Article

The International Journal of Robotics Research
2018, Vol. 37(11) 1376–1394

DOI: 10.1177/0278364918792295
journals.sagepub.com/home/ijr
SAGE

A flexible optimization-based method for synthesizing intent-expressive robot arm motion

Christopher Bodden, Daniel Rakita, Bilge Mutlu and Michael Gleicher

Abstract

We present an approach to synthesize robot arm trajectories that effectively communicate the robot's intent to a human collaborator while achieving task goals. Our approach uses nonlinear constrained optimization to encode task requirements and desired motion properties. Our implementation allows for a wide range of constraints and objectives. We introduce a novel objective function to optimize robot arm motions for intent-expressiveness that works in a range of scenarios and robot arm types. Our formulation supports experimentation with different theories of how viewers interpret robot motion. Through a series of human-subject experiments on real and simulated robots, we demonstrate that our method leads to improved collaborative performance against other methods, including the current state of the art. These experiments also show how our perception heuristic can affect collaborative outcomes.

Keywords

Human–robot interaction, trajectory optimization, legible robot motion, intent-expressive robot motion

1. Introduction

When robots move in close proximity to human collaborators, the people must be able to interpret the robots' movements. Whether a person is trying to collaborate with the robot on a task, monitor the robot's performance, or simply share the same workspace with the robot, the person's ability to interpret the robot's intent and predict its goals may significantly affect safety, comfort, and collaborative performance. For example, being able to predict a robot's movement can help avoid collisions where the human and the robot reach for the same object, allow for smooth handovers, and enable catching mistakes before it is too late. While robots can be designed to explicitly communicate their goals, for instance, through animated eyes or flashing lights, the robot movements themselves can implicitly communicate intent. *Intent-expressive movements* are movements that are designed to effectively communicate intents.

Humans are exquisitely able to interpret each others' motions and have developed movement strategies for improved communication, such as stage acting and pantomime (Blake and Shiffrar, 2007). Similarly, robotics research has shown that the manner in which a robot arm moves as it performs a manipulation task affects how people interpret and interact with it (Bortot et al., 2013). Arm movements can be designed in specific ways to achieve desirable collaborative effects (for a literature review, see Lichtenthäler and Kirsch, 2016). Some of these design goals can encoded as optimization objectives to generate motions that improve collaborative task performance (Dragan et al., 2013, 2015). Studies also show that existing methods do not generalize to different tasks or robots: what is best for one robot or task can be ineffective in another situation (Bodden et al., 2016; Zhao et al., 2014).

In this paper, we introduce a method for synthesizing intent-expressive manipulation motions that better generalizes to a range of tasks and robot designs than prior approaches do. Our method synthesizes arm trajectories that effectively signal their target location to a human viewer. We formulate motion synthesis as nonlinear constrained optimization. Task requirements are encoded as constraints, and desired movement properties are defined as variational objective functions (i.e., functions over a trajectory). Prior approaches to expressive motion synthesis can be re-created in this formulation. We use the flexibility of this formulation to introduce a novel objective function

Department of Computer Sciences, University of Wisconsin–Madison, Madison, WI, USA

Corresponding author:
Daniel Rakita, Department of Computer Sciences, University of Wisconsin–Madison, 1210 West Dayton Street, Madison, WI 53706, USA.
Email: rakita@cs.wisc.edu

that encodes intent-expressiveness without the drawbacks of prior approaches. A general-purpose implementation provides freedom to explore a variety of constraints and objectives, while allowing us to exploit off-the-shelf solvers and make comparisons between approaches. Unlike prior approaches, our method isolates the robot's kinematics, the task space, and the model of the viewer's perception from each other, affording flexibility to adapt each to a wide range of situations.

The key to our method is a novel objective function that encodes intent-expressiveness. We formulate intent-expressiveness in terms of the viewer's estimate of the robot's goal given a configuration; an intent-expressive motion is one that minimizes the error in the viewer's estimate over all states in the trajectory (Figure 1). This formulation relies on a heuristic that predicts the viewer's estimate and enables experimentation with different heuristics. We demonstrate our approach for multiple robot designs, various potential goal sets, and multiple heuristics. Experiments with human participants demonstrate that our approach leads to improved performance in a collaborative task with a continuous workspace and multiple movement constraints. Online and in-person studies with different robot arms establish the reliability of this effect and its generalizability to arms with different kinematics. Experimenting using different heuristics (encodings of the viewer's estimate) show the flexibility of our approach to adapt to new knowledge.

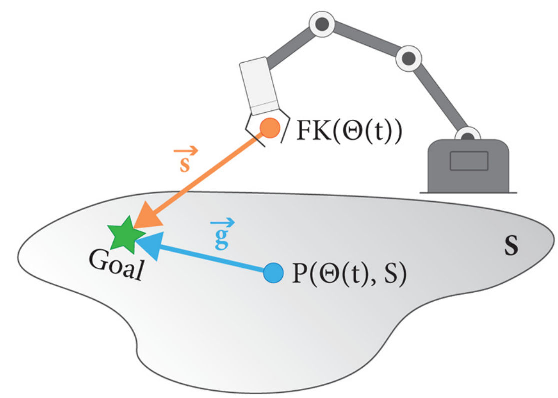

Fig. 1. Our approach encodes intent expressiveness as an objective function that is optimized over the entire trajectory. We assume that, for any state in the trajectory, $\Theta(t)$, the viewer makes a prediction P of the robot's goal within the set S of possible goals. An "intent-expressive" motion is one that minimizes the goal error $\vec{g}$ between P and the actual goal over the course of the motion. The vector, $\vec{s}$, ensures that the trajectory makes progress towards the actual goal.

2. Related Work

The idea that motions inherently communicate intent is rooted in human perception research. Neuroscience research has suggested that the human brain is designed to detect and extract intention from biological motion (Blakemore and Decety, 2001). The work by Blake and Shiffrar (2007) suggests how the visual and motor centers of the brain interpret human actions and gestural cues, and reports on the perceptual implications of such inferences. Also, studies in cognitive science have demonstrated that even infants have the ability to extract intent from motion—whether or not the motion is performed by a human (Gergely et al., 1995) or a humanoid robot (Kamewari et al., 2005). These seminal works provide compelling evidence that the human perception system is hardwired to look for communicative clues in motion itself, and this phenomenon can be leveraged to design motions for more effective collaborations.

Designing communicative motion has a long tradition in character animation (for a history of the development of the modern art, see Thomas and Johnston, 1995). The principles used in animation have been adopted by the robotics community. For example, Takayama et al. (2011) have shown that motions designed using principles of character animation can improve people's perceptions of the robot's intelligence and their confidence in inferring the robot's goal. Other examples of motion design to improve human perceptions of robots have aimed to make social robots more understandable (Van Breemen, 2004), improve perceived safety in mobile robot crossings (Lichtenthaler et al., 2012), and make free-flying robots appear natural and safe (Szafir et al., 2014). The use of communicative motions, where the actions themselves convey information, offers a less intrusive alternative to using explicit signals, which has also been used to improve human–robot collaboration, such as dynamic light displays to show a flying robot's direction (Szafir et al., 2015).

Prior research on improving human perceptions of robot motion has primarily focused on helping the viewer predict a robot's manipulation goal position both quickly and accurately (for a comprehensive survey on legible robot motions, see the work by Lichtenthäler and Kirsch (2016)). Work in this area generally utilizes a technique called *trajectory optimization*, where a robot's motion is optimized to achieve specified motion properties, subject to a set of constraints (Betts, 1998; Ratliff et al., 2009; Schulman et al., 2013). While some work has used explicit communicative signaling, such as pointing gestures (Holladay et al., 2014), to convey intent, our work focuses on implicit intent-expressiveness using just the motion itself. Zhao et al. (2014) considered the effect of standard arm-motion-synthesis methods on communicating intent for manipulation. Dragan et al. (2013, 2015) provide an approach for designing motions to communicate intent, show how these motions can be synthesized using trajectory optimization, and establish their superiority to naïve motions in a collaborative task. Nikolaidis et al. (2016) expanded this approach to consider the user's viewpoint in the optimization. However, Bodden et al. (2016) showed that this approach to legible motion synthesis has limited generalizability to different robot designs, suggesting the need for further research to develop approaches that generalize to a broader range of platforms as well as human–robot collaboration scenarios.

While most prior work in intent-expressive robot motion embeds an *a priori* notion of human motion interpretation within a task model, new work is emerging that attempts to make less prior assumptions. For instance, work by Busch et al. (2017) learns a legibility policy using reinforcement learning that is less dependent on pre-set tasks and still generalizes to numerous users. While our work does not use a data-driven model to inform a policy, our models that encapsulate how users will interpret robot motions are still flexible enough to facilitate many task types. In addition, the models embedded within our objectives and constraints are comprehensible, leading to robot motions and actions that are easily explainable, in contrast to many learned policies. We are interested in comparing the results from such reinforcement learning methods to our intent-expressive synthesis methods in future work.

Lastly, it has been shown that motions and gestures can be interpreted quite differently across cultures (Barrett et al., 2005). We note that our work does not take this key cultural consideration into account, and we highlight this as important future work in the area of designed intent-expressive motions. For instance, models that account for motion and gesture interpretations across cultures would allow robots to seamlessly adapt how they make intentions and actions clear for various users.

3. Synthesizing trajectories

We formulate intent-expressive motion synthesis as a nonlinear constrained optimization problem. We first describe the general framework for optimization-based synthesis, then introduce objective functions to achieve intent-expressive motions, and finally compare this method to prior approaches.

3.1. Spacetime constraints

Trajectory optimization, known in the animation community as *spacetime constraints* (Witkin and Kass, 1988), provides a framework for exploring the properties of communicative motions for robot manipulators. A trajectory is specified as the solution to a variational optimization problem that minimizes some objective function over the trajectory subject to a set of constraints. The constraints allow the requirements of the motion to be defined, while the choice of objective allows the desired properties of the movement to be defined. Objectives typically define movements that are efficient and collision-free (Kalakrishnan et al., 2011; Ratliff et al., 2009), are similar to other motions (Gleicher, 1998), or are minimal in energy (Witkin and Kass, 1988). However, flexibility in the objective allows choosing functions to create other effects, such as the use of this formulation by Bodden et al. (2016) to assess objectives from prior work. The current work explores how this optimization approach can be used to define intent-expressive movements.

A robot trajectory is a function that maps time to robot configurations, $\Theta : \mathbb{R} \rightarrow \mathbb{R}^n$, where n is the number of degrees of freedom of the robot. We denote the configuration at time t as $\Theta(t)$ and forward kinematics, the kinematic function that maps from configuration space to end-effector position, as FK, such that the end-effector position at time t is $FK(\Theta(t))$.

Optimizing the variational problem results in a trajectory:

$$\Theta^* = \operatorname{argmin} g(\Theta) \text{ subject to } c_i(\Theta) \Diamond k_i \qquad (1)$$

for the duration of the motion, t_0 to t_f. The objective function, g, is a function over the entire trajectory that returns a scalar value. Each of the constraints, $c_i(\Theta)$, is either $\Diamond \in \{=, \leq, \geq\}$ to a constant, k_i.

The constraints allow us to define the requirements of the motion. For instance, in the experiments presented in Sections 5, 6, and 7, we consider motions that:

- start at a specific configuration by using an initial pose constraint, $\Theta(t_0) = k_0$;
- end at a given point by using a positional constraint on the end effector at the end of the motion, $FK(\Theta(t_f)) = e_f$;
- keep a grasped object vertical at the end of the motion by specifying an orientation constraint, $\hat{F}K(\Theta(t_f)) \cdot [0, 1, 0] = 1$, where $\hat{F}K(\Theta(t))$ denotes the vertical vector for the end effector orientation.

Note that the end constraint is not the same as solving for the inverse kinematics (IK) at the final frame. The variational problem can choose the final configuration such that it minimizes the overall objective, selecting among the many final poses that meet the end-effector position and orientation constraints. The optimization framework allows us to consider objectives and constraints in both the configuration space of the robot and physical space simultaneously.

A basic objective in trajectory optimization, which we will call *Simple* in the remainder of the paper, is to minimize the length of the trajectory in configuration space, resulting in minimal joint movement:

$$Simple = \int_{t_0}^{t_f} \| \Theta(t)' \|^2 \, dt \qquad (2)$$

This objective yields linear interpolation between fully specified poses. Another possible objective minimizes the total movement of the end effector, leading to straight paths (or as straight as possible subject to constraints), which we will refer to as *Straight*:

$$Straight = \int_{t_0}^{t_f} \| FK(\Theta(t))' \|^2 dt \tag{3}$$

In practice, we add a regularization term to most objectives (such as the *Straight* objective), $\varepsilon \| \Theta(t)' \|^2$. This regularization term avoids discontinuous jumps and erratic motions in joint space by encouraging smooth and short trajectories between waypoints. Our current implementation does not consider robot dynamics, although they can be encoded as constraints (see Witkin and Kass, 1988).

To solve the variational optimization problem, we discretize the trajectory and approximate the objective function with finite differences sampled along time. This discretization leads to a nonlinear programming problem over the variables of the representation of the trajectory that can be modeled using automatic differentiation and solved using commonly available variants of sequential quadratic programming (SQP). Such solutions are described by Witkin and Kass (1988) and Gleicher (1998). Unlike this prior work that used bespoke solvers, we use freely available tools from standard, openly available libraries (the Python `ad`[1] automatic differentiation package and the SLSQP solver from `scipy`[2]). Our implementation halts if the solver is unable to find a feasible solution (i.e., one that satisfies the constraints), but it otherwise finds a minimum of the objective subject to these constraints being satisfied. For the class of spacetime objectives we consider, we cannot use per-frame iterative approaches as our objectives cannot be expressed as low pass filters (Gleicher, 2001).

3.2. Intent-expressive objective

The spacetime framework allows us to define an objective to produce intent-expressive motions. In this paper, we focus on the common class of arm tasks, such as grasping or placing an object, where the intent to be inferred is a target location. An intent-expressive motion is one where the viewer is able to predict this goal early and accurately. Such movements are termed *legible* by Dragan et al. (2013) and goal-predictable by Lichtenthäler and Kirsch (2014).

We introduce a novel objective for synthesizing intent-expressive motions. The key intuition behind our objective is that at any time during the motion, the user's prediction of the goal should be as close as possible to the actual goal. At a particular time t, we estimate the user's prediction using a heuristic (discussed in Section 3.3). Because this prediction is based on the state of the trajectory at this point and the set of possible goals, we denote it as $P(\Theta(t), S) \in \mathbb{R}^3$. This formulation allows us to compute a vector of prediction error $\vec{g}$ between the goal and the estimated prediction, as shown in Figure 1.

Our intent-expressive objective minimizes the magnitude of $\vec{g}$ over the entire trajectory. To promote progress toward the goal, we include an objective term that minimizes the distance between the current end-effector position and the goal, which we denote as $\vec{s}$. While the end constraint described in Section 3.1 enforces that the goal is met at the end of the motion, the goal term, $\vec{s}$, reduces detors and encourages progress over the trajectory.

By splitting the robot's motion to its goal into these two components, we effectively separate the component of the motion that allows legibility (e.g., motion that distinguishes between possible goals) from the component of the motion required to reach the goals. Therefore, we produce intent-expressive, which we will refer to as *Legible* in the remainder of the paper, trajectories to a given goal using our spacetime optimization framework:

$$Legible = \int_{t_0}^{t_f} \alpha \|\vec{g}(\Theta(t))\|^2 + \beta \|\vec{s}(\Theta(t))\|^2 + \varepsilon \| \Theta(t)' \|^2 dt \tag{4}$$

This formulation minimizes two components:

$$\vec{g}(\Theta(t)) = Goal - P(\Theta(t), S) \tag{5}$$

$$\vec{s}(\Theta(t)) = Goal - FK(\Theta(t)) \tag{6}$$

The optimization weights α and β determine the tradeoff between direct, energy-efficient motions and legible ones. All figures and experiments in this work use $\beta = \frac{\alpha}{10}$, which we determined empirically. The third term in Equation (4) is a regularization term (as mentioned in Section 3.1) to ensure continuity in joint angles. We use $\varepsilon = \frac{\alpha}{10}$ for all figures and experiments in this work.

3.3. Predicting the current inference

A key element in our objective is the heuristic P, which estimates the viewer's prediction of the goal for a given robot configuration and set of goals. We assume that there is a set of goals (although it could be an infinite set, for example, if the goal is in continuous space) and that the viewer's prediction will always be a member of this set. Choosing this heuristic allows for encoding knowledge into the objective about how the user perceives motion. The heuristic function abstracts domain-specific parameters away from the objective and is specific to the workspace. We emphasize that the heuristic merely needs to determine a goal position based on a robot configuration and that the designer of a heuristic does not need to consider the entire variational problem (as required in the framework proposed by Dragan and Srinivasa, 2013). This formulation enables the heuristic to be simple and easy to modify, facilitating experimentation. To provide concrete examples of how the heuristic is used to encode the properties of legible motion, we describe three heuristic functions.

1. *Point position:* The predicted goal will be the member of the goal set that is closest to the current position of the end effector based on Euclidean distance. This heuristic works for both discrete goal sets and arbitrary continuous workspaces of goals as long as the closest point to the end effector can be determined. To

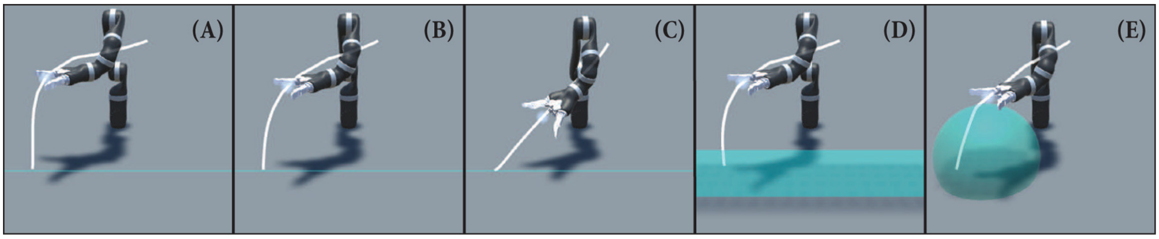

Fig. 2. The heuristic function of our objective allows encoding knowledge of the properties of legible motion. This illustration shows the flexibility of our the heuristic function. (A) Point position, (B) velocity, and (C) pointing show the resulting motions from three different heuristics for the same goal and goal set. (A) Linear, (D) plane, and (E) sphere show the resulting motions from the same heuristic (point position) for three different goal sets.

compute this heuristic, the end-effector position is projected into the set of possible goals. The point-position heuristic leads to motions where the robot quickly moves to configurations for which the prediction is correct and then moves to the goal directly within this subspace. For example, if the potential goal set is a plane (e.g., the surface of a table), the points vertically above the goal all have zero prediction error. Generated motions will move (roughly) parallel to the surface of the table until the end effector is directly above the target and then descend directly toward the target.

2. *Pointing:* The direction of the end effector aligns toward the actual goal under the assumption that the viewer interprets the robot as pointing toward its goal. To compute this heuristic, we determine the ray emanating in the direction of the opening of the end effector and find the point in the goal set that is closest to any point on the ray. The pointing heuristic tends to point the end effector toward the goal and move directly along this direction.
3. *Velocity:* The velocity of the end effector is used to displace the point position under the assumption that the viewer will extrapolate the current end-effector motion into the goal set. Computing this heuristic requires augmenting the state considered to phase space (i.e., configuration and derivatives) and using the position and velocity of the end effector to define a vector that is projected into the set of possible goals.

Figure 2 illustrates the flexibility of our approach with motions generated for different heuristics and for different goal sets. In Section 7, we compare these heuristics in an experiment with human participants. The heuristics are parameterized by the goal set. All heuristics project into the set of possible goals under the assumption that viewers will only choose from among valid goals.

3.4. Comparison to previous work

Prior research includes studies of and alternative approaches to creating legible arm movements. Zhao et al. (2014) considered the expressiveness of existing movement approaches, including "shortest" movement that corresponds to our *Simple* objective (Equation (2)), "straight" movement that corresponds to our *Straight* objective (Equation (3)), and "curved" movement that is an approximation of the technique proposed by Dragan and Srinivasa (2013), which is discussed below.

Dragan and Srinivasa (2013) provided a trajectory optimization approach to creating legible motion and introduce an objective function with similar goals to ours. However, their objective follows a different intuition than ours, suggesting that a motion should be maximally dissimilar from motions to other goals. Figure 3 illustrates a comparison of trajectories produced with the two objectives. As the objective proposed by Dragan and Srinivasa (2013) attempts to maximize distance from non-goals, its implementation must use a trust-region approach to avoid creating excessively long motions. This implementation uses a gradient-ascent solver and expresses all constraints and objectives in joint-angle space. This approach reduces flexibility and does not consider robot kinematics in the optimization.

The trust-region term proposed by Dragan and Srinivasa (2013) can be re-written as a Tikhonov regularization by observing that, at any step, a properly chosen regularization term will have the same derivative as the damping factor (for a derivation, see Gleicher, 1994). This reformulation allows us to re-create the approach proposed by Dragan and Srinivasa (2013) by reformulating its objective function using the methods discussed in Section 3.1:

$$\underset{\xi \in \Xi_{S \to G_r}}{\operatorname{argmin}} \frac{-\int P(G_r|\xi_{S\to Q}) f(t)\, dt}{\int f(t)\, dt} \tag{7}$$

where

$$P\left(G_r|\xi_{S\to Q}\right) = \frac{1}{Z} \frac{e^{-\left(C\left(\xi_{S\to Q}\right) + C^*(Q, G_r)\right)}}{e^{-C^*(S, G_r)}} P(G) \tag{8}$$

Here, G_r is the goal; $\xi_{S\to Q}$ is the trajectory from S to Q; C is the cost of a trajectory; C^* is the optimal cost between two positions; and finally $f(t)$ is a linearly decreasing weight over time.

Having both our proposed objective and that proposed by Dragan and Srinivasa (2013) in the same system allows

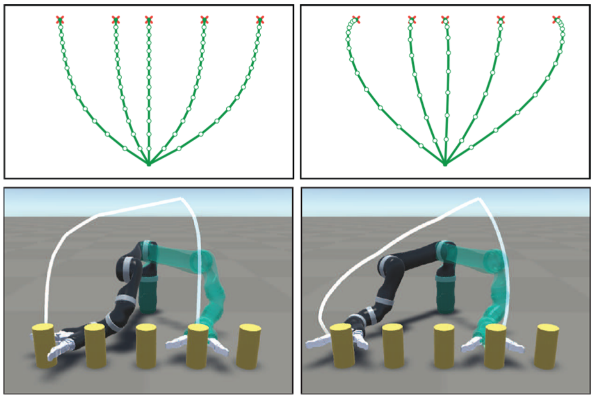

Fig. 3. A comparison for both a point robot (top) and a Kinova Mico robot (bottom) of our legible motion objective (left) and the objective proposed by Dragan and Srinivasa (2013) (right). Note that points are not equally spaced along the line for the point robot.

us to directly compare our approach with the state of the art. Figure 3 shows two simple examples, and Section 6 details an empirical test involving multiple robot designs. In the following, we discuss three ways in which the two approaches differ.

3.4.1. Generality. A primary advantage of the approach proposed here is its generality: it is simpler to adapt and extend. In the approach proposed by Dragan and Srinivasa (2013), the estimate of the user's expectation (termed "predictable") is embedded within the variational problem. While this formulation offers the advantage of a closed-form solution, using the objective with different heuristics requires reasoning about variational objectives. In contrast, our heuristics are simple functions of the robot's current configuration and goal set.

Figure 2 illustrates that our approach can be easily adapted to continuous spaces of goals, including 1D lines, 2D planes, and 3D spheres and can incorporate other heuristics. While such adaptations may be possible in the approach proposed by Dragan and Srinivasa (2013), the realization of these adaptations can be challenging. For example, adapting this approach to continuous goal spaces requires more than just re-writing the summations as integrals, as the objective will become ill-conditioned as many non-goals will be infinitesimally close to any goal. To apply this approach to the continuous goal space required for the experiment described in Section 6, we sampled the goals densely. In this 1D case, this sampling was computationally inefficient; for example, the motions in the bottom of Figure 3 took 2.5 and 19 seconds to solve for our and the state-of-the-art approach, respectively. For larger goal spaces, such sampling may be infeasible.

3.4.2. Movement characteristics. Both Equations (4) and (7) allow for controlling the expressiveness that is necessary to synthesize legible motions using a regularization parameter. However, Equation (7) leads to motions that overshoot the goal, potentially going infinitely far from the goal and stopped only by the trust region or regularization. This overshooting may have both positive and negative consequences. While it may produce more expressive movements, the resulting movements may be misleading. It also has a cost, potentially including additional travel time, that may or may not be worth its benefits. In contrast, our approach provides emphasis without overshooting the goal.

3.4.3. Robot kinematics. Our objective considers the end effector of the robot, while the approach proposed by Dragan and Srinivasa (2013) only uses joint angles. Therefore, our method explicitly considers what the user sees, independent of the robot's kinematics. For a simple point robot (Figure 3, top), task and configuration space are the same, allowing the differences in approaches to be seen independently of the kinematics. However, for arms with more complex kinematics, the space of the formulation may be a factor contributing to the poor performance of Dragan and Srinivasa's method with some robot designs (Bodden et al., 2016), while our method offers consistently superior performance across robot designs, which will be further discussed in Section 6. Extension of their approach to a task-space formulation may be possible future work.

3.5. Solver performance

We tested the performance of our legible-motion method using the various heuristic functions mentioned in Section 3.3 on a testbed of 65 goal locations. The generated trajectories ranged between 3 and 5 seconds in duration, and all were discretized using 15 waypoints. The robot started in a known configuration, and the remaining waypoints were uniformly seeded as the linearly interpolated average between the robot's initial state and a precomputed IK solution corresponding to the goal configuration. Using these initial conditions, only two *Pointing* heuristic motions did not successfully converge, leading to a convergence rate of 99% over all heuristics and goals. For the *Point Position, Pointing*, and *Velocity* heuristics discussed in Section 3.3, our Spacetime Constraints solver executed in 2.4, 109, and 2.7 seconds on average, respectively. All tests were run on an HP Pavilion laptop with an Intel I7-6700HQ 2.6 GHz CPU and 32 GB RAM.

3.6. Trajectory execution

Executing the trajectories synthesized by our method involved sending them to the robot as a series of waypoints in the robot's configuration space. These waypoints were linearly interpolated, and the corners were smoothed (from the intersection $\pm\delta$) using quintic polynomials with a maximum curvature, C. This piecewise function was then executed by the robot controller (for both simulated and physical robots, as discussed in Section 4) as fast as possible within the joint-velocity limits of the robot arm. For the physical robot, we used the controller provided by the `wpi_jaco`[3] Robot Operating System (ROS)[4]

4. Experimental evaluation

To evaluate our legible-motion-synthesis approach, we developed a human–robot collaboration task and conducted three experiments with human participants to evaluate human collaborative performance with robots using different motions generated within our framework. The first experiment (Section 5) compares legible motions synthesized by our approach against simple and straight motion types. We conducted the first experiment online using crowdsourced participants and repeated it in-person in the laboratory with a physical robot. Our second experiment (Section 6) compares legible motions generated using our method and the state-of-the-art method, specifically the objective proposed by Dragan and Srinivasa (2013) as formulated within our approach (as described in Section 3.4). The second experiment was conducted online using crowdsourced participants. Finally, the third experiment (Section 7), conducted online using crowdsourced participants, compares the effects of different heuristics on human collaborative performance.

4.1. Task design

To perform the experiments described above, we devised a human–robot-collaboration task in which participant performance would be directly affected by the ability to interpret the robot's intent. The task required the participant to move a plate underneath where the robot would place a cylinder. The robot randomly chose a position along a linear track (a continuous space), and the participant moved the plate along the track with a mouse to "catch" the cylinder. We described the task scenario to the participants as conducting chemical experiments involving two reactive components (a plate and a cylinder) that needed to be joined for the reaction to occur. The robot's job was to place cylinders at a location along a linear track, and the participant's portion of the task was to place the plate where the cylinder was being placed.

Participants were instructed to always use their best estimate of the robot's target as soon as they were able to make an estimate and not wait until they were sure of the goal. To further encourage this behavior, they were told that the controls were unreliable due to interference in the communication with the robot and might stop working in the middle of the trial, leaving the plate at their last best guess. Some unscored "contact loss" trials were randomly interspersed to remind participants of the need to update their guesses continually as the robot moved toward its goal. This task design allowed us to assess how well the robot communicated its intent by measuring how quickly and accurately the participants placed the plate (before the motion was complete). To ensure that our measurement started from a consistent location, participants were asked to move the plate to the midpoint of the line prior to starting each trial.

To approximate sampling the continuous space of goals in an off-line and reproducible way, we pre-computed robot motions for a dense sampling of goal positions. For each participant, we sampled from this possible set of motions. To ensure consistency of measurement across participants and to avoid potential bias in sampling toward one side of the workspace or toward closer/farther distances, we divided goals into six buckets (three buckets on either side of the center at close, medium, and far distances). Each participant saw randomly sampled motions from each of the six buckets to ensure that the entire workspace was tested for each participant.

To understand whether or not the kinematics of the robot affected the legibility of the motions, we used multiple robot arms with very different kinematics to perform the task. The Kinova Robotics Mico[5] has six degrees of freedom with a series of non-orthogonal "wrists." The Trossen Robotics PhantomX Reactor[6] has only five degrees of freedom and is anthropomorphic in joint configuration. Finally, the Universal Robots UR5[7] has six degrees of freedom with each axis of rotation either orthogonal or parallel to each other.

We used this task in a mix of online and in-person experiments with human participants. In both versions of the experiments, participants controlled the plate using a mouse. In the online version, the participants interacted with one of the simulated robots implemented using the Unity game engine[8] and presented in a web browser. While in-person studies with a real robot may have advantages (Kiesler et al., 2008), simulation has proven to be a valid platform for human–robot interaction studies (Szafir et al., 2014; Woods et al., 2006) and has been used to evaluate the state-of-the-art approach used for comparison (e.g., Dragan and Srinivasa, 2013; Dragan et al., 2013). To confirm the validity of results obtained from online studies and their applicability to real-world human–robot collaborations, our first experiment replicated the online study using a simulated robot as an in-person study with a physical robot. Figure 4 shows the task setups for both the online and the in-person versions of the study.

4.2. Measures

We utilized objective and subjective measures that captured user performance, particularly how quickly and accurately participants inferred the robot's goal from its motion, and perceptions of the robot's motions, including how natural, fluent, and legible they perceived the motions to be. While the main focus of the evaluation was on collaborative performance, the first experiment also measured user perceptions to ensure that any performance gains were not at the expense of user experience, as poor user experience could eventually hurt collaboration and limit acceptance of the robot. These measures are further detailed below.

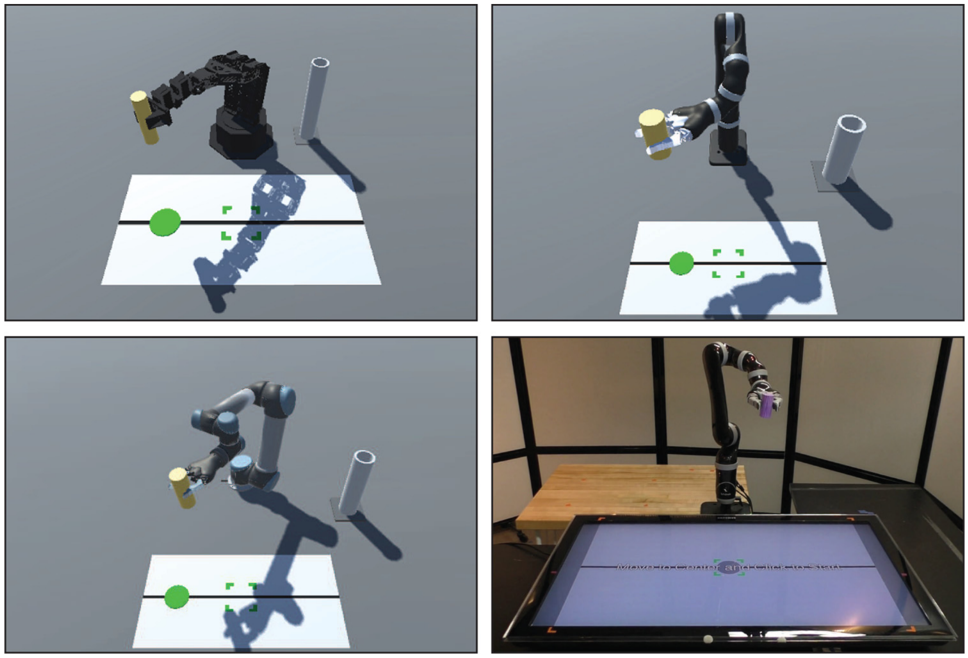

Fig. 4. The collaborative task for the simulated Reactor (top, left), simulated UR5 (bottom, left), simulated Mico (top, right), and actual Mico (bottom, right) robot arms. Participants were asked to place the plate where they thought the robot arm would set the object.

Table 1. Objective measures for collaborative task performance.

Score: The percentage of the trajectory duration remaining when the participant places the plate in the correct position. Here $Score = 1 - (TimeCorrect/Duration)$

Time Correct: The time when the participant places the plate in the correct position ($\pm\varepsilon$) and does not move outside the correct zone again.

Total Error: The integral of the error from the plate to the goal over the duration of the motion.

User Idle Time: The time between the start of the motion and the time the participant begins to move the plate.

4.2.1. Objective measures. To evaluate the objective performance of participants, measures, shown in Table 1, captured how quickly and accurately the participant predicted the target as well as the cost (in total time) of longer motions. *Score, time correct*, and *idle time* were measures of quickness. *Total error* was a measure of accuracy.

4.2.2. Subjective measures. For our first experiment, subjective measures, inspired from previous work (Dragan et al., 2015; Hoffman, 2015), were recorded on a seven-point rating scale for eight questions in four categories shown in Table 2. *Fluency* measured how participants felt about the collaboration with the robot; *legibility* captured how clearly the robot's intent was conveyed; *predictability* measured how well the robot's motions matched participants' expectations; and *naturalness* assessed how human-like participants felt the motions were.

Table 2. Subjective measures and reliabilities for Experiment 1.

Fluency
$\alpha = \{$Online (Mico): 0.76, Online (Reactor): 0.81, In-Person: 0.88$\}$
(1) The robot and I worked fluently together as a team.
(2) The robot contributed to the effectiveness of our team.

Legibility
$\alpha = \{$Online (Mico): 0.87, Online (Reactor): 0.85, In-Person: 0.85$\}$
(3) It was easy to predict where the robot was placing the cylinder.
(4) The robot moved in a manner that made its intention clear.

Predictability
$\alpha = \{$Online (Mico): 0.82, Online (Reactor): 0.88, In-Person: 0.86$\}$
(5) The robot's motion matched what I would have expected if I knew the target before hand.
(6) The robot's motion was not surprising.

Naturalness
$\alpha = \{$Online (Mico): 0.79, Online (Reactor): 0.67, In-Person: 0.81$\}$
(7) The robot's motion looked natural to me.
(8) The robot moved in a human-like way.

4.3. Method of analysis

Our analysis of data obtained from the measures described above revealed that the data was non-normal because it was skewed by trials in which the user was not able to correctly predict the target (zero score). In each of the three experiments, none of the measures for the data we collected met the normality assumptions for analysis of variance (ANOVA) after testing the goodness-of-fit of a normal distribution for each measure using the Shapiro–Wilk W test ($p < 0.001$ for all measures). Therefore, we analyzed our data using the non-parametric Kruskal–Wallis test by ranks for motion type. Pairwise significance tests were performed using Wilcoxon rank-sum tests for each pair of motion types. Because we used non-parametric tests, we also include the median value (denoted as "Mdn") for measures in our results.

5. Comparing legible versus Functional

In our first experiment aimed to show that, in fact, our legible motions were more intent-expressive than standard "functional" motions. We conducted two studies: an online study with simulated Rector and Mico robots and an in-person study with a physical Mico robot. In these studies, we considered three objectives as follows.

1. *Simple* motions (using the objective expressed in Equation (2)), called "shortest" by Zhao et al. (2014) and "predictable" by Dragan and Srinivasa (2013).
2. *Straight* motions (using Equation (3)), termed "straight" by Zhao et al. (2014) and not considered by Dragan et al. (2013, 2015).
3. *Legible* motions using our legibility objective (Equation (4)) with our Point Position heuristic.

We included the *Simple* motions as they are a common baseline and are the most functionally simple method of trajectory execution. *Straight* motions were included, because prior work (e.g., Bortot et al., 2013; Zhao et al., 2014) show them to be superior for collaborative performance, which supports findings from research in neuroscience that humans plan in end-effector space and move their hands in straight lines (Atkeson and Hollerbach, 1985; Morasso, 1981). We hypothesized that our *Legible* motions would be superior to both, as they are designed to improve intent-expressiveness. While Zhao et al. (2014) show legible paths (called "curved") to be inferior to straight paths, this finding may result from an approximated *ad hoc* notion of legibility.

5.1. Hypotheses

We expected *Legible* motion to improve collaborative task performance and motion type to affect subjective measures and, thus, developed the following hypotheses based on *Legible* motion being designed to help participants infer the task and *Straight* motion most closely matching human motion and expectations.

- *H1.1. Task Performance*: Motion type will significantly affect collaborative task performance. *Legible* will result in the best task performance followed by *Straight* then *Simple*.
- *H1.2. Perceptions of Fluency*: Motion type will significantly affect perceived performance. Legible will be rated highest followed by Straight then Simple.
- *H1.3. Perceptions of Legibility*: Legible motions will be perceived as most legible followed by Straight then Simple.
- *H1.4. Perceptions of Predictability*: Straight motions will be perceived as most predictable followed by Legible then Simple.
- *H1.5. Perceptions of Naturalness*: Motion type will affect perceived naturalness. Straight will be rated highest followed by Legible then Simple. An alternative hypothesis is that, because any movement of a robot with three radial wrists is unnatural and unfamiliar, the unnaturalness of the robot will interfere with any effects of the motion and that we will not see significant measurable differences in perceived naturalness.

5.2. Online evaluation

The evaluation of our approach started in the simulated environment, as described in Section 4.1. We conducted two studies with different simulated robots to compare collaborative task performance and participant perceptions of *Legible, Straight*, and *Simple* motions.

5.2.1. Experimental design. Each online study followed a 3×1 between-participants design with *motion type* as the manipulated variable. The first online study was conducted with a simulated Mico robot arm. Each participant observed a total of 24 simulations following one of the motion types (motions toward four random targets that fall within each of the six buckets). After observing all motions, participants filled in a subjective questionnaire. In our second online study, we replicated the first online study with a simulated Reactor robot arm.

5.2.2. Participants. We recruited 54 participants through Amazon Mechanical Turk for each of the two studies. The Mico study involved 18 females and 36 males with an average age of 32.19 ($SD = 9.07$) and a range of 18–61. Participants reported moderate familiarity with robots ($M = 3.20$, $SD = 1.56$) on a seven-point rating scale. Eleven participants reported participating in prior robotics research studies. For the Reactor study, we recruited 21 females and 32 males (one participant did not report gender) aged 20–62 with an average of 33.85 ($SD = 9.65$). Participants reported moderate familiarity with robots ($M = 3.59$, $SD = 1.39$). Nine participants reported participating in prior robotics research studies. Participants were paid $2.00 USD for approximately 20 minutes of participation.

5.3. In-person evaluation

To determine whether our results would translate from simulation to real-world interaction, we conducted an in-person study that replicated our simulation-based studies as closely as possible. We used a Mico robot arm and displayed our simulated plate and track system on an LCD screen laid out on a table. To make recruiting feasible, we modified the study to follow a 3×1 within-participants design. We reduced the number of motions from 24 to 12 per motion type to prevent fatigue in participants. Each participant observed a total of 36 motions (12 motions $\times$ 3 motion types) in a counterbalanced order to account for ordering effects.

5.3.1. Participants. We recruited 24 participants (8 females, 16 males) from the University of Wisconsin–Madison campus. The average participant age was 20.8 ($SD = 1.41$) with a range of 18–24. Participants were moderately familiar with robots ($M = 3.71$, $SD = 1.73$), eight reporting prior participation in robotics research studies. They received $5.00 USD for approximately 30 minutes of time.

5.4. Results

Table 3 provides descriptive statistics and highlights the ordering of the results as follows: green color indicates the top-performing motion type, followed by yellow, which is followed by orange (worst performing). Inferential statistics for the effects of motion type on all measures are shown in Table 4, and Table 5 shows pairwise comparisons for significant effects. At a high level, our analyses showed that, for all measures in all experiments, legible motion was either the best performing motion type or was not significantly worse than the best motion type. For specific results, we refer the reader to Tables 3, 4, and 5 and Figures 5 and 6 and provide a textual summary of the findings below for readability.

The results from our experiments show partial support for H1.1, highlighting *Legible* motion as the top performing motion type. While the score metric provided partial support for the prediction that *Straight* would perform better than *Simple*, results across other metrics and different robots are mixed for this comparison.

Our hypotheses regarding participant perceptions of fluency and legibility (H1.2 and H1.3) were generally supported by our results. *Legible* motion was rated highest for fluency and legibility for all studies with the Mico robot. The results with the Reactor robot were not significant. There was partial support for H1.4; *Legible* motion was rated highest for predictability on the Mico robot, although the difference compared with *Straight* motions were not significant. There were no significant differences in naturalness scores, providing no support for H1.5.

Table 3. Experiment 1: descriptive statistics. Color depicts the ordering of results from dark blue (best performing) to light blue (worst performing).

Online Mico									
Objective Measures	Legible			Straight			Simple		
	M	*SD*	*Mdn*	*M*	*SD*	*Mdn*	*M*	*SD*	*Mdn*
Score	0.61	0.21	0.65	0.48	0.25	0.47	0.23	0.24	0.16
Time Correct	1.34	0.75	1.25	1.69	0.82	1.71	2.41	0.76	2.61
Total Error	0.23	0.14	0.22	0.25	0.13	0.23	0.38	0.22	0.33
Idle Time	0.81	0.47	0.68	0.78	0.41	0.67	1.08	0.50	1.00
Subjective Measures	*M*	*SD*	*Mdn*	*M*	*SD*	*Mdn*	*M*	*SD*	*Mdn*
Fluency	6.06	1.17	6.25	5.83	1.04	6.00	5.08	1.31	5.25
Legibility	6.28	1.11	6.50	5.83	0.99	5.50	4.36	1.38	4.50
Predictability	6.11	1.22	6.50	5.64	1.38	6.00	4.78	1.46	5.00
Naturalness	5.53	1.05	5.50	5.44	1.16	5.50	4.44	1.72	4.75
Online Reactor									
Objective Measures	Legible			Straight			Simple		
	M	*SD*	*Mdn*	*M*	*SD*	*Mdn*	*M*	*SD*	*Mdn*
Score	0.63	0.19	0.66	0.54	0.16	0.54	0.46	0.19	0.45
Time Correct	1.36	0.67	1.25	1.88	0.63	1.88	1.78	0.56	0.81
Total Error	0.22	0.11	0.21	0.30	0.13	0.29	0.28	0.12	0.27
Idle Time	0.61	0.20	0.58	0.86	0.36	0.82	0.92	0.38	0.85
Subjective Measures	*M*	*SD*	*Mdn*	*M*	*SD*	*Mdn*	*M*	*SD*	*Mdn*
Fluency	5.58	1.11	6.00	5.36	0.92	5.50	5.47	1.40	6.00
Legibility	5.92	1.20	6.00	5.44	0.82	5.75	5.64	1.14	6.00
Predictability	5.75	1.32	6.00	5.56	1.37	6.00	5.61	1.23	6.00
Naturalness	4.92	1.02	5.00	4.58	1.50	5.00	4.97	1.47	5.25
In-Person Mico									
Objective Measures	Legible			Straight			Simple		
	M	*SD*	*Mdn*	*M*	*SD*	*Mdn*	*M*	*SD*	*Mdn*
Score	0.66	0.12	0.67	0.43	0.24	0.37	0.37	0.21	0.30
Time Correct	1.36	0.51	1.29	2.14	0.90	2.33	2.27	0.75	2.50
Total Error	0.22	0.10	0.21	0.32	0.16	0.31	0.35	0.17	0.32
Idle Time	0.73	0.22	0.75	0.92	0.45	0.93	0.80	0.31	0.75
Subjective Measures	*M*	*SD*	*Mdn*	*M*	*SD*	*Mdn*	*M*	*SD*	*Mdn*
Fluency	6.48	0.63	7.00	5.85	0.97	6.00	5.50	0.92	5.50
Legibility	6.50	0.69	7.00	5.75	1.11	6.00	4.83	1.19	4.50
Predictability	5.98	0.95	6.25	5.58	1.32	5.75	4.60	1.63	4.25
Naturalness	4.73	1.54	4.50	5.13	1.45	5.00	4.33	1.50	4.00

Table 4. Experiment 1: effects of motion type.

Objective Measures	*H(2)*	*p*	*Subjective Measures*	*H(2)*	*p*
Online Mico					
Score	360.6	0.001	Fluency	7.60	.022
Time Correct	293.5	0.001	Legibility	19.54	< .001
Total Error	137.6	0.001	Predictability	12.13	.002
Idle Time	108.4	0.001	Naturalness	4.44	.11
Online Reactor					
Objective Measures	*H(2)*	*p*	*Subjective Measures*	*H(2)*	*p*
Score	155.6	< .001	Fluency	1.21	.55
Time Correct	142.0	< .001	Legibility	4.54	.10
Total Error	72.1	< .001	Predictability	.50	.78
Idle Time	197.6	< .001	Naturalness	.62	.73
In-Person Mico					
Objective Measures	*H(2)*	*p*	*Subjective Measures*	*H(2)*	*p*
Score	262.1	< .001	Fluency	13.93	< .001
Time Correct	209.5	< .001	Legibility	24.33	< .001
Total Error	97.7	< .001	Predictability	10.14	.006
Idle Time	25.2	< .001	Naturalness	3.78	.15

Table 5. Experiment 1: pairwise comparisons for significant effects of motion type.

Measure	*Straight vs. Simple*	*Straight vs. Legible*	*Simple vs. Legible*
Online Mico			
Score	< .001	< .001	< .001
Time Correct	< .001	< .001	< .001
Total Error	< .001	.023	< .001
Idle Time	< .001	.69	< .001
Fluency	.057	.34	.010
Legibility	.005	.011	< .001
Predictability	.020	.078	.002
Online Reactor			
Measure	*Straight vs. Simple*	*Straight vs. Legible*	*Simple vs. Legible*
Score	< .001	< .001	< .001
Time Correct	.054	< .001	< .001
Total Error	.021	< .001	< .001
Idle Time	.093	< .001	< .001
In-Person Mico			
Measure	*Straight vs. Simple*	*Straight vs. Legible*	*Simple vs. Legible*
Score	< .001	< .001	< .001
Time Correct	.15	< .001	< .001
Total Error	.23	< .001	< .001
Idle Time	< .001	< .001	.20
Fluency	.19	.019	< .001
Legibility	.007	.008	< .001
Predictability	.023	.37	.003

The results from the first experiment provide strong evidence that our legible motion approach helps people better predict the robot's intent and, thus, show better performance in collaborative tasks. Our results also show that our legible motions are perceived as being more capable, legible, and predictable by participants. We found that the results on user perceptions differed across robot platform, highlighting the need for further research to examine the effects of robot kinematics and design on perceptions of legible motion.

6. Comparing legibility methods

The second experiment aimed to compare our method for synthesizing legible motions to the approach proposed by Dragan and Srinivasa (2013), discussed in Section 3.4. To enable this comparison, we extended the experiment performed by Bodden et al. (2016) to include the legible objective presented in Section 3.2. Bodden et al. (2016) found that the approach presented by Dragan and Srinivasa (2013) is not always the most intent-expressive when compared with *Simple* and *Straight* motion types. In particular, different objectives resulted in superior collaborative performance for different robot designs and for different measures of performance. The goal of our approach is to provide a more general method that can achieve superior performance across a broad range of situations.

Our simulated task from Section 4 for each simulated robot was used to conduct online experiments with human participants recruited from the Amazon Mechanical Turk marketplace. In these studies, we considered four objectives.

1. *Simple* motions (using the objective from Equation (2)), called "shortest" by Zhao et al. (2014) and "predictable" by Dragan and Srinivasa (2013).
2. *Straight* motions (using Equation (3)), called "straight" by Zhao et al. (2014) and not considered by Dragan et al. (2013, 2015).
3. *Legible* motions using our legibility objective (Equation (4)) with our Point Position heuristic.
4. *Dragan et al.* (using Equation (7)), approximated as "curved" by Zhao et al. (2014) and called "legible" by Dragan et al. (2013, 2015) and Dragan and Srinivasa (2013).

6.1. Hypotheses

Bodden et al. (2016) hypothesized that both *Straight* and *Dragan et al.* motions would have positive effects on collaborative task performance compared with *Simple* motion, finding mixed results. In our expanded study, we

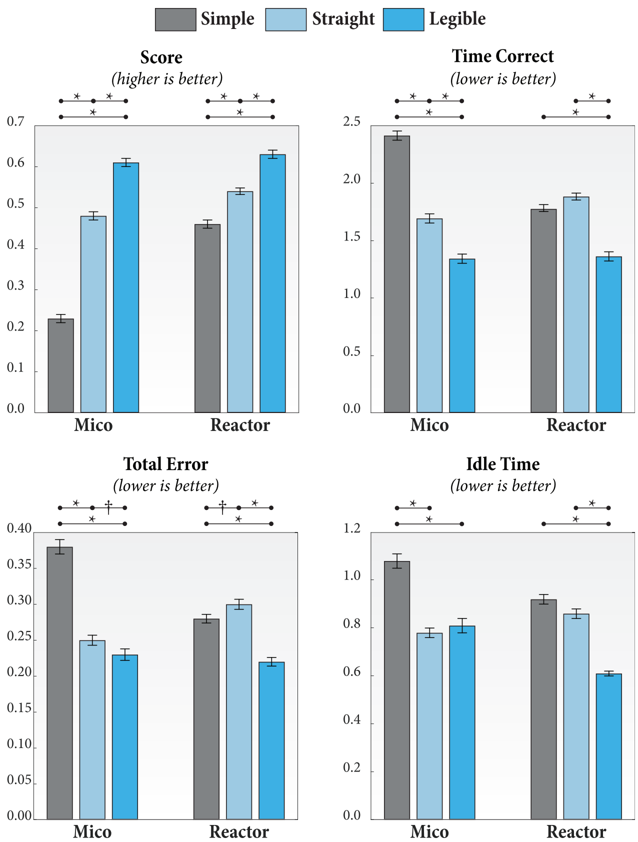

Fig. 5. The results from our first online experiments with both the Reactor and Mico robots, comparing our *legible, straight*, and *simple* motions. Both experiments suggest that our *Legible* motions outperform *Straight* and *Simple* motions. *Error bars indicate standard error.* $*$ *denotes* $p \leq .001$, † *denotes* $p \leq .05$.

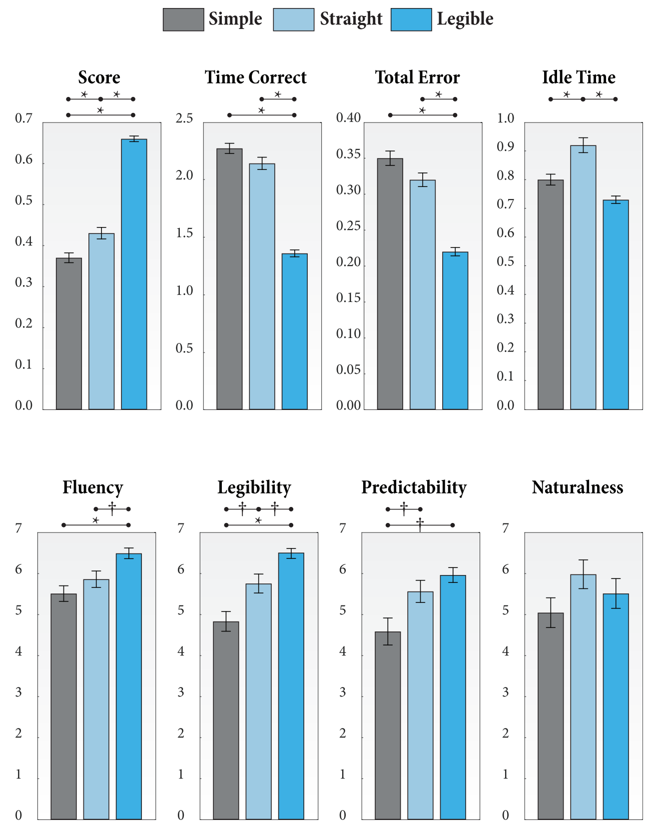

Fig. 6. Results from our in-person experiment to compare our *Legible, Straight*, and *Simple* motions. The results show that our *Legible* motions outperform and are perceived more favorably than *Straight* and *Simple* motions. Error bars indicate standard error. $*$ denotes $p \leq 0.001$, † denotes $p \leq 0.05$.

hypothesized that *Legible* would have positive effects on collaborative task performance compared with all other motion types. For completeness, the hypotheses tested by Bodden et al. (2016) are included (H2.2, H2.3). In addition, we provide analysis to show that the hypotheses from the first experiment (Section 5) hold.

- *H2.1. Legible versus Dragan et al.*: We hypothesize that *legible* motion will have positive effects when compared with *Dragan et al.* We believe that our *Legible* formulation has more intent-expressive power because it separates the intent-expressive and functional components of motion. We predict that our method will perform consistently across robot designs, because it considers the end effector of the robot rather than the joint parameters. We expect our method to perform consistently across performance measures, because it achieves expressiveness without time-consuming exaggeration.
- *H2.2. Dragan et al. versus Simple: Dragan et al.* will positively affect collaborative task performance when compared with Simple motion. This hypothesis based on the results obtained by Dragan et al. (2013, 2015).
- *H2.3. Straight versus Simple: Straight* motion will positively affect collaborative task performance when compared with Simple motion. This hypothesis stems from previous work in neuroscience suggesting Straight motions most closely match human arm motion (Atkeson and Hollerbach, 1985; Morasso, 1981) and is also supported by the findings of our first experiment (Section 5).

6.2. Study design

To test our hypotheses, we conducted three separate 4×1 between-participants experiments (one for each of the three robots: Reactor, Mico, and UR5). Each experiment manipulated motion type (*Legible, Dragan et al., Straight*, and *Simple*). Each participant observed a total of 24 simulated motions (four random targets within each of the six buckets) for one of the three motion types.

6.3. Participants

For each experiment, we recruited 48 participants (12 per motion type) through Amazon Mechanical Turk. For the Reactor, 16 females and 32 males were recruited with an average participant age of 31.9 ($SD = 8.60$, $Max = 67$,

$Min = 20$). Participants reported moderate familiarity with robots ($M = 3.48$, $SD = 1.53$) on a seven-point rating scale. Nineteen participants reported prior robotics-research experience. For the Mico, 24 females and 24 males were recruited with an average participant age of 32.8 ($SD = 8.16$, $Max = 54$, $Min = 18$). Participants reported moderate familiarity with robots ($M = 3.52$, $SD = 1.54$), and 11 participants reported prior participation in robotics research. Finally, for the UR5, 25 females and 23 males were recruited with an average participant age of 35.6 ($SD = 9.01$, $Max = 57$, $Min = 21$). Participants reported moderate familiarity with robots ($M = 3.02$, $SD = 1.73$), and 12 participants reported prior robotics-research experience. Participants were paid \$2.00 USD for approximately 20 minutes of participation.

6.4. Results

Figure 8 summarizes the results of all three experiments. Table 6 shows all descriptive statistics, highlighting the ordering of the results as follows: green color shows the top-performing motion type, followed by yellow, followed by orange, which is followed by red (worst performing). Inferential statistics for the effects of motion type on all measures are shown in Table 7. Table 8 shows pairwise comparisons for significant effects. The results for each experiment are briefly summarized in the paragraphs below.

6.4.1. New results. The data from experiments with all three robots strongly support H2.1. *Legible* motion significantly outperforms *Dragan et al.* across all robots and all measures. In addition, *Legible* motion significantly outperforms both *Simple* and *Straight* motion for all robots and all measures providing more support for the first experiment (H1.1).

6.4.2. Original results. In this section, we briefly present the results for the hypotheses tested by Bodden et al. (2016) (H2.2 and H2.3) for each simulated robot.

The results obtained using the simulated Mico robot support both hypotheses. Both *Dragan et al.* (H2.2) and *Straight* (H2.3) outperform *Simple* in all metrics except idle time. For idle time, only *Straight* (H2.3) outperforms *Simple*. Comparison of *Dragan et al.* and *Straight* showed no significant differences for any measures. Overall, findings with the Mico robot supports findings by Dragan et al. (2013, 2015) that *Dragan et al.* motion outperforms *Simple* motion.

Our results for the Reactor robot tell a different story. While *Straight* and *Dragan et al.* both outperform *Simple* when considering score, *Dragan et al.* significantly outperforms both *Straight* and *Simple* in time correct (absolute time). The difference in time correct is not significant between *Straight* and *Simple*. The performance of *Dragan et al.* and *Simple* are not significantly different for idle time and total error, but both outperform *Straight* motion for idle time. Finally, the difference between *Simple* and *Straight* is not significant for total error. Therefore, H2.2 is only supported by the score and time-correct metrics. H2.3 is supported by the score metric but contradicted by the idle-time metric.

Finally, the results obtained using the simulated UR5 robot tell a third story: *Straight* and *Simple* both outperform *Dragan et al.* when considering score. The difference in score between *Straight* and *Simple* is not significant. *Simple* outperforms both *Straight* and *Dragan et al.* in time correct, and *Dragan et al.* outperforms *Straight* in time correct. The performance of *Dragan et al.* and *Simple* are not significantly different for idle time and total error, but both outperform *Straight* motion for total error. Finally, the difference between *Dragan et al.* and *Straight* is not significant for idle time. Thus, H2.2 is contradicted by the score and time-correct metrics, and H2.3 is contradicted by the time-correct, total-error, and idle-time metrics.

7. Comparing heuristics

Our third and final experiment aimed to show that the choice of heuristic function using our legible objective (Equation (4)) affects the intent-expressiveness of the resulting motion. We utilized our simulated Mico robot to compare each of the three heuristics described in Section 3.3 (*Point Position, Velocity*, and *Pointing*) against a baseline of *Simple* motion synthesized using Equation (2).

7.1. Hypotheses

We hypothesized that each of the heuristics will improve collaborative performance compared to *Simple* motion. We developed the following specific hypotheses based on prior research and our design goals.

- *H3.1. Comparisons across heuristics*: The *Pointing* heuristic will have the highest collaborative task performance of the three heuristics followed by *Velocity* then *Point Position*. Holladay et al. (2014) showed that pointing can be utilized to increase the intent-expressiveness of arm motion. The *Velocity* heuristic is similar to the *Point Position* heuristic except that it encodes the end-effector velocity into the heuristic. Because it provides the participant with more information, we expect *Velocity* to outperform *Point Position*.
- *H3.2. Heuristics versus baseline*: Each heuristic will significantly affect collaborative task performance compared with the baseline Simple motion. The heuristics described in Section 3.3 were each developed as a reasonable assumption of how a participant may pick their current guess. Therefore, when synthesized using our legibility objective, we expect them to increase the intent-expressiveness of the motion compared with Simple motion.

Table 6. Experiment 2: descriptive statistics. Color depicts the ordering of results from dark blue (best performing) to light blue (worst performing).

		Legible			*Dragan et al.*			*Straight*			*Simple*		
Robot	*Objective Measures*	*M*	*SD*	*Mdn*	*M*	*SD*	*Mdn*	*M*	*SD*	*Mdn*	*M*	*SD*	*Mdn*
Reactor	Score	0.66	0.15	0.69	0.53	0.20	0.57	0.55	0.19	0.57	0.47	0.20	0.44
	Time Correct	1.26	0.54	1.15	1.60	0.66	1.42	1.83	0.75	1.73	1.73	0.62	1.79
	Total Error	0.22	0.10	0.21	0.26	0.15	0.24	0.29	0.14	0.28	0.27	0.12	0.26
	Idle Time	0.70	0.33	0.66	0.76	0.35	0.70	0.90	0.37	0.85	0.78	0.35	0.72
	Objective Measures	*M*	*SD*	*Mdn*	*M*	*SD*	*Mdn*	*M*	*SD*	*Mdn*	*M*	*SD*	*Mdn*
Mico	Score	0.64	0.14	0.67	0.41	0.22	0.42	0.44	0.22	0.45	0.28	0.22	0.20
	Time Correct	1.27	0.52	1.11	1.88	0.73	1.83	1.84	0.74	1.79	2.25	0.69	2.49
	Total Error	0.23	0.12	0.21	0.28	0.15	0.27	0.27	0.14	0.25	0.33	0.15	0.28
	Idle Time	0.79	0.39	0.71	0.92	0.52	0.85	0.88	0.43	0.75	0.96	0.50	0.90
	Objective Measures	*M*	*SD*	*Mdn*	*M*	*SD*	*Mdn*	*M*	*SD*	*Mdn*	*M*	*SD*	*Mdn*
UR5	Score	0.59	0.16	0.62	0.35	0.24	0.36	0.42	0.24	0.39	0.40	0.23	0.41
	Time Correct	1.25	0.49	1.15	1.80	0.66	1.80	1.99	0.83	2.13	1.61	0.61	1.58
	Total Error	0.21	0.10	0.20	0.26	0.10	0.25	0.31	0.16	0.26	0.25	0.13	0.24
	Idle Time	0.77	0.40	0.67	0.90	0.39	0.81	1.03	0.56	0.86	0.87	0.40	0.78

7.2. Study design

To test our hypotheses, we conducted an experiment with a 4×1 between-participants design. We manipulated heuristic type (*Point Position, Velocity, Pointing*, and *Simple*). Each participant observed a total of 24 simulated motions (four random targets within each of the six buckets) for one of the four heuristic types.

7.3. Participants

We recruited 48 participants (12 per heuristic type) through Amazon Mechanical Turk. The study included 19 females and 29 males with an average participant age of 31.5 ($SD = 7.99$, $Max = 56$, $Min = 22$). Participants reported moderate familiarity with robots ($M = 3.25$, $SD = 1.67$) on a seven-point rating scale. Twenty-two participants reported prior experience with robotics research. Each participant was paid $2.00 USD for approximately 20 minutes of participation.

7.4. Results

Figure 8 summarizes the results of the experiment. Descriptive statistics are shown in Table 9, which highlights the ordering of the results as follows: green color shows the top-performing heuristic type, followed by yellow, followed by orange, which is followed by red (worst performing). Inferential statistics for the effects of heuristic type on all measures are shown in Table 10, as well as pairwise comparisons for significant effects. The results of the experiment are briefly summarized below.

Table 7. Experiment 2: effects of motion type.

Robot	*Measure*	*H(3)*	*p*
Reactor	Score	123.7	<.001
	Time Correct	109.0	<.001
	Total Error	40.2	<.001
	Idle Time	63.5	<.001
Mico	Score	258.4	<.001
	Time Correct	194.7	<.001
	Total Error	55.0	<.001
	Idle Time	21.4	<.001
UR5	Score	143.0	<.001
	Time Correct	130.7	<.001
	Total Error	52.2	<.001
	Idle Time	49.3	<.001

The results from our experiments contradict our hypothesis comparing heuristics (H3.1). For all measures, except total error when compared with *Velocity*, the *Point Position* heuristic significantly outperforms both the *Velocity* and *Pointing* heuristics. In addition, the *Velocity* heuristic significantly outperforms *Pointing* for all measures. These results show the reverse of the pattern predicted by H3.2. On the other hand, all of the motions using heuristics with our legible-motion objective significantly outperformed the baseline *Simple* motion across all measures, providing strong support for H3.2.

Table 8. Experiment 2: pairwise comparisons for significant effects of motion type.

Robot	*Measures*	*Legible vs. Dragan et al.*	*Legible vs. Straight*	*Legible vs. Simple*	*Straight vs. Simple*	*Straight vs. Dragan et al.*	*Simple vs. Dragan et al.*
Reactor	Score	< .001	< .001	< .001	< .001	0.59	< .001
	Time Correct	< .001	< .001	< .001	.29	< .001	< .001
	Total Error	< .001	< .001	< .001	.08	.006	.24
	Idle Time	.02	< .001	.002	< .001	< .001	.50
Mico	Score	< .001	< .001	< .001	< .001	.21	< .001
	Time Correct	< .001	< .001	< .001	< .001	.57	< .001
	Total Error	< .001	< .001	< .001	< .001	.54	.003
	Idle Time	< .001	.03	< .001	.01	.19	.31
UR5	Score	< .001	< .001	< .001	.29	< .001	.02
	Time Correct	< .001	< .001	< .001	< .001	.01	.002
	Total Error	< .001	< .001	< .001	< .001	.005	.51
	Idle Time	< .001	< .001	< .001	0.006	.05	.36

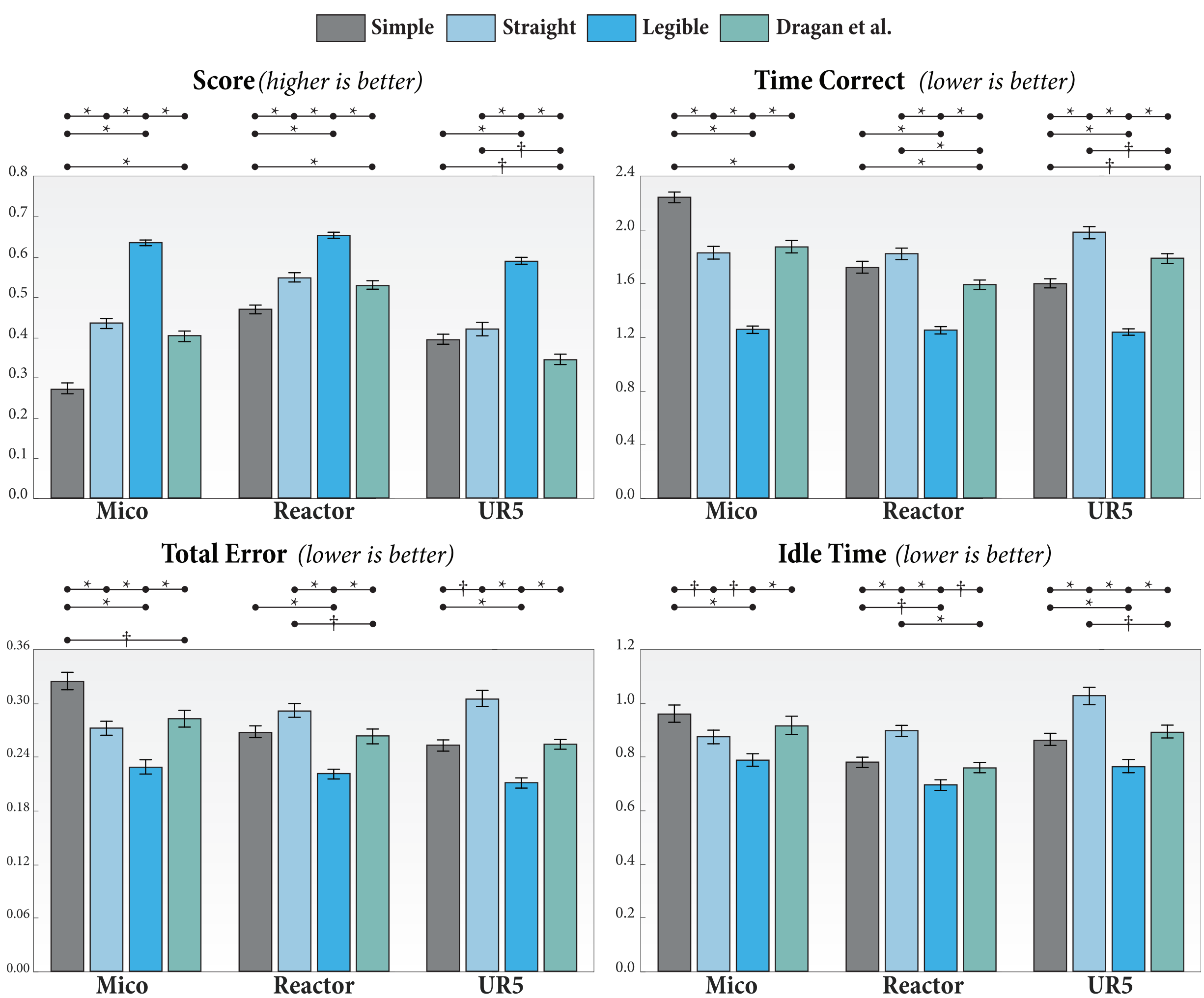

Fig. 7. Results from the second experiment for all three simulated robots comparing *Legible, Dragan et al., Straight*, and *Simple* motions. Error bars indicate standard error. $*$ denotes $p \leq 0.001$, † denotes $p \leq 0.05$.

While the results were unexpected, they can inform the selection of heuristics in designing motion and further examination of the relationship between heuristics and intent expressiveness. One potential explanation of our unexpected findings is that simpler heuristics are more effective for conveying intent, because the resulting

Table 9. Experiment 3: descriptive statistics. Color depicts the ordering of results from dark blue (best performing) to light blue (worst performing).

Objective Measures	*Point*			*Velocity*			*Pointing*			*Simple*		
	M	*SD*	*Mdn*	*M*	*SD*	*Mdn*	*M*	*SD*	*Mdn*	*M*	*SD*	*Mdn*
Score	0.65	0.16	0.68	0.59	0.23	0.62	0.46	0.26	0.42	0.29	0.23	0.22
Time Correct	1.21	0.59	1.08	1.39	0.79	1.29	1.79	0.89	1.92	2.22	0.71	2.43
Total Error	0.22	0.14	0.18	0.25	0.16	0.22	0.28	0.16	0.26	0.33	0.14	0.30
Idle Time	0.68	0.32	0.60	0.81	0.46	0.70	0.95	0.53	0.82	1.09	0.59	0.91

Table 10. Experiment 3: effects of motion type and pairwise comparisons.

Measures	*H(3)*	*p*	Level Difference (*p*-values)					
			Point vs Velocity	*Point vs Pointing*	*Point vs Simple*	*Velocity vs Pointing*	*Velocity vs Simple*	*Pointing vs Simple*
Score	277.2	< .001	< .001	< .001	< .001	< .001	< .001	< .001
Time Correct	223.8	< .001	.012	< .001	< .001	< .001	< .001	< .001
Total Error	88.0	< .001	.059	< .001	< .001	.006	< .001	< .001
Idle Time	92.1	< .001	.02	< .001	.002	.002	< .001	.010

behavior is less complex. For example, the motions resulting from the *Point Position* heuristic tend to move over the goal and then toward it. This behavior may reduce the cognitive burden on the observer, because he/she can conclude that once the robot starts moving toward the goal set, the goal is below the robot. This explanation also implies that the best heuristic can be dependent on the specific task. More exploration is needed to determine which heuristics work best and when.

8. Discussion and future work

In this work, we have presented a flexible method for generating intent-expressive robot arm motions that accommodates different ways a user interprets motion as well as knowledge of a goal set to optimize a trajectory. Our experiments show the effectiveness of our method in quickly and accurately conveying a goal position to a collaborator.

Our first set of studies focused on comparing our new method with naïve *Simple* and *Straight* motion paths. We ran three studies, two online and one in-person, to assess results across a wide population (132 participants) with two robot platforms. Our results confirm that our method leads to improved task performance over the simpler baselines. We assessed task performance using four objective measures, and all of these measures in each of the three studies, except for idle time in the Mico online study, showed a significant advantage over the other motion types. We note that idle time was shown to be significantly better in the Mico in-person study. The main takeaway from these objective results is the effectiveness of our intent-expressive motions over naïve motion types, even across different robots. Prior work by Bodden et al. (2016) has shown that state-of-the-art legibility methods have not displayed benefits across different robot arm designs, so our results here indicate our method's improved ability to generalize across various platforms. We speculate that this is due to our objective function considering the end-effector path in operational space, which can generally be considered independent of the robot's kinematics, rather than just reasoning in the robot's joint-space.

We also show significant subjective measure results from the first set of studies. We measured user perceptions along four subjective measures, *Fluency, Legibility, Predictability*, and *Naturalness*. Users found robots exhibiting our intent-expressive motions to be more fluent and legible compared with the naïve motion types. This effect indicates that users not only noticed that robot motions were easier to predict in the task, but they perceived a positive sense of teamwork and collaboration as a result. These results suggest that the task benefits exhibited by our intent-expressive motions do not come at the expense of user experience. While users did not find intent-expressive motions to be more natural, this was to be expected as prior work shows that people naturally move their hands in straight lines for manipulations (Atkeson and Hollerbach, 1985; Morasso, 1981). However, even though users found the motion to be less natural, they still found the intent-expressive motion to be easier and more enjoyable to work with. We note that this effect could be because the robot arms used did not have many anthropomorphic qualities, and it is possible that the more anthropomorphic a robot arm is, the more users expect and favor natural motion.

Our second set of studies also assesses our work against the state-of-the-art method by Dragan and Srinivasa (2013).

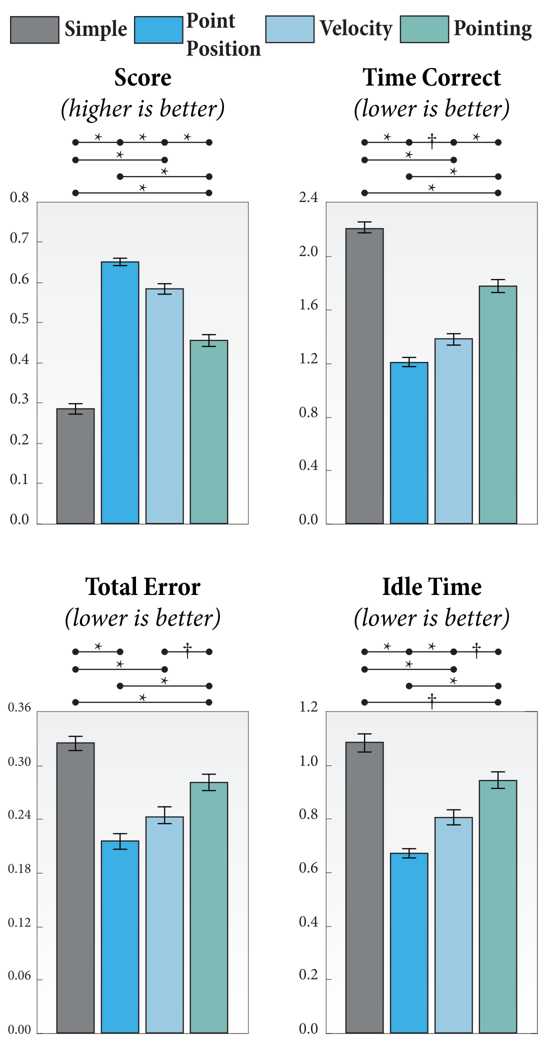

Fig. 8. Results from the third experiment comparing different heuristics *Point, Velocity*, and *Pointing* against *Simple* motions. Error bars indicate standard error. $*$ denotes $p \leq 0.001$, † denotes $p \leq 0.05$.

We ran three separate studies, recruiting a total of 144 participants, and assessed effects across three different robots. We show our method's improved task performance over the method by Dragan *et al.* along all objective measures and all robots. The motion qualities that contribute to the performance difference between the two approaches has not been empirically tested, and is left as important future work. However, we speculate that the overshooting of the goal using the method by Dragan *et al.* may sometimes be misleading, for instance, by perhaps suggesting another goal before veering back to the actual goal. This effect may be more evident with a densely sampled goal set where there are many non-goals crowding the actual goal position. In contrast, our approach does not repel non-goals, but rather attracts to the single goal position by minimizing the user's anticipated prediction error at every instant. Thus, our approach may convey more information about the goal position by not suggesting erroneous goal positions early in the trajectory. Exaggerated motions may have advantages in situations without potential for the exaggerations to be misinterpreted or for the extra movements to be problematic (e.g., cause collisions). For example, in the binary forced choice task of (Dragan and Srinivasa, 2013) with two goal objects placed close together, exaggeration may amplify the small differences in motion without suggesting other possible goals (as they do not exist). Understanding the tradeoffs of exageration would be interesting future work.

The work presented here has a number of limitations, including computational limitations of the proposed approach and limitations of our research in generalizability and applicability to real-world settings. A computational limitation of our approach is that it relies on constrained nonlinear optimization that may become slow or unreliable as the problems get larger and more complex. Our approach is also limited because it plans the entire trajectory at once, and its use in dynamic settings will require substantial improvements in the method. Our empirical validation offers limited generalizability, as our experiments have been limited to a single task; future work should explore how the proposed method extends to other tasks, particularly more complex collaborative scenarios. In addition, our implementation has been limited to a small set of constraints. In principle, the optimization-based approach readily admits richer problems through new constraints and objectives. For example, a cluttered environment can be encoded as a constraint using potential-field methods, and multiple goal constraints could be placed. In practice, however, these additions may be challenging, as the necessary functions must be derived and will most likely result in nonlinear programs that are difficult to solve. Finally, while we have not demonstrated our approach on robots other than arms or goals other than positional targets, such extensions are possible given the flexibility of our constraint-based approach and provide important areas of future research.

We have made a prototype implementation of our approach available as open source.[9]

9. Conclusion

We have provided a method for synthesizing intent-expressive (legible) motions. Our nonlinear-constrained-optimization-based approach allows for movement qualities to be encoded as objective functions. We have provided an objective that encodes legibility for robot arm motions. Our method provides flexibility in adapting to different task situations, flexibility in choosing heuristics, and adaptability to different robot designs. In three experiments, we have demonstrated that the method improves performance in a collaborative task over prior approaches and consistently provides these advantages over different robot designs and performance metrics. Our experimentation with different heuristics has shown that different encodings of the properties of legibility affect the collaborative performance of a human collaborator. Specifically, in some cases, the motion itself offers more legibility than an explicit signal such as pointing. Our work adds to the body of knowledge that shows that without any explicit prompting, participants naturally pick up movement cues and use them to infer the robot's intent, allowing them to improve their task performance.

Funding

This work was supported by the National Science Foundation (National Robotics Initiative Award Number 1208632) and the University of Wisconsin–Madison Office of the Vice Chancellor for Research and Graduate Education with funding from the Wisconsin Alumni Research Foundation.

Notes

1. `Python ad library`: https://pypi.python.org/pypi/ad/1.3.2https://pypi.python.org/pypi/ad/1.3.2
2. `Python scipy library`: https://www.scipy.org/https://www.scipy. org/
3. `wpi_jaco package`: https://github.com/RIVeR-Lab/wpi_ja cohttps://github.com/RIVeR-Lab/wpi_jaco
4. See http://www.ros.org/http://www.ros.org/package.
5. See http://www.kinovarobotics.com/http://www.kinovarobotics.com/
6. See http://www.trossenrobotics.com/p/phantomx-ax-12-reactor-robot-arm.aspxhttp://www.trossenrobotics.com
7. See http://www.universal-robots.com/products/ur5-robot/http://www.universal-robots.com/products/ur5-robot/
8. See https://unity3d.com/https://unity3d.com/
9. Open Source Implementation of Our Approach: https://github.com/uwgraphics/trajectoryoptimizer-public

References

Atkeson CG and Hollerbach JM (1985) Kinematic features of unrestrained vertical arm movements. *Journal of Neuroscience* 5(9): 2318–2330.

Barrett HC, Todd PM, Miller GF and Blythe PW (2005) Accurate judgments of intention from motion cues alone: A cross-cultural study. *Evolution and Human Behavior* 26(4): 313–331.

Betts JT (1998) Survey of numerical methods for trajectory optimization. *Journal of Guidance Control and Dynamics* 21(2): 193–207.

Blake R and Shiffrar M (2007) Perception of human motion. *Annual Review of Psychology* 58: 47–73.

Blakemore S and Decety J (2001) From the perception of action to the understanding of intention. *Nature Reviews Neuroscience* 2(8): 561–567.

Bodden C, Rakita D, Mutlu B and Gleicher M (2016) Evaluating intent-expressive robot arm motion. In: *2016 RO-MAN*.

Bortot D, Born M and Bengler K (2013) Directly or on detours? How should industrial robots approximate humans? In: *ACM/IEEE HRI*, pp. 89–90.

Busch B, Grizou J, Lopes M and Stulp F (2017) Learning legible motion from human–robot interactions. *International Journal of Social Robotics*, in press.

Dragan A, Bauman S, Forlizzi J and Srinivasa S (2015) Effects of robot motion on human–robot collaboration. In: *ACM/IEEE HRI*, pp. 51–58.

Dragan A, Lee K and Srinivasa S (2013) Legibility and predictability of robot motion. In: *ACM/IEEE HRI*, pp. 301–308.

Dragan A and Srinivasa S (2013) Generating Legible Motion. In: *Proceedings of Robotics: Science and Systems*.

Gergely G, Nádasdy Z, Csibra G and Bíró S (1995) Taking the intentional stance at 12 months of age. *Cognition* 56(2): 165–193.

Gleicher M (1994) *A Differential Approach to Graphical Manipulation*. PhD Thesis, Carnegie Mellon University, Pittsburgh, PA.

Gleicher M (1998) Retargeting Motion to New Characters. In: *ACM SIGGRAPH*, pp. 33–42.

Gleicher M (2001) Comparing constraint-based motion editing methods. *Graphical Models* 63(2): 107–134.

Hoffman G (2015) Evaluating fluency in human–robot collaboration. In: *RSS Workshop on Human–Robot Collaboration*.

Holladay R, Dragan A and Srinivasa S (2014) Legible robot pointing. In: *2014 RO-MAN*, pp. 217–223.

Kalakrishnan M, Chitta S, Theodorou E, Pastor P and Schaal S (2011) Stomp: Stochastic trajectory optimization for motion planning. In: *IEEE ICRA*, pp. 4569–4574.

Kamewari K, Kato M, Kanda T, Ishiguro H and Hiraki K (2005) Six-and-a-half-month-old children positively attribute goals to human action and to humanoid-robot motion. *Cognitive Development* 20(2): 303–320.

Kiesler S, Powers A, Fussell S and Torrey C (2008) Anthropomorphic interactions with a robot and robot-like agent. *Social Cognition* 26(2): 169–181.

Lichtenthäler C and Kirsch A (2014) Goal-predictability vs. trajectory-predictability. In: *Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction (HRI '14)*. New York: ACM Press, pp. 228–229.

Lichtenthäler C and Kirsch A (2016) Legibility of robot behavior: a literature review. Working paper or preprint.

Lichtenthaler C, Lorenzy T and Kirsch A (2012) Influence of legibility on perceived safety in a virtual human–robot path crossing task. In: *IEEE RO-MAN*, pp. 676–681.

Morasso P (1981) Spatial control of arm movements. *Experimental Brain Research* 42(2): 223–227.

Nikolaidis S, Dragan A and Srinivasa S (2016) Viewpoint-based legibility optimization. In: *The Eleventh ACM/IEEE International Conference on Human Robot Interaction (HRI '16)*. Piscataway, NJ: IEEE Press, pp. 271–278.

Ratliff N, Zucker M, Bagnell J and Srinivasa S (2009) Chomp: Gradient optimization techniques for efficient motion planning. In: *IEEE International Conference on Robotics and Automation, 2009, ICRA'09*, pp. 489–494. IEEE.

Schulman J, Ho J, Lee AX, Awwal I, Bradlow H and Abbeel P (2013) Finding locally optimal, collision-free trajectories with sequential convex optimization. *Robotics: Science and Systems* 9: 1–10.

Szafir D, Mutlu B and Fong T (2014) Communication of intent in assistive free flyers. In: *ACM/IEEE HRI*, pp. 358–365.

Szafir D, Mutlu B and Fong T (2015) Communicating directionality in flying robots. In: *ACM/IEEE HRI*, pp. 19–26.

Takayama L, Dooley D and Ju W (2011) Expressing thought: Improving robot readability with animation principles. In: *ACM/IEEE HRI*, pp. 69–76.

Thomas F and Johnston O (1995) *The Illusion of Life: Disney Animation*. Hyperion.

Van Breemen A (2004) Bringing robots to life: Applying principles of animation to robots. In: *CHI Workshop on Shaping Human–Robot Interaction*.

Witkin A and Kass M (1988) Spacetime constraints. *ACM SIGGRAPH Computer Graphics* 22(4): 159–168.

Woods S, Walters M, Koay K and Dautenhahn K (2006) Comparing human robot interaction scenarios using live and video based methods: towards a novel methodological approach. In: *IEEE International Workshop on Advanced Motion Control*, pp. 750–755.

Zhao M, Shome R, Yochelson I, Bekris KE and Kowler E (2014) An experimental study for identifying features of legible manipulator paths. In: *ISER*, pp. 639–653.