Properties for Efficient Demonstrations to a Socially Guided Intrinsically Motivated Learner

Sao Mai Nguyen, Pierre-Yves Oudeyer

Sao Mai Nguyen and Pierre-Yves Oudeyer. Properties for Efficient Demonstrations to a Socially Guided Intrinsically Motivated Learner, in Proceedings of the 21st IEEE International Symposium on Robot and Human Interactive Communication, 2012, Paris, France.

Properties for Efficient Demonstrations to a Socially Guided Intrinsically Motivated Learner

Sao Mai Nguyen[1] and Pierre-Yves Oudeyer[1]

***Abstract*— The combination of learning by intrinsic motivation and social learning has been shown to improve the learner's performance and gain precision over a wider range of motor skills, with for instance the SGIM-D learning algorithm [1]. Nevertheless, this bootstrapping a-priori depends on the demonstrations made by the teacher. We propose in this paper to examine this dependence: to what extend the quality of the demonstrations can influence the learning performance, and which are the characteristics of a good demonstrator. Results on a fishing experiment highlights the importance of the difficulty of the demonstrated tasks, as well as the structure of the actions demonstrated.**

I. Introduction

Developmental robots, similarly to animal or human infants, need to be endowed with exploration mechanisms which continuously push them toward learning new skills and new situations [2], [3] in order to adapt to their changing environment and users' needs. Exploration strategies developed in the recent years can be classified into two broad interacting families: 1) socially guided exploration [4]–[7]; 2) internally guided exploration and in particular intrinsically motivated exploration [2], [8]–[10].

A. Intrinsic Motivation vs Social Guidance

In developmental robotics, intrinsic motivation, which consist in meta-exploration mechanisms monitoring the evolution of learning performances [11]–[13] with heuristics defining the notion of interest used in an active learning framework [14]–[16], is often studied separately from socially guided learning where the learner can interact with teaching agents by mimicry, emulation or stimulus enhancement [17], [18]. While many forms of socially guided learning can be seen as extrinsically driven learning, in the daily life of humans, the two strongly interact, and on the contrary push their respective limits (cf. table I).

	Intrinsically Motivated Exploration	Socially Guided Exploration
Pros	Independent from human, broad task repertoire	transfer knowledge from human to robot
Cons	High-dimensionality, unboundedness	Teacher's patience & ambiguous input, correspondence problem

TABLE I: Advantages and disadvantages of the two exploration strategies.

Social guidance can drive a learner into new intrinsically motivating spaces or activities which it may continue to explore alone and for their own sake. Robots may acquire new strategies for achieving intrinsically motivated activities while observing examples. One might either search in the neighbourhood of the good example, or eliminate from the search space the bad example.

Conversely, as learning that depends highly on the teacher is limited by ambiguous human input or the correspondence problem [19], and would require too much time from the teacher, some autonomous learning is needed. While self-exploration fosters a broader task repertoire of skills, exploration guided by a human teacher tends to be more specialised, resulting in fewer tasks that are learnt faster. Combining both can thus bring out a system that acquires a wide range of knowledge which is necessary to scaffold future learning with a human teacher on specifically needed tasks, as proposed in [20]–[22].

Social learning has been combined with reinforcement learning [21]–[23]. However, these approaches are restricted to a single task. We would like a system that learns not only for a single task, but for a continuous field of tasks. Such a multi-goal system has been presented in [20], [24], where unfortunately the representation of the environment and actions is symbolic and discrete in a limited and preset world, with few primitive actions possible.

B. SGIM-D Combines Social Guidance and Intrinsic Motivation

In an initial work to address multi-task learning, we proposed the Socially Guided Intrinsic Motivation by Demonstration (**SGIM-D**) algorithm which merges socially guided exploration and intrinsic motivation based on SAGG-RIAC algorithm [25], to reach goals in a continuous task space, in the case of a complex, high-dimensional and continuous environment [1]. SGIM-D has been shown to efficiently take advantage of the demonstrations to explore unknown subspaces, and to focus on interesting subspaces of the task space. It also takes advantage of the autonomous exploration of SAGG-RIAC to improve its performance and gain precision in the absence of the teacher in a wide range of tasks. The possible tasks to complete are yet of the same nature, still they belong to multi-tasks problems, because they are infinite in number and belong to a continuous space. Two tasks may require very different actions to reach them.

Nevertheless, in that first study, we did not examine how dependent on the demonstrations the learner's performance is. We propose in this paper to study to what extend the quality of the demonstrations can influence the learning performance, and which are the characteristics that make a good demonstrator.

[1]Flowers Team, INRIA and ENSTA ParisTech, France. nguyensmai at gmail.com, pierre-yves.oudeyer at inria.fr

Sao Mai Nguyen and Pierre-Yves Oudeyer. Properties for Efficient Demonstrations to a Socially Guided Intrinsically Motivated Learner, in Proceedings of the 21st IEEE International Symposium on Robot and Human Interactive Communication, 2012, Paris, France.

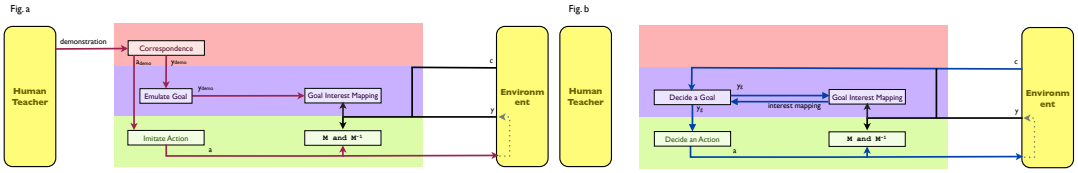

Fig. 1: Data flow of the SGIM-D learner with its environment and teacher. (a): social learning regime. (b): intrinsic motivation regime.

II. SGIM-D Framework

A. Formalisation

In this subsection, we describe the learning problem that we consider. Following Csibra's theory of human action [26], [27], we represent episodes as [context][action][effect] sets.

Let us consider a robotic system which states are described in both a state/context space C, and an effect/task space Y. In contexts $c \in C$, actions $act \in ACT$ output an effect $y \in Y$ (cf fig. 1). In this work, we only consider episodes without planning, and thus do not describe the change in context. For the learning agent, the actions act are parameterised dynamic motor primitives, i.e. temporally extended macro-actions controlled by parameters $a \in A$ (while the actions of the teacher are a priori of unknown structure).

Our agent learns a policy through an inverse model $\mathbf{M^{-1}} : (c, y) \mapsto a$ by building local mappings of $\mathbf{M} : (c, a) \mapsto y$, so that from a context c and for any achievable effect y, the robot can produce y with an action a. The association (c, a, y) corresponds to a learning exemplar which will be memorised. We can also describe the learning in terms of tasks, and consider y as a desired task or *goal* which the system reaches through the *means* a in a given *context* c. In the following, both descriptions will be used interchangeably.

B. SGIM-D Overview

SGIM-D learns by episodes during which it learns either by intrinsically motivated or social learning exploration.

In an episode under intrinsic motivation (fig. 1b), it actively self-generates a goal $y_g \in Y$ where its competence improvement is maximal, then explores which actions a can achieve the goal y_g in context c, following the SAGG-RIAC algorithm [25]. The exploration of the action space gives a local forward model $\mathbf{M} : (c, a) \mapsto y$ and inverse model $\mathbf{M^{-1}} : (c, y) \mapsto a$, that it can use later on to reach other goals. SGIM-D explores preferentially goals easy to reach and where it makes progress the fastest. It tries different actions to approach the self-determined goal, re-using the action repertoire of its past autonomous and imitative explorations. The episode ends after a fixed duration.

When the learner observes a demonstration by the teacher, it starts to learn by social guidance. In an episode under social learning (fig. 1a), our SGIM-D learner observes the demonstration $[c_d, act_d, y_d]$, memorise this effect y_d as a possible goal, and imitates the demonstrated action act_d for a fixed duration.

Its architecture is detailed in the next section.

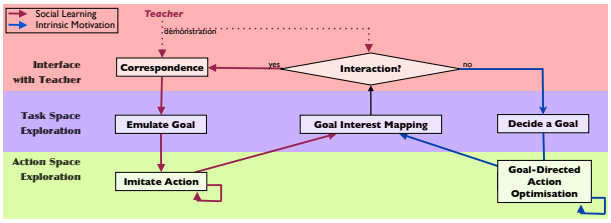

Fig. 2: Time flow chart of SGIM-D, which combines Intrinsic Motivation and Social Learning into 2 layers that pertain the task space exploration and the action space exploration respectively.

Algorithm II.1 SGIM-D

Initialization: $\mathcal{R} \leftarrow$ singleton $C \times Y$, $flagInteraction \leftarrow false$, $Memo \leftarrow$ empty episodic memory
loop
  $flagInteraction \leftarrow$ check if the teacher makes a demonstration
  if $flagInteraction$ **then**
    Social Learning Regime
    repeat
      $(c_d, a_d, y_d) \leftarrow$ Correspondence of the teacher's demonstration
      Emulate Goal: $y_g \leftarrow y_d$
      $Memo \leftarrow$ Imitate Action (a_d, c)
    until End of social interaction
  else
    Intrinsic Motivation Regime
    Measure current configuration c
    $y_g \leftarrow$ Decide a goal
    repeat
      $Memo \leftarrow$ Goal-Directed Action Optimisation(c, y_g)
    until Terminate reaching of y_g
  end if
  $\mathcal{R} \leftarrow$ Update Goal Interest Mapping$(\mathcal{R}, Memo, c, y_g)$
end loop

III. SGIM-D Architecture

SGIM-D (Socially Guided Intrinsic Motivation by Demonstration) is an algorithm that merges interactive learning as social interaction, with the SAGG-RIAC algorithm of intrinsic motivation [25], to learn local inverse and forward models in complex, redundant, high-dimensional and continuous spaces. Its architecture is separated into two layers (fig. 2) :

- The *Task Space Exploration*, a level of active learning which drives the exploration of the task space. With the autonomous learning regime, it sets goals y_g depending on the interest level of previous goals (*Decide a Goal*). With the social learning regime, it retrieves from the teacher information about demonstrated effects y_d (*Emulate a Goal*). Then, it maps $C \times Y$ in terms of interest level (*Goal Interest Mapping*).
- The *Action Space Exploration*, a lower level of learning that explores the action space A to build an action repertoire and local models. With the social learning regime, it imitates the demonstrated actions act_d (*Imitate an Action*), while during self-exploration, the *Goal-Directed Action Optimisation* function attempts to reach the goals y_g set by the *Task Space Exploration* level,

then, it returns the measure of competence at reaching y_d or y_g.

A. Task Space Exploration

1) Goal Interest Mapping: $C \times Y$ is partitioned according to interest levels. For each effect y_g explored in context c, it assigns a competence γ_{c,y_g} which evaluates how close it can reach y. A high value of γ_g (i.e. close to 0) represents a system that is competent at reaching the goal y_g in a context c.

$C \times Y$ is partitioned so as to maximally discriminate areas according to their competence progress, as described in [25]. For a region $R_i \subset C \times Y$, we compute the interest as *the local competence progress, over a sliding time window of the ζ most recent goals attempted inside R_i*:

$$interest_i = \frac{\left| \left(\sum_{j=|R_i|-\zeta}^{|R_i|-\frac{\zeta}{2}} \gamma_j \right) - \left(\sum_{j=|R_i|-\frac{\zeta}{2}}^{|R_i|} \gamma_j \right) \right|}{\zeta} \quad (1)$$

2) Emulate a Goal: This function observes the effect y_d that the teacher demonstrated, and computes its competence using the learner's past action repertoire and model it has built.

3) Decide a Goal: This function uses the interest level mapping to decide which goal is interesting to focus on. It stochastically chooses effects in regions for which its empirical evaluation of learning progress is maximal.

B. Action Space Exploration

1) Imitate an Action: This function tries to imitate the teacher with movement parameters $a_{imitate} = a_d + a_{rand}$ with a random movement parameter variation $|a_{rand}| < \epsilon$. After a short fixed number of times, SGIM-D computes its competence at reaching the goal indicated by the teacher y_d.

2) Goal-Directed Action Optimisation: This function searches for actions a that guide the system toward the goal y_g in the given context c by 1) building local models during exploration that can be re-used for later goals and 2) optimising actions to reach for the current goal. In the experiments below, the exploration mixes local optimisation with the Nelder-Mead simplex algorithm [28] and global random exploration to avoid local minima, in order to build memory-based local direct and inverse models, using locally weighted learning with a gaussian kernel such that presented in [29].

For a detailed description of the architecture, please refer to [1] for more details. In the following section, we apply SGIM-D to an illustration experiment.

IV. Combining Imitation and SAGG-RIAC Improves the Learning Performance

A. Fishing Arm Experiment

We consider a simulated 6 degrees-of-freedom robotic arm holding a fishing rod (fig. 3). It learns how to reach any point on the surface of the water with the hook at the tip of the flexible fishing line.

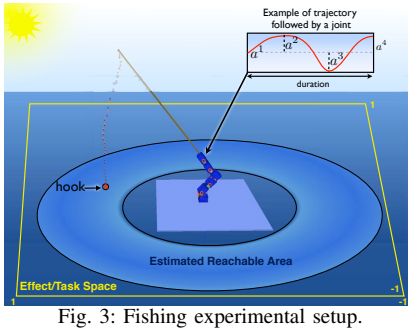

Fig. 3: Fishing experimental setup.

$Y = [-1, 1]^2$ is a 2-D space that describes the position of the hook when it reaches the water. The robot always starts with the same configuration c_{org}, and performs actions that are parametrized motor primitives. For each joint are defined 4 positions: u_1 at $t = 0$, u_2 at $t = \frac{\tau}{3}$,u_3 at $t = \frac{2\tau}{3}$ and u_4 $t = \tau$. The trajectory for each joint is generated by Gaussian distance weighting:

$$u(\mathbf{t}) = \sum_{i=0}^{4} \frac{w_i(\mathbf{t}) u_i}{\sum_{j=0}^{4} w_j(\mathbf{t})} \text{ with } w_i(\mathbf{t}) = e^{\sigma * |\mathbf{t} - \frac{\mathbf{i}\tau}{3}|^2}, \sigma > 0 \quad (2)$$

4 parameters determine the trajectory of each of the 6 joints. Another parameter sets τ. Therefore A is a 25-D space. The robot learns an inverse model in a continuous space, and deals with high-dimensional and highly redundant models. Our setup is all the more interesting since a fishing rod's and wire's dynamics are very difficult to model. Thus learning directly the effect of one's actions is all the more advantageous. This simulation environment is analysed in detail in [1].

B. Experimental Protocol

To assess the efficiency of SGIM-D, we decide to compare the performance of several exploration algorithms (fig. 4):

- Random exploration : throughout the experiment, the robot picks actions randomly in the action space A.
- SAGG-RIAC: the robot explores autonomously, without taking into account any demonstration by the teacher, and is driven by intrinsic motivation .
- Imitation learning: every time the robot sees a new demonstration a_d of the teacher, it repeats the action while making small variations: $a_{imitate} = a_d + a_{rand}$ with $|a_{rand}| < \epsilon$ a small random movement. It keeps on repeating this demonstration until it sees a new demonstration every N actions , and then starts imitating the new demonstration.
- Observation learning: the robot does not make any action, but only watches the teacher's demonstrations.
- SGIM-D: the robot's behaviour is a mixture between Imitation learning and SAGG-RIAC. When the robot sees a new demonstration, it imitates the action, but only for a short while. Then, it resumes its autonomous exploration, until it sees a new demonstration by the teacher. Its autonomous exploration phases take into account all its history from both the autonomous and imitation phases.

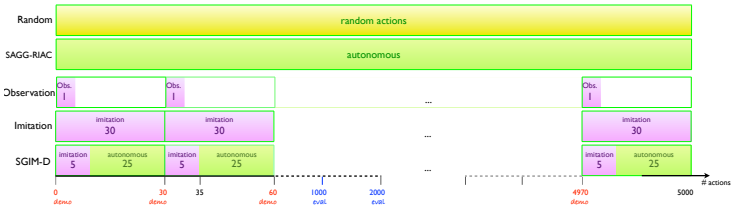

Fig. 4: The experiment compares the performance of several exploration algorithms: Random exploration of the action space A, autonomous exploration SAGG-RIAC, Learning from Observation, Imitation learning and SGIM-D. The comparison is made through the same experimental duration (5000 actions performed by the robot), through the same teaching frequency (every 30 actions) and through regular evaluation (every 1000 actions).

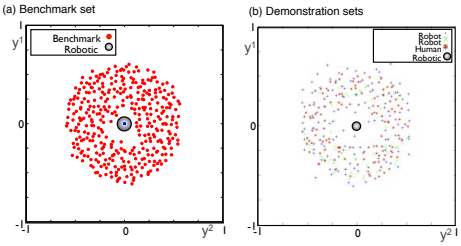

Fig. 5: (a): Map in the 2D task space Y of the benchmark points used to assess the performance of the robot: by measuring how close they can reach each of these points.
(b): Maps in the 2D task space Y of the teaching sets used in SGIM-D, by three demonstrators. Demonstrator 1 is a SAGG-RIAC learner, while demonstrator 2 is an optimised SAGG-RIAC learner, and demonstrator 3 is a human teacher.

For each experiment, we let the robot perform 5000 actions in total, and evaluate its performance every 1000 actions, using the method described below.

C. Evaluation and Demonstration

1) Evaluation: After several runs of Random explorations, SAGG-RIAC and SGIM-D, we determined the apparent reachable space as the set of all the reached points in the effect/task space, which makes up some 300.000 points. We then tile the reachable space into small tiles, and generated a random point in each tile. We thus obtained a set of 358 goal points in the task space, representative of the reachable space, (fig. 5a), to assess the learning precision.

2) Demonstrations: We use 5 demonstration sets (fig. 5b):

- demo 1: the demonstration set is evenly distributed in the reachable space, and taken from a pool of data from several runs of SAGG-RIAC, using the previous SAGG-RIAC learners as teachers. This demo set is chosen randomly among the pool but evenly distributed in the reachable space, as for the evaluation set.
- demo 2: SAGG-RIAC learners who now teach in return our SGIM-D, as for demo 1. But it carefully chooses among their memory exemplars (c, a, y) the most reliable, minimising the variance of y over several re-executions of the same action a in the same context c.
- demo 3: a human teacher gives demonstrations (act_d, y_d) evenly distributed in the reachable space of Y by tele-operating a simulated robot through a physical robot (`http://youtu.be/Ll_S-uO0kD0`). We obtained a teaching set from an expert teacher of 127 samples.

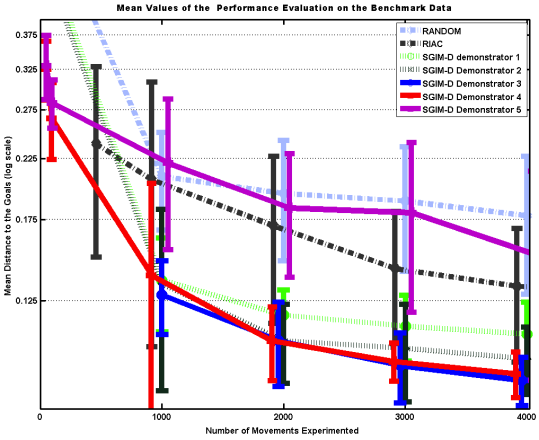

Fig. 6: SGIM-D's performance depends on the demonstrator

- demo 4: in this set, the demonstrator 3 only selects demonstrations where $y_d^1 < 0$ (in the bottom part)
- demo 5: the demonstrator 3 only selects demonstrations where $y_d^1 > 0$ (in the upper part).

As with the evaluation set, we define a tile of the reachable space. The teacher observes the exploration of the learner, and gives a demonstration belonging to a subspace randomly chosen among those it has explored the least.

We showed in [1] that the combination of intrinsic motivation and social guidance improves the learner's performance, compared to learning by imitation or learning by intrinsic motivation only.

Nevertheless, like any social learning method, SGIM-D's performance depends on the quality of the demonstrations. In the next sections, we examine further this dependency,

V. Task Space Exploration

A. Dependence of the Performance on the Teacher

Let us examine how the learning of the same SGIM-D algorithm differs in the case of various teachers. Fig.6 shows that error rates depend on the teachers. The difference between teachers 1, 2 and 3 will be examined in the following section. We here examine the more interesting contrast between demonstrators 3, 4 and 5. All three demonstration sets come from human teacher teleoperation, with demonstrations 4 and 5 being the subsets of demonstrations 3 for $y_d^1 < 0$ and $y_d^1 > 0$ respectively. Nevertheless, the error plot for demonstrator 4 is similar to that of demonstrator 3, whereas the error rate for demonstrator 5 is in between the error plot of a random or a SAGG-RIAC learner. Therefore, the subspace of Y covered by demonstrations is a main factor to the learner's performance.

B. Difference in the Explored Task Spaces

To visualise how the teachers influence the subspaces explored by each learning algorithm, we plot the histogram of the positions y in the effect space Y of the hook when it reaches the water (fig. 7). Each column represents a different algorithm or teacher. We represent for each 2 example experiment runs. The 1st column shows that a natural position lies around $y_c = (0, 0.5)$ in the case of an exploration with random movement parameters. Most

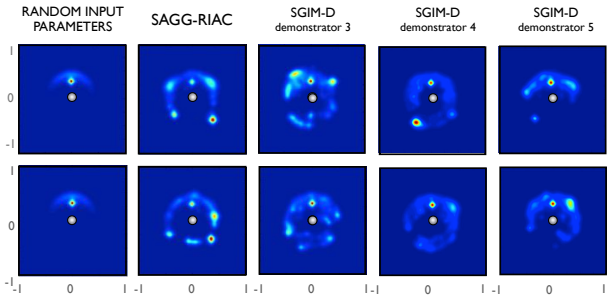

Fig. 7: Histogram of the tasks explored by the fishing rod inside the 2D effects space. Each algorithm is illustrated by 2 example experiments.

movement parameters map to a position of the hook around that central position. This is due to the configuration of the fishing rod which initial state is close to the water surface. Therefore, most random movements would easily drop the hook into the water. On the contrary, to reach positions far from y_c, the robot has to make quite specific movements to lift the rod and make the hook reach farther areas. The second column shows the histogram in the task space of the explored points under SAGG-RIAC algorithm. Compared to random exploration, SAGG-RIAC has increased the explored space, and most of all, covers more uniformly the explorable space. Besides, the exploration changes through time as the system finds new interesting subspaces to focus on and explore. Intrinsically motivated exploration has resulted in a wider repertoire for the robot. SGIM-D (demonstrator 3 and 4) even emphasises this effect: the explored space even increases further, with a broader range of radius covered: the minimum and maximum distances to the centre have respectively decreased and increased. Furthermore, the explored space is more uniformly explored, around multiple centres. The examination of the explored parts of Y show that random exploration only reaches a restricted subspace of Y, while SGIM-D increases this explored space owing to its task space exploration and to demonstrations. However, the case of demonstrator 5 (SGIM-D), demonstrations are given only in subspaces $y_d^1 > 0$ of Y that are often reached by random or SAGG-RIAC exploration. Fig. 7 shows a task space exploration which is broader than the random learner, but still more restricted than the SAGG-RIAC learner. Indeed, this SGIM-D learner only explores around the demonstrated area and neglects other parts of the task space. Demonstrations for easy tasks entail poor performance for the learner, whereas demonstrations for difficult tasks enhance better progress.

Therefore, one of the main bootstrapping factors of SGIM-D is the task space exploration. The teacher influences the exploration of difficult tasks, either by encouraging it with demonstrations of difficult tasks, or by hindering it by focusing attention too much on the easy tasks.

VI. Action Space Exploration

Fig.6 also shows that there are differences in the error plots for the case of teachers 1, 2 and 3, even though their demonstrations cover the same subspace in Y. Let us examine the difference between the teachers 1, 2 and 3.

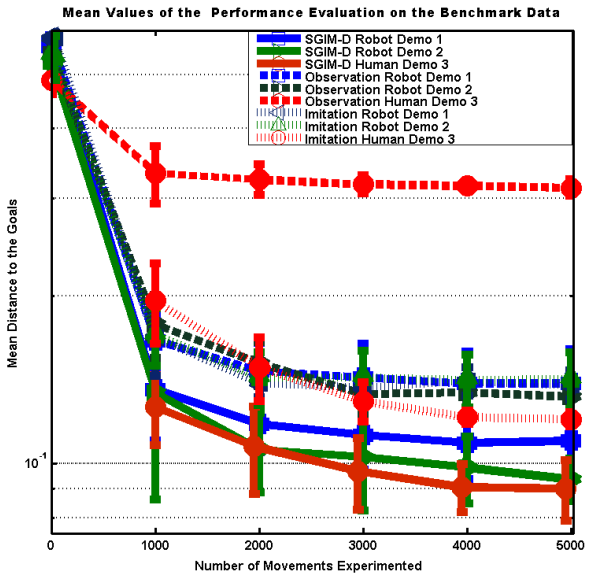

Fig. 8: The performance of the SGIM-D learner depends on the demonstrator.

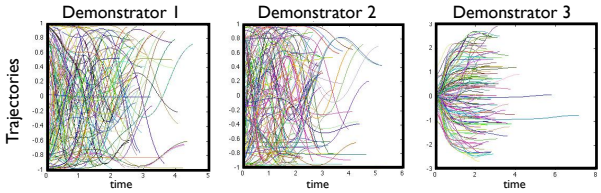

Fig. 9: Plot for the demonstrations of the trajectories for joint 1 (vertical axis: joint angles, horizontal axis: time). Demonstrator 3 gives demonstrations of trajectories that have particular structures.

A. Dependence of SGIM-D Performance on the Quality of Demonstrations

We plot the mean error of the social learning algorithms for our 3 different demonstrators (fig. 8). First of all, we notice that for all 3 teachers, SGIM-D performs better than the other algorithms (t-test with $p < 0.05$ for the error (mean distance to the goals) at t=6000). SGIM-D is therefore robust with respect to the quality of the demonstration as the teacher only guides the learner towards interesting action or effect subspaces, and the learner lessens its dependence on the teacher owing to self-exploration. Still, among the 3 demonstration sets we used, some perform in average better than others. As expected, the demonstrations 1 that are chosen randomly bootstrap less than the demonstrations 2 that have smaller variance (t-test with $p < 0.05$). We also note that the human demonstrations (3), also bootstrap better than demonstrations 1 (t-test with $p < 0.05$). This result seems at first sight surprising, as the results of learning by observation seem to indicate the contrary: demonstrator 1 or 2 are more beneficial to the observation learner (t-test with $p < 0.05$), since demonstrator 3's actions can be not easily reproduced due to correspondence problems.

B. Analysis of the Demonstrated Movements

To understand the reasons of this result, let us examine the different demonstrations. Fig. 9 plots the trajectories of the demonstrations. We can see that demonstrations show

different distribution characteristics in the trajectory profile. The most noticeable difference is the case of demonstrator 3. Whereas the trajectories of demonstrators 1 and 2 seem disorganised, the joint value trajectories of demonstrator 3 are all monotonous, and seem to have the same shape, only scaled to match different final values. Indeed, the comparison of the demonstrations set 3 to random movements with ANOVA [30] indicates that we can reject the hypothesis that demonstration set 3 comes from a random distribution ($p = 4.10^{-40}$). The demonstrations set 3 is not randomly generated but are well structured and regular. Therefore, the human demonstrator shows a bias through his demonstrations to the robot, and orients the exploration towards different subspaces of the action space. Indeed, the ANOVA analysis of the movements parameters a performed during the learning reveals that they have different distributions with separate means. Because his demonstrations have the same shape, they belong to a smaller, denser and more structured subset of trajectories from which is easier for the learner to generalise, and build upon further knowledge. Moreover, this comparative study highlights another advantage of SGIM-D: its robustness to the quality of demonstrated actions. The performance varies depending on the teacher, but still is significantly better than the SAGG-RIAC or imitation learner.

VII. Conclusion

The improvement of SGIM-D over SAGG-RIAC is mainly induced by the teacher's influence on the exploration of the task space. He can hinder the exploration of subspaces difficult to reach by attracting the learner's attention to easy subspaces. On the contrary, he can encourage their exploration by making demonstrations in those subspaces. Therefore, the choice of the task y of the demonstrations (a, y) is crucial. The demonstrations also help the action space exploration. Demonstrations with structured action sets, similar actions shapes, bias the action space exploration to interesting subspaces, that allow the robot to interpolate to reach the most tasks and map to the task space continuously. These conclusions do not only apply to the specific SGIM-D algorithm, but are general to any multi-task learning algorithm who learns by interaction with a teacher. The demonstrator needs both to encourage its goal-oriented exploration to unexplored subspaces of the task space, and to help it generalise by using structured action demonstrations to focus on small and more regular action subspaces. This result underlines the role of social interaction: to bias the exploration by both emulation and mimicking behaviours.

Although SGIM-D is robust to the quality of demonstrations to some extend, this study highlights the importance of the demonstrator. Hence, future work should focus on increasing SGIM-D's robustness, by extending to a learner who can choose to imitate or prefer to learn autonomously, or a learner who can even choose with which teacher to learn. Besides, this present study shows results on only one human teacher. It would be interesting to extend the experiment on several human teachers, especially non-expert robot users.

Acknowledgment

This research was partially funded by ERC Grant EXPLORERS 240007 and ANR MACSi.

References

[1] S. M. Nguyen, A. Baranes, and P.-Y. Oudeyer, "Bootstrapping intrinsically motivated learning with human demonstrations," in *Proceedings of the IEEE International Conference on Development and Learning*, Frankfurt, Germany, 2011.

[2] J. Weng, J. McClelland, A. Pentland, O. Sporns, I. Stockman, M. Sur, and E. Thelen, "Autonomous mental development by robots and animals," *Science*, vol. 291, no. 599-600, 2001.

[3] M. Asada, K. Hosoda, Y. Kuniyoshi, H. Ishiguro, T. Inui, Y. Yoshikawa, M. Ogino, and C. Yoshida, "Cognitive developmental robotics: A survey," *IEEE Trans. Autonomous Mental Development*, vol. 1, no. 1, 2009.

[4] A. Whiten, "Primate culture and social learning," *Cognitive Science*, vol. 24, no. 3, pp. 477–508, 2000.

[5] M. Tomasello and M. Carpenter, "Shared intentionality," *Developmental Science*, vol. 10, no. 1, pp. 121–125, 2007.

[6] A. Billard, S. Calinon, R. Dillmann, and S. Schaal, *Handbook of Robotics*. MIT Press, 2007, no. 59, ch. Robot Programming by Demonstration.

[7] B. D. Argall, S. Chernova, M. Veloso, and B. Browning, "A survey of robot learning from demonstration," *Robotics and Autonomous Systems*, vol. 57, no. 5, pp. 469 – 483, 2009.

[8] E. Deci and R. M. Ryan, *Intrinsic Motivation and self-determination in human behavior*. New York: Plenum Press, 1985.

[9] M. Lopes and P.-Y. Oudeyer, "Active learning and intrinsically motivated exploration in robots: Advances and challenges (guest editorial)," *IEEE Transactions on Autonomous Mental Development*, vol. 2, no. 2, pp. 65–69, 2010.

[10] P.-Y. Oudeyer, F. Kaplan, and V. Hafner, "Intrinsic motivation systems for autonomous mental development," *IEEE Transactions on Evolutionary Computation*, vol. 11, no. 2, pp. 265–286, 2007.

[11] A. G. Barto, S. Singh, and N. Chenatez, "Intrinsically motivated learning of hierarchical collections of skills," in *Proc. 3rd Int. Conf. Development Learn.*, San Diego, CA, 2004, pp. 112–119.

[12] P.-Y. Oudeyer, F. Kaplan, and V. Hafner, "Intrinsic motivation systems for autonomous mental development," *IEEE Transactions on Evolutionary Computation*, vol. 11(2), pp. pp. 265–286, 2007.

[13] J. Schmidhuber, "Curious model-building control systems," in *Proc. Int. Joint Conf. Neural Netw.*, vol. 2, 1991, pp. 1458–1463.

[14] V. Fedorov, *Theory of Optimal Experiment*. New York, NY: Academic Press, Inc., 1972.

[15] D. A. Cohn, Z. Ghahramani, and M. I. Jordan, "Active learning with statistical models," *Journal of Artificial Intelligence Research*, vol. 4, pp. 129–145, 1996.

[16] N. Roy and A. McCallum, "Towards optimal active learning through sampling estimation of error reduction," in *Proc. 18th Int. Conf. Mach. Learn.*, vol. 1, 2001, pp. 143–160.

[17] M. Lopes, F. Melo, L. Montesano, and J. Santos-Victor, *From Motor to Interaction Learning in Robots*. Springer, 2009, ch. Abstraction Levels for Robotic Imitation: Overview and Computational Approaches.

[18] J. Call and M. Carpenter, *Imitation in animals and artifacts*. Cambridge, MA: MIT Press., 2002, ch. Three sources of information in social learning, pp. 211–228.

[19] C. L. Nehaniv and K. Dautenhahn, *Imitation and Social Learning in Robots, Humans and Animals: Behavioural, Social and Communicative Dimensions*. Cambridge: Cambridge Univ. Press, March 2007.

[20] A. L. Thomaz and C. Breazeal, "Experiments in socially guided exploration: Lessons learned in building robots that learn with and without human teachers," *Connection Science*, vol. 20 Special Issue on Social Learning in Embodied Agents, no. 2,3, pp. 91–110, 2008.

[21] J. Peters and S. Schaal, "Reinforcement learning of motor skills with policy gradients," *Neural Networks*, vol. 21, no. 4, pp. 682–697, 2008.

[22] M. Lopes, F. Melo, and L. Montesano, "Active learning for reward estimation in inverse reinforcement learning," in *European Conference on Machine Learning*, 2009.

[23] F. Stulp and S. Schaal, "Hierarchical reinforcement learning with movement primitives," in *Humanoids*. IEEE, 2011, pp. 231–238.

[24] A. L. Thomaz, "Socially guided machine learning," Ph.D. dissertation, MIT, 5 2006.

[25] A. Baranes and P.-Y. Oudeyer, "Active learning of inverse models with intrinsically motivated goal exploration in robots," *Robotics and Autonomous Systems*, in press.
[26] G. Csibra, "Teleological and referential understanding of action in infancy," *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, vol. 358, no. 1431, p. 447, 2003.
[27] G. Csibra and G. Gergely, "Obsessed with goals: Functions and mechanisms of teleological interpretation of actions in humans," *Acta Psychologica*, vol. 124, no. 1, pp. 60 – 78, 2007.
[28] J. C. Lagarias, J. A. Reeds, M. H. Wright, and P. E. Wright, "Convergence properties of the nelder-mead simplex method in low dimensions," *SIAM Journal of Optimization*, vol. 9, no. 1, pp. 112–147, 1998.
[29] C. Atkeson, M. Andrew, and S. Stefan, "Locally weighted learning," *AI Review*, vol. 11, pp. 11–73, April 1997.
[30] W. J. Krzanowski, *Principles of Multivariate Analysis: A User's Perspective*. Oxford University Press, 1988.