

Familiarization to Robot Motion

Anca D. Dragan
Carnegie Mellon University
adragan@cs.cmu.edu

Siddhartha S. Srinivasa
Carnegie Mellon University
siddh@cs.cmu.edu

ABSTRACT

We study the effect of *familiarization* on the *predictability* of robot motion. Predictable motion is motion that matches the observer's expectation. Because of the difficulty robots have in learning motion from user demonstrations, we explore the idea of having *users* learn from *robot* demonstrations — how accurate do users get at predicting how the robot will move? We find that although familiarization significantly increases predictability, its utility depends on how natural the motion is. Overall, familiarization shows great promise, but needs to be combined with other methods that generate appropriate motion with which to be familiarized.

1. INTRODUCTION

We are interested in generating robot motion in the presence of an observer. Imagine a robot like HERB from Fig.1 collaborating with a human on a manipulation task like cleaning up the dining room table. Our goal is to make the motion *predictable*:

DEFINITION 1. *Predictable motion is motion that matches the observer's expectation of how the robot will move [9].*

Predictability is desirable in any interaction, and is especially important for human-robot collaboration [2, 4], where it serves two *goals*: 1) it enables the collaborator to predict how the robot will move (by its very definition), which enables them to plan their own motion accordingly and move simultaneously with the robot; and 2) it increases their *comfort* with the robot (we test this in our experiments).

As the definition implies, to generate predictable motion we must know how an observer expects the robot to move. The traditional paradigm for solving this problem is Learning from Demonstration (LfD), where the observer provides several examples of how they expect the robot to move. The robot learns from these examples using techniques like Gaussian Mixture Models [6], Dynamic Movement Primitives [20], or Inverse Reinforcement Learning [28].

These have shown great success for learning particular skills, or learning general motion in low-dimensional spaces, but suffer in two critical ways: they become computationally intractable in high-dimensional spaces [19], and novice users find it difficult to provide good demonstrations [1].

In this paper, we study the opposite teacher-learner relationship: rather than focusing on the robot learning from the user's demonstrations, we explore the idea of the user

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

HRI'14, March 3–6, 2014, Bielefeld, Germany.
Copyright 2014 ACM 978-1-4503-2658-2/14/03 ...\$15.00.
<http://dx.doi.org/10.1145/2559636.2559674>.

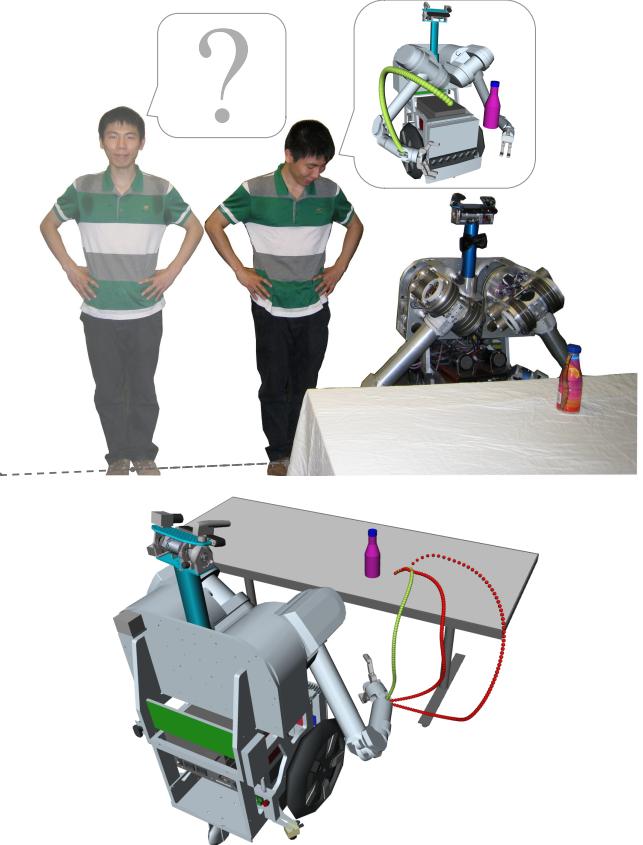


Figure 1: (Top) One of our users getting more comfortable with working/standing next to the robot after familiarization, as he can better predict how the robot will move. (Bottom) Users identify the robot's actual trajectory (we plot here its end effector trace only, in green, but show users the robot actually moving along it) as the one they expect more often after familiarization.

learning from the robot's demonstrations, via *familiarization*:

DEFINITION 2. *Familiarization to robot motion is the process of exposing the observer to how the robot moves in different situations.*

Many times, we take for granted that familiarization works. Familiarization is often used in studies prior to experimental conditions [13], under the assumption that it will adapt the user's mental model of the robot. Studies on sensemaking [27] support this assumption [21, 18, 3, 14, 23], as does the remarkable adaptability of humans: we learn new languages [17], adapt to new ways of communicating [24], and even remap existing sensors like our tongues to new senses, like vision [26].

Here, we study the effects of familiarization to *motion*. On the one hand, the breadth of human adaptability suggests that with familiarization, the robot's motion will become significantly more predictable. On the other hand,

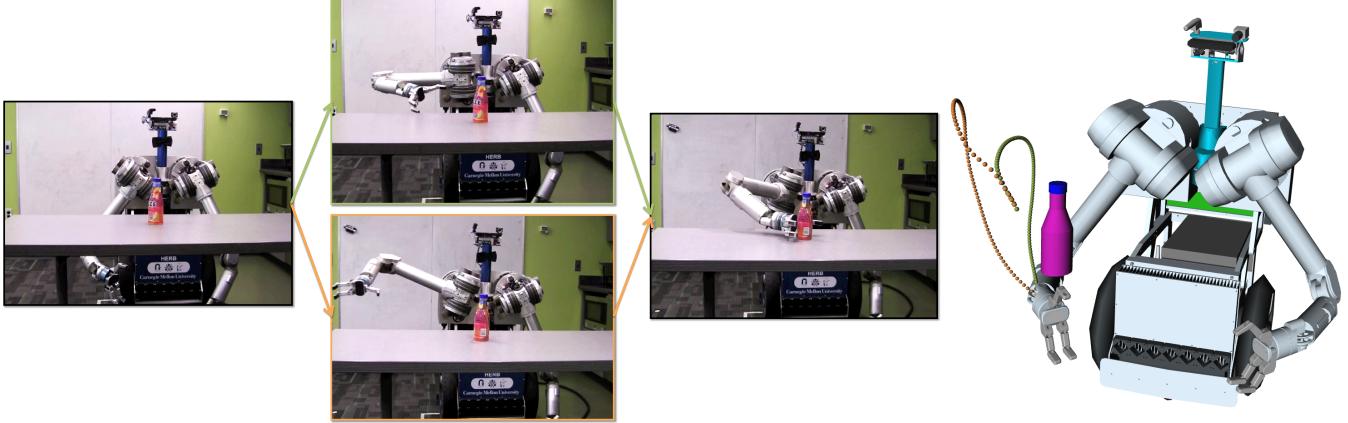


Figure 2: For the same situation, the trajectories for the more natural motion in Sec. 3 (top, green), and for the less natural motion in Sec. 4 (bottom, orange).

the same obstacles robots face when learning motion — the high-dimensionality and complexity of the space — might induce similar limitations in humans.

We ran a series of three experiments investigating the effect of familiarization to two different types of motion, on both the users’ ability to predict how the robot moves (Goal 1), as well as the users’ comfort with the robot (Goal 2).

Our first experiment (Sec. 3) analyzed familiarization to motion produced by a state-of-the-art trajectory optimizer. We evaluated predictability before and after familiarization by testing whether users identify the actual robot motion as the one they expect it to execute (from a set of different motions, see Fig.1, bottom), as well as asking users to rate the motion on a subjective predictability scale.

Our results do support the utility of familiarization — *the motion became significantly more predictable*. However, we came across unexpected limitations of familiarization. We found that despite improving predictability, familiarization can fail to make the motion fully predictable, and can fail to generalize to new situations.

Next, we tested familiarization on a different type of motion. The initial study indicated that the optimizer-generated motion was moderately natural:

DEFINITION 3. *Natural motion is motion that is predictable without (or prior to) familiarization.*

This finding raised an interesting question: would familiarizations still have an effect when the motion is less natural? In a second experiment (Sec. 4), we found that some unnatural motion may *never* reach a high predictability level, even when exposed to over twice the number of motions, suggesting that familiarization saturates.

Next, we tested the practical effects of increasing predictability (Sec. 5) — does the user comfort with working or standing next to the robot also increase? We found a significant effect of familiarization on comfort. However, a lot of users over-trusted the robot, moving closer to it than would be safe. This has a surprising implication: less predictable motion might actually be safer in some situations, as it might prevent over-trust.

Overall, familiarization is an essential aspect of human-robot interactions, and it is important to study it and understand its limitations — sometimes, we cannot rely solely on human adaptability. Our data suggests that familiarization to motion helps, but *cannot be used exclusively for generating predictable motion*. The robot still has the burden of producing motion that is not too unnatural — motion

with which it is easy to familiarize. However, given such motion, familiarization shows great promise for significantly improving predictability and ultimately enabling better human-robot collaboration.

2. GENERATING MOTION

Predictable motion matches what the observer expects. The theory of action interpretation suggests that humans apply the principle of rational action when observing other humans [10], as well as when observing robots [12]: they expect the most efficient action to achieve the goal.

In our previous work, we translated the principle of rational action to robot motion in the form of cost optimization [9]: if humans expect efficient motion, then they expect the robot to optimize some cost function defining efficiency.

Thus, generating predictable motion has two requirements: optimizing cost, and finding the cost to optimize.

Generating Motion by Optimizing Cost. We optimize cost functions over motion by functional gradient optimization [29]. We treat motion trajectories as functions mapping time to robot configurations, $\xi : [0, 1] \rightarrow \mathcal{Q}$. Given a cost functional C defined over the space of trajectories Ξ ,

$$C : \Xi \rightarrow \mathbb{R}^+ \quad (1)$$

we iteratively improve an initial trajectory ξ_0 .

At every iteration, we minimize a first-order approximation of C about the current trajectory, subject to a regularization term which penalizes distances with respect to a metric A on Ξ :

$$\xi_{i+1} = \arg \min_{\xi \in \Xi} C[\xi_i] + \nabla C_{\xi_i}^T (\xi - \xi_i) + \frac{\eta}{2} \|\xi - \xi_i\|_A^2 \quad (2)$$

We iterate until convergence. For state of the art optimizers like CHOMP, this is typically in the order of 1-2 seconds.

The Cost Function. The main challenge in generating predictable motion is finding the cost functional C that the human observer expects. In this paper, we test how well the robot can adapt the observer to its own C , through familiarization.

We test familiarization with a typical cost functional used in CHOMP [29]. This functional is composed of two terms: C_{prior} which measures the efficiency of the motion, and C_{obs} which measures how safe the trajectory is in staying away from obstacles:

$$C[\xi] = C_{prior}[\xi] + \lambda C_{obs} \quad (3)$$

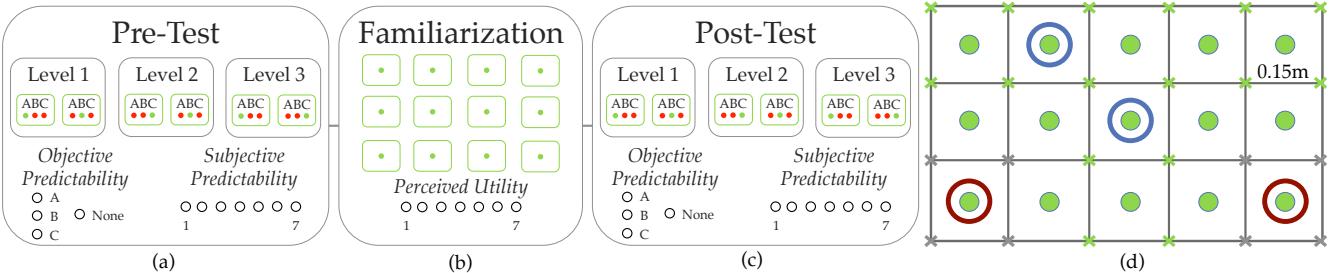


Figure 3: The overall experimental procedure, consisting of a familiarization phase (b), and a pre- and post-test for predictability (a and c). The tests involve three types of examples (**Levels 1–3**), each with two instances to aid robustness. For each example, we show users three trajectories and ask them to identify which one they expect the robot to perform, as well as rate each on a predictability scale. The grid in (d) depicts target object placements on the table (shown in Fig.1 and Fig.2) to produce the familiarization examples. The ones we re-use for testing (**Level 1**) are highlighted in blue, and the ones we set aside for testing-only (**Level 3**) are highlighted in brown. The crosses represent additional example locations we use in Sec. 4.3.

Our implementation of CHOMP uses path length as the measure of efficiency:

$$C_{prior}[\xi] = \int ||\xi(t)'||^2 dt \quad (4)$$

We use CHOMP-generated motions in Sec. 3, when we test how useful familiarization is for state-of-the-art generated motion. Fig.2 (top, green) shows one of the example motions. We find that they are moderately natural, i.e. have good levels of predictability even before familiarization, and that familiarization increases their predictability further. This prompts us to test familiarization for less natural motion in Sec. 4 — would it still work?

A Less Natural Alternative. To test familiarization on less natural motion, we changed the cost function. Rather than using CHOMP’s notion of efficiency from (4), which uses the same weight on each of the robot’s degrees of freedom, we weigh different degrees of freedom differently:

$$C_{prior_W}[\xi] = \int ||\xi'(t)||_W^2 dt = \int \xi'(t)^T W \xi'(t) dt \quad (5)$$

By choosing a W with lower values for the shoulder joints and higher values for the wrist joints, the robot starts penalizing motion in the wrist, and starts moving it less at the expense of moving the shoulder more. This is contrary to what human motion does in reaching tasks [15], which suggests it will also make the robot’s motion less natural. Our results in Sec. 4 support this.

Fig.2 shows a comparison between the original CHOMP cost function and this modified version (bottom, orange).

3. DOES FAMILIARIZATION WORK?

We designed a user study to test the utility of familiarization to robot motion. Does exposing the users to how the robot moves *help* them form the right expectations in the future? And if so, *how good* do users get at predicting the robot’s motion?

3.1 Methods

We exposed users to examples of the robot’s motion (Fig.3, b), and measured improvement in predictability by administering a pre- (Fig.3, a) and post-test (Fig.3, c), using both objective and subjective measures. We detail our procedure below.

3.1.1 Design Decisions

The complexity and high-dimensionality of robot motion are the key obstacle to the utility of familiarization. We

designed our experiment to alleviate this issue: we focus on familiarization by training.

We made familiarization a targeted learning experience, rather than treating it as exposure to the robot “in the wild”. We chose a narrowly scoped task, structured the examples users see by parametrizing the task, and presented users with many examples comprising a good task discretization. **Robot Platform.** We use the robot HERB [22] from Fig.1 as a platform, a bi-manual mobile manipulator with a total of 24 degrees of freedom.

HERB is a good choice for testing familiarization because it is not too anthropomorphic (anthropomorphic robots could inherit the biases of human motion, hindering the effectiveness of familiarization), yet it still has a familiar structure, with a head and two arms. Each arm has 7 DOF, much like human arms: three in the shoulder, one in the elbow, and one in the wrist.

Robot Task. Rather than showing users a snippet of a daily activity, we chose to show them structured examples that better support learning. To do so, we narrowed the scope of our study to a single type of task, and extracted examples by parameterizing the task and discretizing the parameter space.

Of all possible tasks, we focus on *reaching motions*. Reaching for an object (and grasping it) is one of the most common manipulation tasks state-of-the-art robots perform (along with placing): we see it in manufacturing environments [11] as well as in personal [22, 5] and assistive robotics [16].

We designed a typical reaching task, where HERB uses its right arm to reach for a target object on the table (see Fig.2). We parametrize the task by a starting configuration for the arm, a goal configuration where the robot can grasp the target object, and obstacles in the environment which the robot’s motion must not collide with.

We selected these parameters by replicating a scenario in which HERB drives up to the table and reaches for the bottle: we selected HERB’s typical driving configuration as the start, and kept the table in place as the obstacle.

Example Number and Order. We generated examples by varying the goal parameter. We varied the location of the target object on the tabletop, as depicted in Fig.3 (d). To aid familiarization, we discretized this space *finely*, forming a 5×3 grid with 0.15m resolution for where the bottle can be placed. This creates a space of 15 possible examples, 2 of which we kept aside for our pre- and post- test **Level 3** (details in Sec. 3.1.2, “Manipulated Factors”).

We followed human teaching patterns and presented the examples to the users in the order from most simple to most complex [25]. Here, we defined simplicity based on how ef-

ficient each trajectory was relative to the distance between the starting configuration and the goal.

User Instructions. We decided to specifically instruct the users to actively try to learn how the robot moves, in line with our decision of making this a learning task rather than a passive observation task.

3.1.2 Design Overview

Manipulated Factors. We manipulate two factors: **familiarization** and **difference level**.

We manipulate **familiarization** by testing the predictability of motion both **before** and **after** exposing the users to the examples. We use recordings of HERB executing the CHOMP-generated motions.

With **difference level**, we look at test situations that relate in different ways to the examples.

We select two of the 13 possible scenarios the user will see during training and identify these as **Level 1** situations. Next, we select one of the two and change the start configuration or add another obstacle, and identify these as **Level 2** situations. Finally, the user is shown the two scenarios that will not be shown as part of the training set. These scenarios are **Level 3** situations.

Fig.4 (a-c) shows an example situation for each level. Since there is no clear ordering in terms of difficulty between levels 2 and 3, we keep this variable as nominal (as opposed to ordinal) in our analysis below.

We use two situations for each difference level (as opposed to only one) in order to alleviate the risk of introducing confounds in the manipulation. This leads to a total of 6 test situations, which we present to the users in a randomized order both before and after familiarization.

Dependent Measures. We measured the predictability of the robot’s motion in the 6 test situations using both an *objective*, as well as a *subjective* metric.

Objective Predictability. For our objective metric, we measured the accuracy with which users can identify the robot’s actual motion from a set of different motions. This is a way of objectively measuring whether users expect the motion that the robot would execute.

For each test situation, we first presented the users with an image of the robot in the starting configuration, with the bottle placed in the corresponding location. We asked them to spend a minute imagining how they expect the robot to move his arm. To make sure they think the task through, we asked them to describe the motion.

Next, we showed them video recordings of HERB executing three motion trajectories (in randomized order). One of these is the actual trajectory (represented by a green dot in Fig.3 (a) and (c)) produced according to the procedure outlined in Sec. 2.

We selected the other two motions (by varying the goal configuration) such that they are spatially similar either to the actual trajectory from the same situation, or to the actual trajectory from one of the example situations.

We imposed a minimum distance requirement on the test motions: they have to achieve a minimum distance (either at the end effector or at the elbow) from one another. We choose a threshold (of 0.2m) to signify “practical difference”: if the users cannot distinguish among motions that are too similar, this has no practical side effect — at the limit, differences among motions will not even be observable to the naked eye; on the other had, if users mistake the motions for one in which the robot’s arm reaches a different part of the space, this can have severe practical consequences when working next to the robot.

These two motions are represented as red dots in Fig.3 (a) and (c). Fig.1 (bottom) shows the end effector traces for the three candidate trajectories in one of the test situations.

After seeing the three trajectories, we asked the users a multiple-choice question: “Which of the trajectories matched the one you expected?”. The choices were trajectories 1-3, as well as a “None” option (which, despite the strong wording in the question of having “matched” the expected trajectory, was only used in 12% of the cases).

Subjective Predictability. For our subjective metric, we designed a scale for predictability, comprised of three 1-7 Likert scale statements shown in Table I.

TABLE I: PREDICTABILITY SCALE.

Trajectory ‘x’ matched what I expected.

Trajectory ‘x’ is predictable.

I would be surprised if the robot executed Trajectory ‘x’ in this situation.

We asked users, after seeing all three trajectories, to indicate their level of agreement with each statement, for each trajectory (in order to not give away which trajectory HERB would actually execute). In our analysis below, we show that the scale has internal reliability, and combine the ratings for HERB’s actual trajectory (with the third statement reverse-coded) into our subjective metric.

Aside from measuring the motion’s predictability before and after familiarization, we were also interested in two additional measures: whether the users thought that familiarization helped, and what they thought of the robot’s motion.

Perceived Utility. After we showed them the motion examples, users did Likert self reports on utility (whether seeing how HERB moves helps them predict how HERB would move in a new situation), on improvement (whether they are better now at predicting how HERB would move than they were originally), and on confidence (whether they are confident they can predict how HERB would move).

Motion Attributes. We also asked them about the motions that they saw. We were interested in whether the CHOMP motions made sense to them, whether they were more fluid or more machine-like than they originally expected, and whether they would be comfortable working next to the robot if it moved in the way they saw.

Subject Allocation. We opted for a within-subjects design. We explicitly wanted to measure predictability for the same user before and after **familiarization** in order to avoid additional variance. Furthermore, users never get to see what the right answer to the test situations are. This enables us to treat **difference level** as a within-subjects factor as well.

We recruited 25 users (11 female and 14 male, with ages between 19 and 56, $M = 34.68$, $SD = 10.29$, and only 5 reporting having a technical background) via Amazon Mechanical Turk. They performed the study in an average of 50 minutes. To avoid rushed responses, we prevented users from advancing in the task without watching all videos and answering all questions, and we asked control questions at the end to verify attention.

Hypothesis. *Familiarization significantly improves the predictability of motion, as reflected by both the objective accuracy metric, as well as the subjective user ratings.*

3.2 Analysis

Accuracy (objective). Supporting our hypothesis, **familiarization** had a significant effect on the users’ accuracy in recognizing HERB’s actual motion, as indicated by a logistic regression using **familiarization** and **difference level** as factors ($\chi^2(1, 300) = 8.53$, $p = .0035$). There was

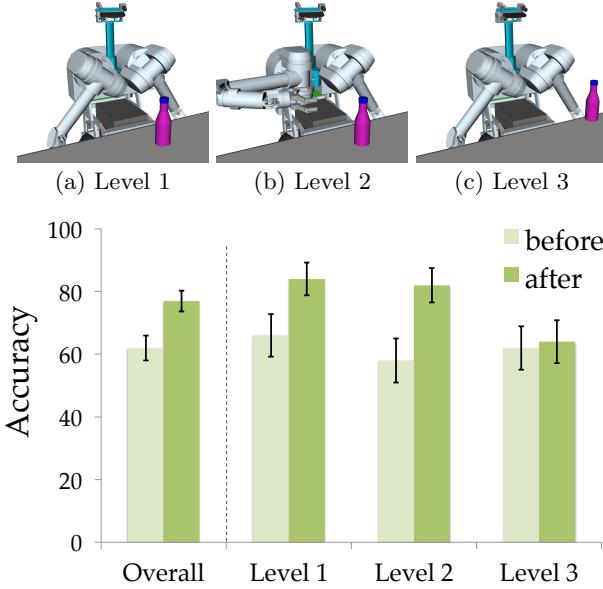


Figure 4: Overall, familiarization significantly improves the accuracy in recognizing the robot’s motion (left). Different test situations, however, show different improvements (right). Error bars show standard error.

no main effect for **difference level**, and no interaction effect. A factorial repeated-measures ANOVA treating accuracy as a 0-1 continuous variable ($F(1, 270) = 10.17, p = .0016$) confirmed the significance of **familiarization**. This test has the advantage of allowing a treatment of the user ID as a random variable, and is considered to be robust to dichotomous data[7]. These results are encouraging, indicating that:

RESULT 1. *Familiarization can make motion more predictable.*

Prior to familiarization, users already had a 62% accuracy (significantly higher than the 33% random choice, Pearson $\chi^2(1, 150) = 55.47, p < .0001$), suggesting that the CHOMP-generated motions were *moderately natural*.

Familiarization did significantly increase accuracy, but, surprisingly, only to 77%.

Although familiarization helps make the CHOMP motion more predictable, our data suggest that it has important limitations: despite the test situations coming from the same task as the training ones, and despite the fine discretization of the task space, users were not able to always identify the correct trajectory from other (spatially different) trajectories. For distance **Level 1**, i.e. testing situations that were also present in the training examples, the accuracy was highest, at 84%.

Fig.4 shows the accuracy improvement after familiarization, both across tasks as well as split by the **difference level**: accuracy is highest on testing on situations that were also used as training examples (**Level 1**), as well as on situations with the same target location as a training example, but different starting/obstacle locations (**Level 2**). The test situations that were at the limits of the task (**Level 3**) did not see an improvement with familiarization (Fig.4, right), suggesting that familiarization can have limited generalization.

RESULT 2. *Familiarization does not always make the motion fully predictable, and can have limited generalization ability.*

Predictability Rating (subjective). Our scale for predictability comprised of ratings for expectedness, predictability, and surprise (reverse-coded) showed internal reliability (Chronbach’s $\alpha = 0.91$), leading to a combined score for predictability based on the three ratings. This score is correlated with the accuracy (Pearson’s $r(288) = .73, p < .0001$).

To test the effects of **familiarization** and **difference level** on this score, we ran a factorial repeated-measures ANOVA. This showed a significant main effect for **familiarization** ($F(1, 270) = 10.17, p = .0016$), but not for **difference level** (and no interaction effect). These results are consistent with our findings for accuracy, and strengthen the utility of familiarization to robot motion.

TABLE II: UTILITY OF FAMILIARIZATION RATINGS.

Utility Type	M	SD	t(24) for $M \neq 4$	p
example helpfulness for prediction	5.76	1.09	8.06	<.0001
improvement in prediction capability	5.6	1.32	6.04	<.0001
confidence in prediction capability	5.76	0.97	9.07	<.0001

Perceived Utility. Table II shows the responses for the perceived utility of familiarization. Participants thought that seeing the videos helps them predict how HERB will move in a new situation, that they are better at predicting how HERB will move in a new situation than they were before seeing the videos, and were confident they can make this prediction accurately. These ratings are significantly different from the neutral stance of 4 (1-7 scale), even after Bonferroni corrections for multiple comparisons.

TABLE III: MOTION RATINGS.

Motion Attribute	M	SD	t(24) for $M \neq 4$	p
makes sense	6.56	0.71	17.98	<.0001
more fluid than expected	5.52	1.66	4.57	<.0001
more machine-like than expected	2.32	1.62	-5.17	<.0001
comfort for collaboration	5.8	1.15	7.79	<.0001

Motion Attributes. Table III shows the responses for the motion attribute questions, together with the results of a *t*-test against the neutral mean of 4. Participants strongly agreed that HERB’s motion made sense. They also agreed that the motions are more fluent than they originally expected, and disagreed that the motions were more machine like. All participants but one reported that they would be comfortable collaborating with HERB on a close-proximity task if it moved in the way they saw. The means are significantly different from the neutral stance, and remain significant after Bonferroni corrections.

These findings, together with the initial accuracy on CHOMP motions, suggest that CHOMP makes a good starting choice for familiarization. The next section will put familiarization to a more difficult test. It will study the effect of familiarization for less natural motions — does it still work, and how predictable do these motions become?

4. FAMILIARIZATION TO UNNATURAL MOTION

Our results showed an improvement with familiarization when the motions are moderately natural. This led us to wonder: what if the robot moved in an unnatural way? Would familiarization still increase predictability?

4.1 Methods

To investigate the effect of familiarization on less natural motion, we ran the same study, replacing the type of motion performed by the robot with the less natural version from Sec. 2, also depicted in Fig.2 (bottom, orange).

For the testing situations, we were interested in whether familiarization would change the users' model and make them select the actual trajectory against the more natural CHOMP one. Thus, we selected the original CHOMP trajectory as one of the alternatives whenever possible, i.e. whenever it was practically different (using our definition from Sec. 3.1.2 of having a difference in the end effector or elbow locations of above 0.2m).

We recruited 25 new users via Amazon Mechanical Turk, and eliminated 1 for failing to answer the control questions correctly, leading to 11 male and 13 female users, with ages between 19 and 45 ($M = 29.16$, $SD = 7.12$).

Hypothesis. *Familiarization significantly improves the predictability of the less natural motion. Furthermore, it brings the less natural motion to the same predictability level as the more natural motion.*

4.2 Analysis

Manipulation Check. The initial accuracy this time was only 34% (close to the random choice mark of 33%¹). This confirms that the motions were less natural (less predictable before familiarization) than the CHOMP motions from the previous section ($\chi^2(1, 588) = 47.38$, $p < .0001$). We call this type of motion *moderately unnatural*: low accuracy without going below the random choice threshold.

Accuracy and Rating Fig.5(left) shows the accuracy before and after familiarization, as compared to the data from the more natural motions in the previous section. Consistent with our previous findings, and with our hypothesis, **familiarization** has a significant positive effect on accuracy, as evidenced by a logistic regression with our two factors ($\chi^2 = 6.95$, $p = .0084$).

Despite this improvement, the accuracy after familiarization is merely 48% — familiarization fails to bring this motion to the same predictability level that it brings the CHOMP motion (i.e. 77% accuracy). This is also supported by the ratings on our predictability scale: although familiarization has a positive main effect on the score ($F(1, 286) = 5.09$, $p = .0248$), the score after familiarization is significantly lower than for the CHOMP motion, as seen in Fig.5(center).

Furthermore, for the test situations where a CHOMP trajectory was one of the options, more users chose the CHOMP trajectory (48%) than the trajectory generated by the cost function with which they were familiarized (43% on these situations).

Given that the initial accuracy on these tests was 29%, familiarization did change the users' model of how the robot moves, *but was not enough to make the true model more likely in their view than the more natural CHOMP model*.

Combined Analysis. The difference between the moderately natural and the moderately unnatural motions is also reflected when looking at the data overall. A logistic regression with **naturalness** (low versus high), **familiarization**, and **difference level** as factors shows significant main effects for all three factors (**naturalness** $\chi^2(1, 588) = 49.71$, $p < .0001$; **familiarization** $\chi^2(1, 588) = 15.46$, $p < .0001$; **difference level** $\chi^2(2, 588) = 14.26$, $p = .0008$). It also shows an interaction effect between difference level and initial predictability ($\chi^2(2, 588) = 10.19$, $p = .0061$).

A factorial repeated-measures ANOVA yielded the same results, and the Tukey HSD post-hoc analysis on the interaction effect revealed that all conditions for the moderately natural motions had significantly better accuracy than all moderately unnatural conditions, with the exception of difference level 2. The tests in this level maintained high accu-

racy, possibly due to a similarity in the motion for the test situations in this difference level).

Overall, we see that lower naturalness of motion results in lower predictability even after familiarization. This suggests a second limit:

RESULT 3. Familiarization can fail to bring less natural motion to the same predictability levels it brings more natural motion.

4.3 Follow-Up: Do we just need more examples?

Upon finding this limitation, we wondered: could we bring predictability levels as high as for the more natural motion by simply increasing the number of examples? Is this limitation caused by the amount of familiarization?

Methods. We tested this in a follow-up study. We created more examples by discretizing the space further, as shown by the grid crosses in Fig.3. After eliminating the ones close to the testing situations from **Level 3** (shown in gray in the figure), we obtained 16 new examples (leading to a possible total of $13+16=29$).

We replicated our previous study, manipulating one additional factor — the **number of examples** — with 3 levels: 13 (previous study), 21, and 29. We added the additional examples after the original ones, maintaining their order and thus avoiding the order of the examples as a confound.

The **number of examples** factor was between-subjects. This was necessary in order to manage the different number of examples in the familiarization stage. We recruited 25 users per level of examples.

Analysis. Fig.5 (right) shows the accuracy before and after familiarization for each case.

RESULT 4. Familiarization can saturate: final predictability does not increase with the number of examples.

Surprisingly, accuracy decreases in the last case, with the largest number of examples. This decrease is significant in a logistic regression over all example levels, which shows a main effect for **number of examples** ($\chi^2(2, 876) = 6.85$, $p = .0325$), and marginally significant in a factorial repeated-measures ANOVA ($F(2, 70) = 2.80$, $p = .0675$).

The accuracy after familiarization with 29 examples is consistently smaller than with 13 or 21, in particular for difference level 1, i.e. tests that appear in the training data.

This could imply that with more examples to learn from, users are more focused on a general model and less able to keep in mind particular cases. Rather than over fitting to the limited number of examples, users might be fitting a more general but less accurate model. There can also be something specific to the examples added that adds confusion. Further investigation is needed in order to understand this drop, and verify it is not produced by chance.

5. FAMILIARIZATION AND COMFORT

In the previous sections, we found that familiarization increases the motion's predictability. Here, we are taking a first step towards analyzing the practical consequences of improved predictability to human-robot collaboration. In particular, does familiarization improve the users' comfort with working next to the robot?

5.1 Methods

We designed an experiment where we evaluated user comfort before and after familiarization, using both an indirect, objective metric, as well as a direct, subjective metric.

¹this is 25% if we take the “none” option into account

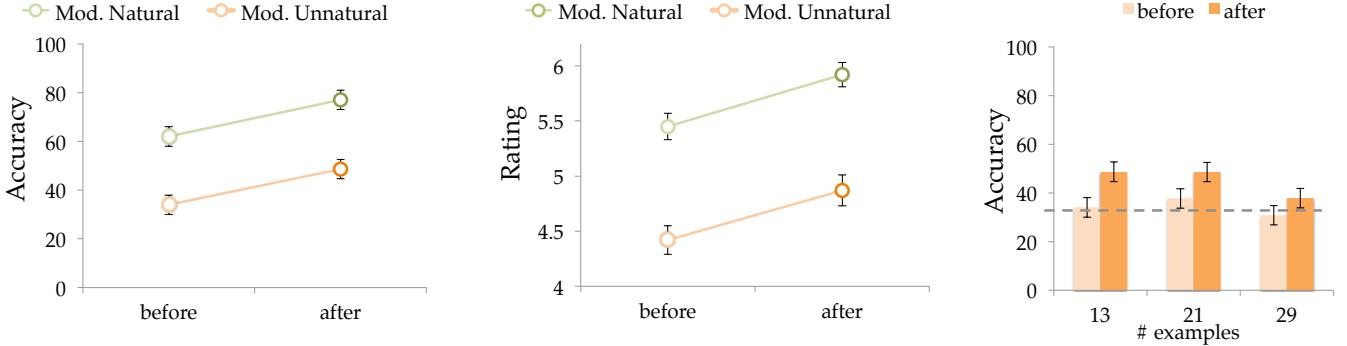


Figure 5: Left and Center: Results for familiarization to a less natural motion, as compared to the more natural CHOMP motion from Fig. 4. The error bars represent standard error on the mean. Familiarization does improve predictability, but not to the level of the CHOMP motions. Right: This limitation is not due to the number of examples, since more examples fail to improve performance.

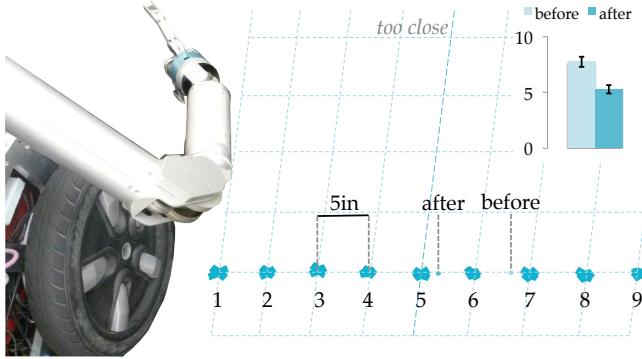


Figure 6: Markers measuring distance to the robot are spaced 5 inches apart. Familiarization brought users 7.35 inches closer to the robot.

Manipulated Factors. We manipulated two factors: **familiarization** and **naturalness** of the motion. We used the same familiarization procedure as before, and the two motions from Sec. 3 and Sec. 4.

We decided against manipulating the **difference level** factor in this study, and only used a **Level 1** situation. We could not use **Level 3**, as familiarization had no effect on the predictability of motion in situations from this level. Furthermore, our pilot for this study (with 6 users) showed no differences between **Level 1** and **Level 2**.

Dependent Measures. We evaluated comfort in two ways: **Objective Comfort.** The robot was set in a **Level 1** situation, in the starting configuration. The experimenter told the users that the robot will move to reach for the target object, and asked them to stand side-by-side with the robot, as close as possible, but far enough away that they felt confident that the arm would not hit them as it moves during the reach (Fig.1).

We marked the floor with 18 marks, starting right next to the robot and moving outward, placed every 5 inches (Fig.6). We measured the distance (marker ID) from the user to the robot.

Although indirect, this metric is of high practical relevance for collaboration: we want users to be comfortable enough to get close to the robot as it is working, in order to be able to do their own tasks simultaneously.

Subjective Comfort. We also directly asked users to indicate (on a Likert scale from 1 to 7), their level of agreement with the statement: “I would feel comfortable working side by side with the robot on a close-proximity task like cleaning

up the dining room table.” (which we augmented with “if it moved in the way I saw” after familiarization).

Subject Allocation. We used a mixed design. We kept **familiarization** within-subjects, measuring improvement in comfort before and after exposure to the robot’s motion. However, **naturalness** was between-subjects, as each user could only familiarize with one type of motion (to avoid confusion and ordering effects).

We recruited 16 users from the local community (9 female and 7 male, with ages between 20 and 64, $M = 36.68$, $SD = 16.6$, with 7 reporting having a technical background).

Hypothesis. *Familiarization significantly improves comfort with working next to the robot, as indicated by both the objective and subjective metrics.*

5.2 Analysis

We were very surprised by how comfortable users were with the robot to begin with: with no prior knowledge of how HERB moves, users stood only 33 inches from the robot’s arm, while the arm could touch them even at 45 inches away. A particularly trusting user stood only 20 inches away, which makes it very difficult for the robot to avoid them even when it knows exactly where they are. Users also rated their comfort with the robot very highly ($M = 6.52$, $SD = 0.61$).

A factorial ANOVA showed a significant main effect for **familiarization** on our objective metric ($F(1, 14) = 12.68$, $p = .0031$): in line with our hypothesis, users were willing to stand closer to the robot after familiarization ($M = 5.28$, $SD = 1.49$) than they were initially ($M = 6.75$, $SD = 1.84$) – a difference of $1.47 \times 5 = 7.35$ inches, indicating that:

RESULT 5. *Familiarization can increase comfort with working side-by-side with the robot.*

We found no effect of **familiarization** on our subjective metric. The mean improved ever-so-slightly ($M = 6.56$, $SD = 0.51$).

Although there was no significant effect for **naturalness**, the means for the objective metric reveal that users did stand slightly further away in the unnatural condition. The means very closely matched the actual safe distances (5.06 for the natural case, and 5.5 for the unnatural case) – users were surprisingly good at estimating the correct spot on which stand, on average.

However, this has an interesting side-effect: familiarization made a lot of users *over-trust* the robot, in that it made them stand *too close* to it (5 out of 8 in the natural condition, and 3 out of 8 in the unnatural condition). Overall, **familiarization** had a marginally significant effect on whether

users over-trusted the robot ($\chi^2(1, 32) = 3.56, p = .0592$), which could have a startling implication:

Less predictable motion might actually be safer in some cases, in that it might prevent over-trusting the robot.

This echoes findings in the trust literature: unreliable behavior increases trust [8]. However, when the robot needs to be conservative about safety (e.g., in the case of industrial arms), this can be a desired effect.

6. DISCUSSION AND LIMITATIONS

In this paper, we explored the effects and limits of familiarizing users to robot motion.

On the one hand, human adaptability would suggest that users would have no issues adapting to any robot motion, as long as it is consistent. We did find that motion becomes more predictable after familiarization, at least when the familiarization process is presented as a learning task.

On the other hand, motion often lies in a complex, high-dimensional space — a possible hindrance for the utility of familiarization. We found that familiarization is not always enough to enable users to identify the robot’s motion (despite choosing among spatially different trajectories), and that less natural motion reaches lower predictability levels. Our data suggests that this limitation can not (at least not always) be overcome by increasing the familiarization length: familiarization can saturate.

Furthermore, our experiments used a pre-test, which could prompt the users’ learning toward test situations, and inflate the effect of familiarization. Predictability after familiarization could be even lower than our measurements indicate.

Our results suggest that users are more comfortable (or more trusting) with the robot after familiarization. However, this could be confounded by the users’ ability to make the translation from the video space to which they familiarize to the physical space in which they are tested.

Of all the factors that could affect the utility of familiarization, our experiments touched upon two: the naturalness of motion, and the number of examples the robot gives the users. Many other factors could impact familiarization: the anthropomorphism of the robot (would users have a harder time with less anthropomorphic robot?), the dimensionality of the space (would they have an easier time with robots with fewer DOF?), the convexity of the cost function the robot optimizes (does non-convexity affect humans as it does machines?), the breadth of examples (one task vs. many), as well as the order or the examples. All of these make very exciting questions for future research.

Although our studies were controlled and focused, they revealed surprising limitations of familiarization. Given that we made optimistic choices for the factors we did not manipulate, aiding familiarizations, we expect to see similar limitations when performing familiarization “in-the-wild”: familiarization improves predictability, but the robot still faces the challenge of producing good motion with which to familiarize.

Acknowledgements

We thank Henny Admoni and the members of the Personal Robotics Lab for their advice. This material is based upon work supported by NSF-IIS-0916557, NSF-EEC-0540865, ONR-YIP 2012, and the Intel PhD Fellowship.

7. REFERENCES

- [1] B. Akgun, M. Cakmak, J. W. Yoo, and A. L. Thomaz. Trajectories and keyframes for kinesthetic teaching: a human-robot interaction perspective. In *HRI*, 2012.
- [2] R. Alami, A. Albu-Schaeffer, A. Bicchi, R. Bischoff, R. Chatila, A. D. Luca, A. D. Santis, G. Giralt, J. Guiochet, G. Hirzinger, F. Ingrand, V. Lippiello, R. Mattone, D. Powell, S. Sen, B. Siciliano, G. Tonietti, and L. Villani. Safe and Dependable Physical Human-Robot Interaction in Anthropic Domains: State of the Art and Challenges. In *Workshop on pHRI*, 2006.
- [3] A. Arita, K. Hiraki, T. Kanda, and H. Ishiguro. Can we talk to robots? ten-month-old infants expected interactive humanoid robots to be talked to by persons. *Cognition*, 95, 2005.
- [4] M. Beetz, F. Stulp, P. Esden-Tempski, A. Fedrizzi, U. Klank, I. Kresse, A. Maldonado, and F. Ruiz. Generality and legibility in mobile manipulation. *Autonomous Robots*, 28:21–44, 2010.
- [5] D. Bertram, J. Kuffner, R. Dillmann, and T. Asfour. An integrated approach to inverse kinematics and path planning for redundant manipulators. In *ICRA*, 2006.
- [6] S. Calinon, F. Guenter, and A. Billard. On learning, representing, and generalizing a task in a humanoid robot. *IEEE Trans. on Systems, Man, and Cybernetics*, 37(2):286–298, 2007.
- [7] R. B. D’AGOSTINO. A second look at analysis of variance on dichotomous data. *Journal of Educational Measurement*, 8(4):327–333, 1971.
- [8] M. Desai, M. Medvedev, M. Vázquez, S. McSheehy, S. Gadea-Omelchenko, C. Bruggeman, A. Steinfeld, and H. Yanco. Effects of changing reliability on trust of robot systems. In *HRI*, 2012.
- [9] A. Dragan, K. Lee, and S. Srinivasa. Legibility and predictability of robot motion. In *HRI*, 2013.
- [10] G. Gergely, Z. Nadasdy, G. Csibra, and S. Biro. Taking the intentional stance at 12 months of age. *Cognition*, 56(2):165 – 193, 1995.
- [11] T. Huang, Z. Li, M. Li, D. G. Chetwynd, and C. M. Gosselin. Conceptual design and dimensional synthesis of a novel 2-dof translational parallel robot for pick-and-place operations. *Journal of Mechanical Design*, 126:449, 2004.
- [12] K. Kamewari, M. Kato, T. Kanda, H. Ishiguro, and K. Hiraki. Six-and-a-half-month-old children positively attribute goals to human action and to humanoid-robot motion. *Cognitive Development*, 20(2):303 – 320, 2005.
- [13] C. D. Kidd and C. Breazeal. Human-robot interaction experiments: Lessons learned. In *Proceeding of AISB*, volume 5, pages 141–142, 2005.
- [14] T. Komatsu and S. Yamada. Adaptation gap hypothesis: How differences between users’ expected and perceived agent functions affect their subjective impression. *Journal of Systemics, Cybernetics and Informatics*, 9(1):67–74, 2011.
- [15] F. Lacquaniti and J. Soechting. Coordination of arm and wrist motion during a reaching task. *J Neurosci.*, 2:399–408, April 1982.
- [16] D. P. Miller. Assistive robotics: an overview. In *Assistive Technology and Artificial Intelligence*, pages 126–136. 1998.
- [17] R. Mitchell and F. Myles. *Second language learning theories*. 2004.
- [18] B. Mutlu and J. Forlizzi. Robots in organizations: the role of workflow, social, and environmental factors in human-robot interaction. In *HRI*, 2008.
- [19] N. Ratliff, J. A. Bagnell, and M. Zinkevich. Maximum margin planning. In *ICML*, 2006.
- [20] S. Schaal. Dynamic movement primitives—a framework for motor control in humans and humanoid robotics. In *Adaptive Motion of Animals and Machines*, pages 261–280. 2006.
- [21] R. M. Siino and P. J. Hinds. Robots, gender & sensemaking: Sex segregation’s impact on workers making sense of a mobile autonomous robot. In *ICRA*, 2005.
- [22] S. Srinivasa, D. Berenson, M. Cakmak, A. Collet, M. Dogar, A. Dragan, R. Knepper, T. Niemueller, K. Strabala, M. V. Weghe, and J. Ziegler. Herb 2.0: Lessons learned from developing a mobile manipulator for the home. *IEEE*, 2012.
- [23] K. Stubbs, D. Wettergreen, and I. Nourbakhsh. Using a robot proxy to create common ground in exploration tasks. In *HRI*, 2008.
- [24] K. Sundin, L. Jansson, and A. Norberg. Communicating with people with stroke and aphasia: understanding through sensation without words. *Journal of clinical nursing*, 9(4):481–488, 2000.
- [25] A. L. Thomaz and M. Cakmak. Learning about objects with human teachers. In *HRI*, 2009.
- [26] S. Upson. tongue vision. *Spectrum, IEEE*, 44(1):44–45, 2007.
- [27] K. E. Weick. *Sensemaking in organizations*, volume 3. 1995.
- [28] B. D. Ziebart, A. Maas, J. A. Bagnell, and A. Dey. Maximum entropy inverse reinforcement learning. In *AAAI*, 2008.
- [29] M. Zucker, N. Ratliff, A. Dragan, M. Pivtoraiko, M. Klingensmith, C. Dellin, J. A. D. Bagnell, and S. Srinivasa. Chomp: Covariant hamiltonian optimization for motion planning. *IJRR*, 2013.