_t$. For each set $\vec{o}_s$, the robot uses its own action planner to compute an action sequence $\vec{a}_t$ leading from o_0 to o_T through the intermittent states o_n in $\vec{o}_s$. In planning the action sequence $\vec{a}_t$, the robot should take into account the physical constraints C experienced by the demonstrator. Hence, the action sequence $\vec{a}_t$ is the action plan the robot would come up with itself (1) *if it were in the same situation as the demonstrator* and (2) *wanted to attain each of the selected subgoals in* $\vec{o}_s$. As such, the notation for the planned action sequence, $\vec{a}_t$, should be considered as shorthand for $\vec{a}_t = f(\vec{o}_s, C)$ indicating that the planned action sequence is a function of (1) the currently selected action states $\vec{o}_s$ and (2) the physical constraints C. In terms of the behavioural experiments discussed above, physical constraints could include the fact that the demonstrator's hands are occupied (e.g. as in Gergely et al., 2002).

For each set of selected states $\vec{o}_s$ and resulting action sequence $\vec{a}_t$, the imitator estimates the cost of $\vec{a}_t$. We tentatively suggest the cost is expressed in terms of energy expenditure. The estimated energetic cost $\hat{E}(\vec{a}_t)$ is compared with the estimated cost of the demonstrated action sequence $\hat{E}(\vec{o}_t)$ calculating the cost difference ΔE as,

$$\Delta E = |\hat{E}(\vec{o}_t) - \hat{E}(\vec{a}_t)| \cdot S(\vec{o}_t) \quad (1)$$

In equation 1, the parameter $S(o_t)$ indicates the saliency of the demonstrated state $\vec{o}_t$. This weighing allows discounting part of the demonstrated action sequence $\vec{o}_t$ in favour of salient action outcomes. The saliency of (part of) a demonstration could be computed using existing approaches to visual saliency methods developed in the field of human-machine interaction (e.g. Scassellati, 2002; He et al., 2014). In the experiments reported in the current paper, we do not vary this parameter and fix it at a value of 1. However, experimental evidence strongly suggests saliency is an important factor (e.g., Carpenter et al., 2005; Liu and Spelke, 2017) and we plan to expand the model in this direction.

At first, the set of selected states $\vec{o}_s$ only contains the initial and final observed states, i.e., $\vec{o}_s = \{o_0, o_T\}$. However, the set is iteratively expanded by adding more intermediate states. Therefore, the set of selected states $\vec{o}_s$ will eventually approach the observed action sequence $\vec{o}_t$. In consequence, ΔE approaches zero as the set $\vec{o}_s$ is expanded. When the value of ΔE is below a certain threshold τ_E, expanding $\vec{o}_s$ is terminated and the current set $\vec{o}_s$ (with the exception of the initial state o_0) is taken to contain the subgoals in the observed behaviour. The set $\vec{o}_s$ contains the minimum number of subgoals that are required to explain the (cost of the) observed behaviour $\vec{o}_t$. Also, notice that the iterative process implies that when $\Delta E(\vec{o}_s = \{o_0, o_T\}) < \tau_E$, the imitator will simply plan an action sequence to attain the final state demonstrated – hence, no imitation of any intermediate goal will take place. In this case, the imitator assumes that the observed behaviour $\vec{o}_t$ can be inadequately explained by assuming the demonstrator is simply attempting to reach the final goal. No subgoals need to be assumed.

Obviously, expanding the set $\vec{o}_s$ can be done in many ways. Here, we propose that on each iteration additional states are selected at time instances intermediate between the currently selected states. At first, only two states will be selected,

$$\vec{o}_s = \{o_0, o_T\}. \tag{2}$$

On the next iteration, an additional state in between these two will be added: $\vec{o}_s = \{o_0, o_{\frac{T}{2}}, o_T\}$. Next, the set will be expanded to $\vec{o}_s = \{o_0, o_{\frac{T}{4}}, o_{\frac{T}{2}}, o_{\frac{3T}{4}}, o_T\}$. In other words, at the nth iteration the length of $\vec{o}_s$ is given by $|\vec{o}_s| = 1 + 2^{n-1}$.

In equation 1, $\vec{a}_t$ denotes the action sequence planned to attain the selected states $\vec{o}_t$. Hence, we assume that the agent can plan an action sequence passing through a number of selected goal states. In addition, we assume that the agent can plan this taking into account the physical constraints C of the demonstrator. This assumption represents the most challenging cognitive ability supposed under our model. However, evidence suggests that infants are capable of planning actions under physical constraints (Upshaw and Sommerville, 2015; Claxton et al., 2003).

Figure 1 illustrates the process outlined above. Figure 1b depicts a hypothetical path followed by a demonstrator (depicted as a black line) from start to goal. Observing this path, an imitator iteratively selects an increasing number of states (here: n = 2, 3 and 4, respectively) from the demonstrated path. Selecting only the start and goal position (fig. 1c) leads to a large cost difference ΔE (fig. 1f).

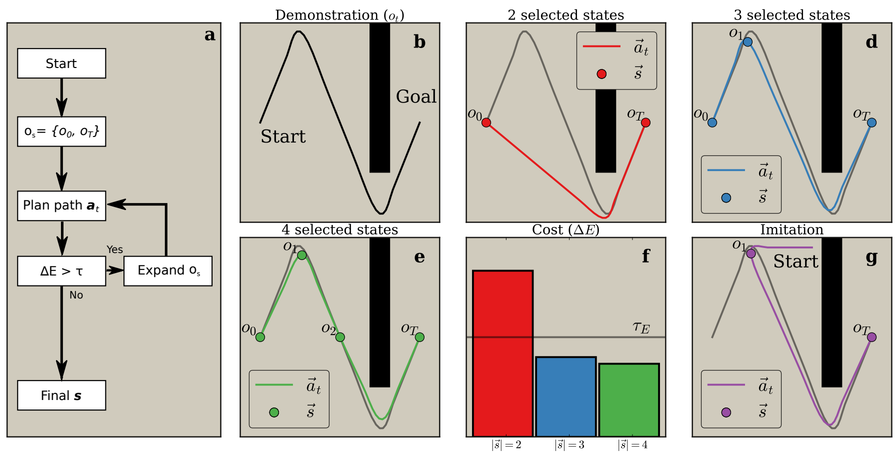

Figure 1: Illustration of the process of selecting states $\vec{o}_s$ of the demonstrated action sequence $\vec{o}_t$. (a) flow chart depicting the process of selecting $\vec{o}_s$. (b) The hypothetical path taken by a demonstrator (black line) from start to goal. Notice the demonstrated path consists of both an unnecessary curve (first) and necessary curve (to negotiate the black obstacle). (d) This panel illustrates the planned path $\vec{a}_t$ for $\vec{o}_s$ containing only the initial state and final state. Notice that this results in a discrepancy between the paths $\vec{a}_t$ and $\vec{o}_t$. In particular, the first curve is not included in $\vec{a}_t$. This will result in a value for ΔE that is larger than τ_E. Hence, additional states will be added to $\vec{o}_s$. This is illustrated in panels d-e where $\vec{o}_s$ contains 3 and 4 selected states respectively. By selecting a single additional state in panel d, the match between paths $\vec{a}_t$ and $\vec{o}_t$ increases (and $\Delta E < \tau_E$, panel f). At this point, the iterative expansion of $\vec{o}_s$ is terminated and adding further states does not markedly decrease ΔE (panels e and f). Finally, panel g depicts the path the imitator would follow (note, it starts from a different location than the demonstrator). Omitting state o_0 from $\vec{o}_s$, it goes to o_T via o_1, thereby imitating the unnecessary (and energetically demanding) detour shown by the demonstrator.

The reason is that the planned action $\vec{a}_t$ does not include the deviation present in the demonstrator's path. However, by including an additional third state (fig. 1d), the imitator's planned action sequence $\vec{a}_t$ better matches the demonstrated path (and energetic cost). Adding more states does not improve the match (fig. 1 e and f). Hence, the imitator will copy the three states (depicted in fig. 1d). The imitated path is shown in fig. 1g.

2.3 Accounting for experimental data

In this section, we explain how the CDM can account for the relevant findings in the literature on rational imitation in human infants. In particular, we discuss the results of Carpenter et al. (2005) mentioned above because these allow us to illustrate all aspects of the CDM. The relevant findings of these authors are depicted in figure 2. To recapitulate, these authors reported (among other results) that 18-month old children were most prone to copy the actions demonstrated by an experimenter when a toy mouse was moved across a table top using a hopping motion (Figure 2a, condition 1). They copied the action less faithfully when the mouse was slid across the table (Figure 2a, condition 2) and even less so when a small toy house was present at the final location (Figure 2a, condition 3). Finally, moving the mouse to the toy house using a hopping motion was more likely to be copied (Figure 2a, condition 4) than when it was moved in a sliding motion (Figure 2a, condition 3).

First, the CDM accounts for the increased action copying associated with the hopping motion with respect to the sliding motion (conditions 1 and 3 vs. 2 and 4) by assuming that the former is more energetically demanding. In other words, the hopping motion is assumed to result in a large value for the first term in equation 1 if not faithfully modelled using sufficient number of states $\vec{o}_t$. Hence, the CDM predict the hopping motion should be more faithfully copied.

Second, the CDM can account for the reduction in copying due to the introduction of the house (conditions 1 and 2 vs. 3 and 4) in terms of the saliency parameter, $S(o_t)$. We assume that the event of inserting the toy into the house is more salient than the preceding actions. Hence, the saliency function $S(o_t)$ discounts the preceding action. In absence of the house, no such discounting occurs (see fig. 2b).

Finally, we briefly discuss how the CDM accommodates the experimental results using the popular head touch paradigm. The model assumes that whenever a demonstrator with free hands performs a head touch, the first term of equation 1 will be large. Indeed, the energetic demand of the head touch will be compared with that of a simple hand touch. In contrast, when the demonstrator's hands are occupied Gergely et al. (2002), the infant is assumed to plan an action taking into account these constraints (remember that $\vec{a}_t$ in equation 1 should be regarded as shorthand for $\vec{a}_t = f(\vec{o}_s, C)$ with C representing the physical constraints of the demonstrator). We assume that this will result in infants covertly planning a head-touch themselves. As such, this will result in lower values for the first term of equation 1 and, therefore, a lower degree of action copying. It could be objected that is unlikely that children come up with a head touch as a way of dealing

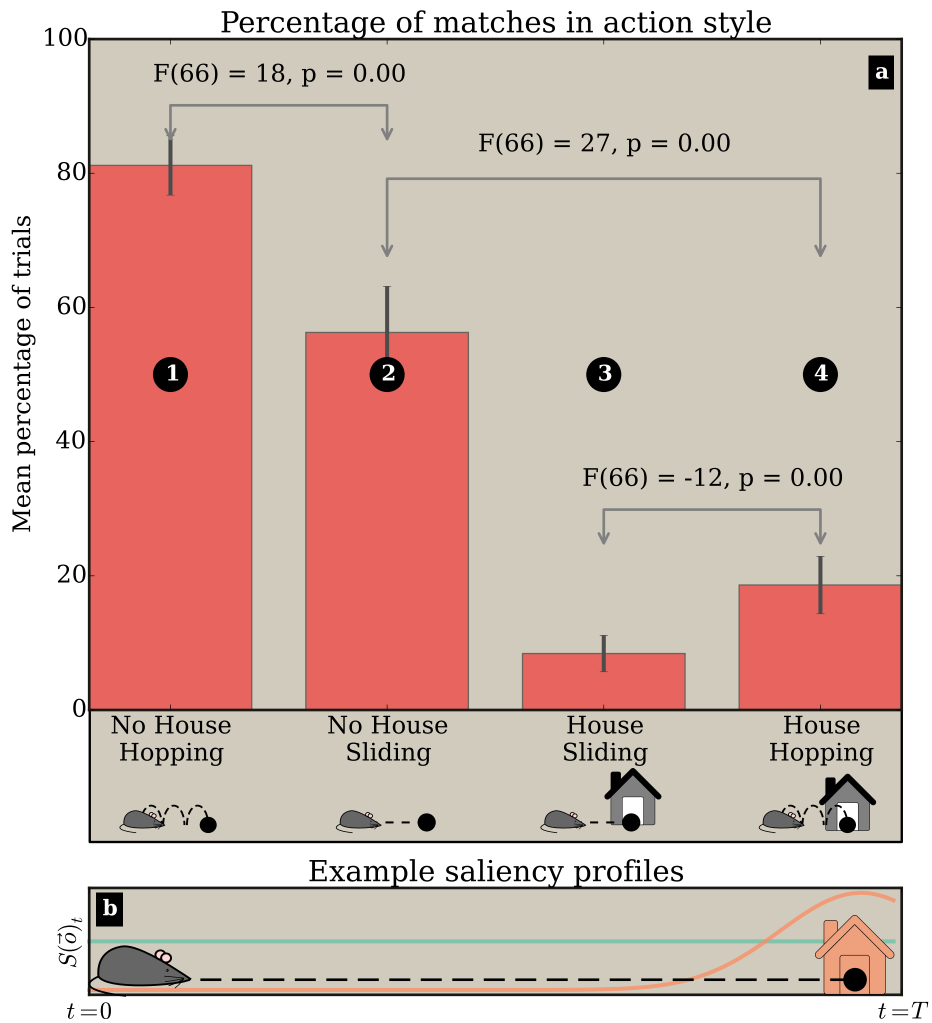

Figure 2: (a) Data from Carpenter et al. (2005). The statistical tests are our post-hoc tests, i.c., t-tests based on the reported means and standard deviations. (b) Examples of assumed salience functions, $S(o_t)$. See text for details.

with the constraints. However, a small percentage of infants who have not been shown the head touch still choose to touch the lamp with their heads (Paulus et al., 2013), especially younger infants (Zmyj et al., 2009). Hence, it is not beyond plausibility that the apparatus used in these experiments spontaneously elicits head pushing as a solution to deal with the constraint of occupied hands. Incidentally, perceiving the lamp being switched might induce discounting the preceding action through the saliency. However, this would not result in head touch being ignored as the end state in these experiments involves the experimenter touching the lamp with her head. Hence, even if the saliency parameter results in only the final state of the demonstration to be copied, the head touch will still be imitated.

In contrast to an account based on attentional processes (Beisert et al., 2012), the CDM does not require conditions to recruit different levels of attention for rational imitation to occur (Paulus et al., 2013; Kolling et al., 2014). However, attentional processes can be accounted for using the term $S(\vec{o}_t)$ (eq. 1). Our model also differs in its predictions with the 'motor resonance' account of rational imitation (Paulus et al., 2011). As mentioned, 12-month old – but not 9-month old – infants have been shown to ignore the head touch action of a model with hands fixed to the table (Zmyj et al., 2009). Our model could explain these findings by assuming that 12-month olds are better at accounting for a model's constraints. In contrast, the motor resonance account would need to account for this by assuming that infants are more susceptible to 'a lack of motor resonance' at 12 months than at 9 months. This would imply that infants are less good at copying motor behavior at 12 months than at 9 months.

3 Methods

We used two NAO humanoid robots (Aldebaran) in this study, a blue and a red version. The blue robot was assigned the role of the demonstrator. The red robot was assigned the role of the imitator. Experiments were carried out in a 3 by 2.5 m arena. An overhead 3D tracking system (Vicon) consisting of 4 cameras was used to monitor the position and orientation of the robots at a rate of 30 Hz. The robots were equipped with a clip-on helmet fitted with a number of reflective beads used by the tracking system to localize the robots. In addition to the robots, the arena contained three small tables each with a unique pattern of reflective beads. These served as obstacles and a target position.

The custom-written Python software controlling the robots implemented a path planning algorithm (figure 7). This algorithm overlaid the arena with a rectangular graph with nodes spaced 10 cm apart (Schult and Swart, 2008). Nodes closer than 0.5 m to an obstacle were removed from the graph. A path between the current position of a robot and the desired goal location was planned by finding the shortest path of connected nodes between the node closest to the robot's current position and the node closest to the goal position. By removing the nodes closer than 0.5 m to an obstacle, the path planning algorithm ensured the robots steered well clear of obstacles. In the current paper, the estimated energetic costs $\hat{E}(\vec{o}_t)$ and $\hat{E}(\vec{a}_t)$ are approximated by the length of the planned and observed paths, respectively. For robots moving at a constant speed, this is a fair approximation.

4 Experiment 1: Modelling Experimental Findings

Figure 3 illustrates the four conditions of experiment 1. In the first condition, the demonstrator is not hampered by obstacles. Hence, it moves towards the goal position using a direct path (fig. 3a). In the second condition (fig. 3b), the demonstrator could approach the goal using a direct path. However, the demonstrator approaches the goal by a detour. In the third condition, obstacles between the demonstrator and the goal prevent a direct path. The path planning algorithm yields a path circumventing the obstacles (fig. 3c). Finally, in the fourth condition (fig. 3d), the demonstrator was sent to the goal by the same path as in condition 2. Hence, in condition 4, the detour was not planned by the path planner but explicitly programmed. Condition 3 and 4 should lead to the same outcome. However, methodologically, condition 4 confirms that differences between conditions 1 & 2 and 2 & 3 are not due to the way the motion of the demonstrator is planned. In other words, condition 4 demonstrates that the (internal) intention of the demonstrator is not taken into account by the imitator.

The critical conditions, in modelling the experimental results regarding rational imitation in infants (e.g., Gergely et al., 2002; Meltzoff, 1988), are conditions 2 and 3. In both conditions, the demonstrator does not take the direct path to the goal. The difference between these conditions, however, is the presence of an obstacle in condition 3. In this condition, the obstacle forces the demonstrator to take the longer path. This is analogous to a demonstrator switching on the lamp with her head when her hands are occupied in the sense that the constraints of the situation necessitate the less direct (and energetically inefficient) mode of operation. Critically, the CDM assumes that the robot (infant) plans an indirect path (head touch) to cope with the constraints introduced by the obstacle (occupied hands). Hence, the robot (infant) is predicted not to imitate the indirect path (head touch). In contrast, in condition 2, given no obstacle (analogous to the free hands condition in behavioural experiments) the imitator will plan a direct path (a hand touch). The planned direct path (head touch) is assumed to differ sufficiently (in terms of energy expenditure) from the demonstrated indirect path (head touch) to incur imitation.

Figure 4 depicts the results of experiment 1. In condition 1, the demonstrator takes the direct route to the goal position (fig 4a). Calculating Δ_E for $\vec{o}_t$ with two states results in a value lower than τ_E (fig 4e and fig. 6). Hence, imitator only retains the final goal o_T as policy. Therefore, the imitator proceeds directly to the goal, using a direct path (fig 4i).

In condition 2, the demonstrator takes a detour to the goal, in spite of a direct path being possible (fig 4b). Calculating Δ_E for $\vec{o}_t$ with two states results in a

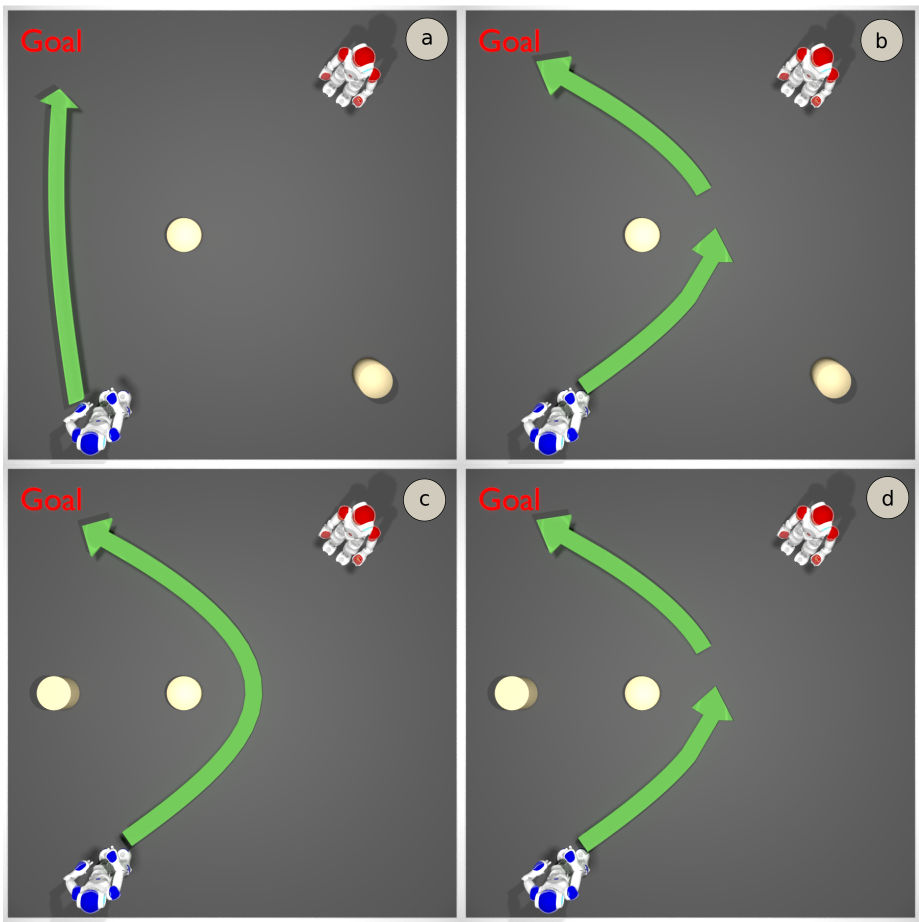

Figure 3: Illustration of the four conditions in experiment 1. The blue robot is the demonstrator. The red robot is the imitator. The green arrows depict the path taken by the demonstrator. Note that in panel c the demonstrator cannot pass between the two round obstacles. Details in text.

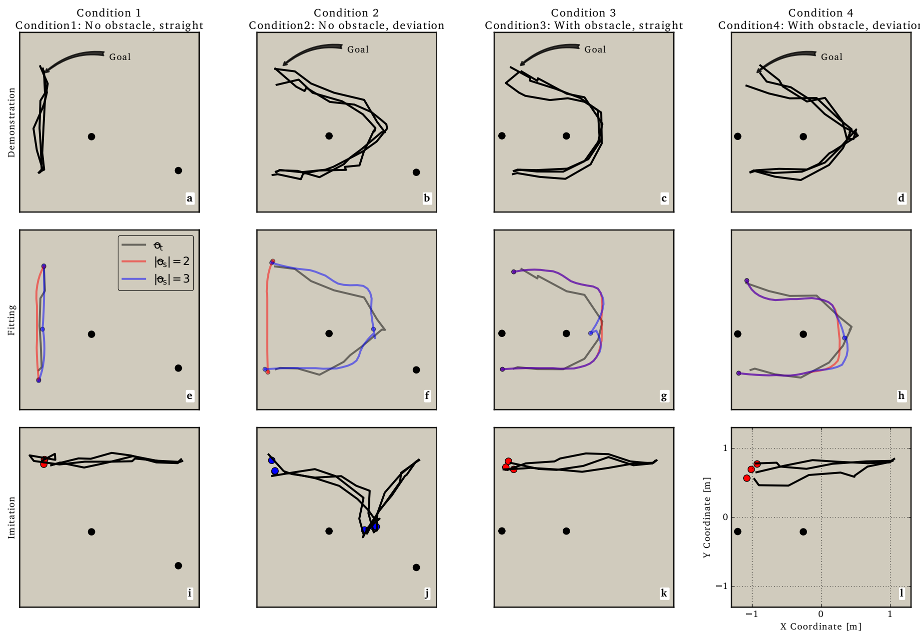

Figure 4: Results of experiment 1. Panels a-d: traces of the paths taken by the demonstrator for conditions 1-4, respectively. The black circles denote the position of two obstacles. Panels e-h depict the process of iteratively expanding $\vec{o}_t$. In red, the planned path $\vec{a}_t$ is shown for $\vec{o}_s$ with two states, i.e., $\vec{o}_t = \{o_0, o_T\}$. In blue, the planned path $\vec{a}_t$ is shown for $\vec{o}_s$ with three states, i.e., $\vec{o}_s = \{o_0, o_{T/2}, o_T\}$. In conditions 1,3 & 4, the red path $\vec{a}_t$ matches the demonstrated path $\vec{o}_t$ well. This is, $\Delta E < \tau_E$. In condition 2 red path $\vec{a}_t$ does not match the demonstrated path $\vec{o}_t$ ($\Delta E > \tau_E$). In contrast, the blue path $\vec{a}_t$ satisfies the requirement $\Delta E < \tau_E$. Here only the resulting paths $\vec{a}_t$ for $|\vec{o}_s|$ equal to 2 and 3 are shown. However, the $\vec{a}_t$ for $|\vec{o}_s|$ equal to 5, 9 and 17 were also evaluated. Their resulting weighted cost differences Δ_E are plotted in figure 6. Panels i-l depict the imitated behaviour for each of the four conditions. Notice that the imitator does not start from the same position as the demonstrator. In conditions 1, 3 & 4, the imitator proceeds to the goal (i.e., o_T) by a direct path. In condition 2, the set of selected states contains three states. Hence, the imitator proceeds to o_T via an intermediate state, i.e., $o_0 \to o_{T/2} \to o_T$.

value higher than τ_E (fig 4f and fig. 6). In contrast, calculating Δ_E for $\vec{o}_t$ with three states results in a value lower than τ_E (fig 4f and fig. 6). Hence, the policy copied will include an additional sub goal en route to the goal. The imitator proceeds to this intermediate goal before going to the final goal (fig 4j). The blue path $\vec{a}_t$, based on $\vec{o}_s$ with three states, in fig. 4e satisfies the requirement $\Delta E < \tau_E$. Hence, the policy copied will include an additional subgoal en route to the goal. The imitator proceeds to this intermediate goal before going to the final goal (fig. 4h).

In conditions 3 & 4, the demonstrator reaches the goal by a detour fig 4c & d). However, the presence of an obstacle makes this necessary. Indeed, the planned path $\vec{a}_t$ from o_0 to o_T will also contain this detour. As such, the value of Δ_E will be small, even for $\vec{o}_s = \{o_0, o_T\}$ (fig 4g & h and fig. 6). As such, the imitator proceeds directly to the final goal (fig 4k & l).

In condition 3, the demonstrator reaches the goal via a detour fig. 4c). However, the presence of an obstacle makes this necessary. Indeed, the path $\vec{a}_t$ planned by the imitator from o_0 to o_T (i.e. $|\vec{o}_s| = 2$) will also contain this detour. As such, the value of ΔE will be small, even for $|\vec{o}_s| = 2$ (fig. 4f and j). The red path $\vec{a}_t$ for $|\vec{o}_s| = 2$ (fig. 4f) matches the demonstrated path $\vec{o}_t$ sufficiently. As a result, the imitator proceeds directly to the final goal (fig. 4i), as it did in condition 1.

Experiment 1 was aimed at modelling the basic findings of the behavioural experiments regarding rational imitation in infants (Meltzoff, 1988; Gergely et al., 2002; Zmyj et al., 2009; Beisert et al., 2012; Paulus et al., 2011). As mentioned above, these authors showed that children copy the head-touch demonstrated by adults only if the adult's hands were unrestricted. In our robot experiments, the imitator only copied the demonstrated detour if the demonstrator was not forced to take this detour by the obstacles (Condition 2, fig. 4b, e and h). In contrast, when the demonstrator took the same path – but was forced to do so on account of an obstacle – the imitator disregarded the detour (Condition 3, fig. 4c, f and i). As such, conditions 2 and 3 reveal our robots modelling the behaviour of infants in the behavioural experiments discussed above.

5 Experiment 2: Learning Better Policies

In our view, the behavioural experiments concerning rational imitation cited above can be considered as cases of pathological imitation (Winfield and Erbas, 2011). That is, the behavioural experiments are set up to induce imitation in spite of the behaviour being inefficient, i.e., the head touch is a less efficient way of switching on the light than a hand touch. The experiments of Lyons et al. (2007) and Keupp et al. (2013) illustrate how easily children can be tricked into imitating inefficient behaviour. In these experiments, the demonstrating adult exhibited a range of action irrelevant to attain a given goal. Nevertheless, the infants tended to copy these actions – even when explicitly instructed not to copy any 'silly' behaviour. However, when not experimentally controlled, adults' behaviour can generally be assumed to be more efficient or more adaptive than

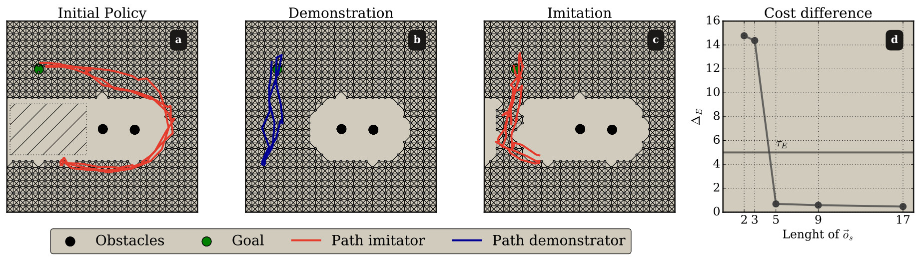

Figure 5: Results of experiment 2. The paths of both the imitator (red paths) and demonstrator (blue paths) for three trials are plotted.The grids in the background of panels a-c represent the graph used in path planning by the imitator (panels a & c) and the demonstrator (panel b). Panel a: the initial policy of the imitator in reaching the goal position involves a detour. Part of the graph used by the imitator for path planning has been taken out (the hatched region). Panel b: the demonstrator approaches the goal in a straight line (its path planning graph has not been lesioned). Panel c: the imitator, based on observing the demonstrator's policy, adopts a more efficient policy. Panel d: cost difference ΔE as a function of the number of states in $\vec{o}_s$ averaged over the three trials

that of infants. Under these conditions, as will be shown below, the mechanism proposed above for selecting policies for imitation is adaptive.

In this section of the paper, we present a robotic experiment showing that the CDM can also select more efficient policies if these are observed in a demonstrator. Indeed, by virtue of equation 1, the CDM can select policies for explorative imitation that are *less* costly than the current policy. The current policy of the robot amounts to the planned route $\vec{a}_t$ for $\vec{o}_s$ with only two states (o_0 and o_T). For $|\vec{o}_s| = 2$, the robot will generate a plan reaching the end goal without taking into account the demonstrated behaviour. If the observed policy $\vec{o}_t$ is significantly *less* costly than the currently held policy, ΔE will be larger than τ_E (by virtue of the absolute value operator in equation 1). This will trigger the expansion of the set of intermediate goals $\vec{o}_s$ until ΔE is smaller than τ_E.

In experiment 2, the imitator starts with a policy that is clearly not optimal. When going from the start position to the goal, the imitator takes an unnecessary detour (fig. 5a). This detour is caused by the imitator's path planning algorithm not considering the locations in the hatched area (fig. 5a). In effect, the hatched area is not part of the search space considered by the path planning algorithm. In contrast, panel b of figure 5 shows the demonstrator moving in a straight line from start to goal – as depicted in this panel, the whole arena is part of the demonstrator's search space. As such, the demonstrator can find a shorter path to the goal. Considering the observed behaviour $\vec{o}_t$, the imitator iteratively expands a set of selected states $\vec{o}_s$ from the demonstrated states $\vec{o}_t$. Each state o_s in $\vec{o}_s$ corresponds to a position of the demonstrator in the arena. By adding states o_s to $\vec{o}_s$ the imitator effectively expands its path planning search space. Iteratively expanding the set of selected states $\vec{o}_s$ will eventually lead to filling in the part of the search space that was initially not available to the imitator (in panel a). Indeed, in effect, a corridor between start and goal position is built (figure 5c). When this corridor is established the value $\Delta E < \tau_E$ (at $|\vec{o}_s| = 5$, panel d) and expansion of $\vec{o}_s$ is stopped. Eventually, the imitator imitates the shorter path, as shown in fig. 5c.

6 Discussion

Selective and rational imitation shown by children would be a beneficial capacity for robots (Gergely, 2003). Unfortunately, no computationally explicit model of rational imitation in infants is available. In this paper, we have presented a formalization that captures the most relevant aspects of the behaviour of infants in experiments. The CDM can be considered as a formalized version of the teleological reasoning hypothesis, which is underspecified (See Zmyj and Buttelmann, 2014, for references). As such, the CDM is explicit enough to be implemented on robots, as demonstrated above.

While our model is primarily conceived as a practical method for supporting rational imitation in robots, it can also be evaluated for its ability to explain infant behavior. Considering the CDM as a psychological model of rational imitation in infants allows making a number of predictions. First, the CDM predicts that the surface structure of the observed action is not important in determining whether the action will be imitated by infants. Observed actions that have similar associated predicted costs, $\hat{E}(\vec{o}_t)$, will induce similar levels of imitation. Experimental work, using paradigms akin to those used to evaluate over-imitation (Lyons et al., 2007; Keupp et al., 2013), could test this prediction. These experiments use arbitrary complex action sequences and evaluate the extent to which they are copied by the child. According to the CDM, changing the order of the actions in a sequence should not influence the level of imitation. A second prediction that follows from our model is that the sign of the cost difference, $\hat{E}(\vec{o}_t) - \hat{E}(\vec{a}_t)$, does not influence the level of imitation. Indeed, we postulated that only the absolute value of the difference is taken into account in calculating ΔE. Therefore, the CDM predicts that both actions that are more costly and more efficient than the current strategy known to infants should lead to imitation. Again, this is a testable prediction of the CDM. A third prediction of the CDM is that the two previous predictions can be modulated by targeted manipulations of the saliency of parts of the action sequences used.

Acknowledgements

This work was supported by grant EP/L024861/1 ('Verifiable Autonomy') from the Engineering and Physical Sciences Research Council (EPSRC).

References

Argall, B. D., Chernova, S., Veloso, M., and Browning, B. (2009). A survey of robot learning from demonstration. *Robotics and autonomous systems*, 57(5):469–483.

Baldwin, D. A., Baird, J. A., Saylor, M. M., and Clark, M. A. (2001). Infants Parse Dynamic Action. *Child Development*, 72(3):708.

Beisert, M., Zmyj, N., Liepelt, R., Jung, F., Prinz, W., and Daum, M. M. (2012). Rethinking 'rational imitation' in 14-month-old infants: A perceptual distraction approach. *PLoS ONE*, 7(3):1–5.

Billard, A., Calinon, S., Dillmann, R., and Schaal, S. (2008). Robot programming by demonstration. In *Handbook of robotics*, pages 1371–1394.

Breazeal, C. and Scassellati, B. (2002). Robots that imitate humans. *Trends in Cognitive Sciences*, 6(11):481–487.

Buttelmann, D., Carpenter, M., Call, J., and Tomasello, M. (2007). Enculturated chimpanzees imitate rationally. *Developmental Science*, 10(4):F31–F38.

Carpenter, M. and Call, J. (2006). The question of 'what to imitate': inferring goals and intentions. In Nehaniv, C. L. and Kirstin, D., editors, *Imitation and Social Learning in Robots, Humans and Animals Behavioural, Social and Communicative Dimensions*, pages 135–152. Cambridge University Press.

Carpenter, M., Call, J., and Tomasello, M. (2005). Twelve- and 18-month-olds copy actions in terms of goals. *Developmental Science*, 8(1):13–20.

Claxton, L. J., Keen, R., and McCarty, M. E. (2003). Evidence of motor planning in infant reaching behavior. *Psychological Science*, 14(4):354–356.

Dautenhahn, K. and Nehaniv, C. (2002). *Challenges in Building Robots That Imitate People*, pages 363–390. MIT Press.

Erbas, M. D., Winfield, A. F., and Bull, L. (2013). Embodied imitation-enhanced reinforcement learning in multi-agent systems. *Adaptive Behavior*, page 1059712313500503.

Gariépy, J.-F., Watson, K. K., Du, E., Xie, D. L., Erb, J., Amasino, D., and Platt, M. L. (2014). Social learning in humans and other animals. *Frontiers in Neuroscience*, 8(58).

Gattis, M., Bekkering, H., and WohlSchlager, A. (2002). Goal-directed imitation. In Meltzoff, A. N. and Prinz, W., editors, *The Imitative Mind: Development, Evolution and Brain Bases Edited*, pages 183–205. Cambridge University Press.

Gellén, K. and Buttelmann, D. (2017). Fourteen-Month-Olds Adapt Their Imitative Behavior in Light of a Model ' s Constraints. 2017.

Gergely, G. (2003). What should a robot learn from an infant? Mechanisms of action interpretation and observational learning in infancy. *Connection Science*, 15(4):191–209.

Gergely, G., Bekkering, H., and Kiraly, I. (2002). Rational imitation in preverbal infants. *Nature*, 415(6873):755.

He, H., Ge, S. S., and Zhang, Z. (2014). A saliency-driven robotic head with bio-inspired saccadic behaviors for social robotics. *Autonomous Robots*, 36(3):225–240.

Hesslow, G. (2002). Conscious thought as simulation of behaviour and perception. *Trends in Cognitive Sciences*, 6(6):242–247.

Hesslow, G. (2012). The current status of the simulation theory of cognition. *Brain Research*, 1428:71–79.

Jones, S. S. (2009). The development of imitation in infancy. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1528):2325–2335.

Keupp, S., Behne, T., and Rakoczy, H. (2013). Why do children overimitate? Normativity is crucial. *Journal of Experimental Child Psychology*, 116(2):392–406.

Kolling, T., Óturai, G., and Knopf, M. (2014). Is selective attention the basis for selective imitation in infants? An eye-tracking study of deferred imitation with 12-month-olds. *Journal of Experimental Child Psychology*, 124:18–35.

Laland, K. N. (2004). Social learning strategies. *Animal Learning & Behavior*, 32(1):4–14.

Legare, C. H. and Nielsen, M. (2015). Imitation and innovation: the dual engines of cultural learning. *Trends in cognitive sciences*, 19(11):688–699.

Liu, S. and Spelke, E. S. (2017). Six-month-old infants expect agents to minimize the cost of their actions. *Cognition*, 160:35–42.

Lyons, D. E., Young, A. G., and Keil, F. C. (2007). The hidden structure of overimitation. *Proceedings of the National Academy of Sciences*, 104(50):19751–19756.

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.

Nehaniv, C. L. (2007). Nine billion correspondence problems. In Nehaniv, C. L. and Dautenhahn, K., editors, *Imitation and Social Learning in Robots, Humans and Animals*, pages 35—-46. Cambridge University Press.

Nehaniv, C. L. and Dautenhahn, K. (2001). Like me?-measures of correspondence and imitation. *Cybernetics & Systems*, 32(1-2):11–51.

Paulus, M., Hunnius, S., and Bekkering, H. (2013). Examining functional mechanisms of imitative learning in infancy: does teleological reasoning affect infants' imitation beyond motor resonance? *Journal of experimental child psychology*, 116(2):487–98.

Paulus, M., Hunnius, S., Vissers, M., and Bekkering, H. (2011). Imitation in infancy: Rational or motor resonance? *Child Development*, 82(4):1047–1057.

Poulin-Dubois, D., Brooker, I., and Polonia, A. (2011). Infants prefer to imitate a reliable person. *Infant Behavior and Development*, 34(2):303–309.

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

Scassellati, B. (2002). Theory of mind for a humanoid robot. *Autonomous Robots*, 12(1):13–24.

Schaal, S., Ijspeert, A., and Billard, A. (2003). Computational approaches to motor learning by imitation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 358(1431):537–547.

Schult, D. A. and Swart, P. (2008). Exploring network structure, dynamics, and function using networkx. In *Proceedings of the 7th Python in Science Conferences (SciPy 2008)*, volume 2008, pages 11–16.

Schwier, C., Maanen, C. V., Carpenter, M., and Tomasello, M. (2006). Rational Imitation in 12-Month-Old Infants. *Infancy*, 10(3):303–311.

Tomasello, M. (2009). *The cultural origins of human cognition*. Harvard University Press.

Upshaw, M. B. and Sommerville, J. A. (2015). Twelve-month-old infants anticipatorily plan their actions according to expected object weight in a novel motor context. *Frontiers in public health*, 3.

Vanderelst, D., Ahn, R. M., and Barakova, E. I. (2009). Simulated trust: A cheap social learning strategy. *Theoretical population biology*, 76(3):189–196.

Winfield, A. F. and Erbas, M. D. (2011). On embodied memetic evolution and the emergence of behavioural traditions in robots. *Memetic Computing*, 3(4):261–270.

Zmyj, N. and Buttelmann, D. (2014). An integrative model of rational imitation in infancy. *Infant Behavior and Development*, 37(1):21–28.

Zmyj, N., Buttelmann, D., Carpenter, M., and Daum, M. M. (2010). The reliability of a model influences 14-month-oldsâĂŹ imitation. *Journal of Experimental Child Psychology*, 106(4):208–220.

Zmyj, N., Daum, M. M., and Aschersleben, G. (2009). The Development of Rational Imitation in 9- and 12-Month-Old Infants. *Infancy*, 14(1):131–141.

A Appendix

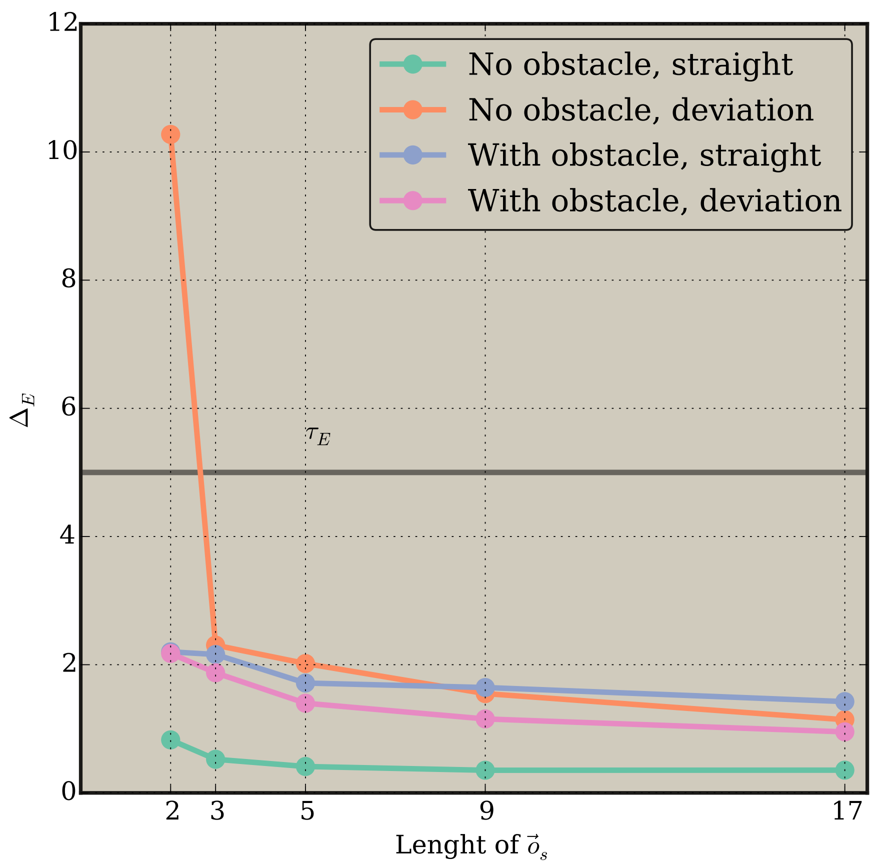

Figure 6: The values of ΔE as function of the number of selected states in $\vec{o}_s$ for the four conditions in experiment 1.

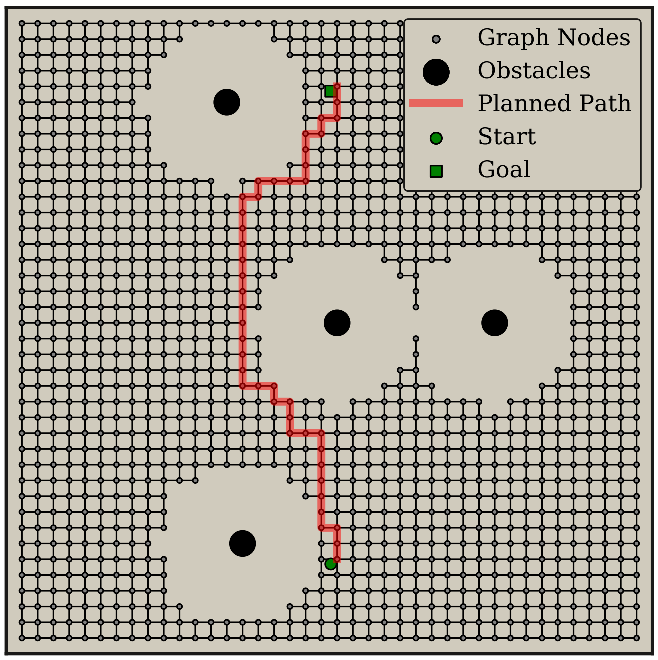

Figure 7: Plot illustrating the path planning algorithm used by the robots. The plot depicts a hypothetical arena featuring 4 obstacles. The path planning algorithm overlay the arena with a graph of closely nodes spaces. The path planning algorithm searches for the shortest path of graph nodes between (1) the node closest to the current position of the robot and (2) the node closest to the goal position. Nodes that are too close near an obstacle are removed from the network to force the path planning to steer clear of obstacles.