Foundations and Trends® in Robotics
Vol. 5, No. 4 (2014) 261–349

DOI: 10.1561/2300000052

A Survey of Methods for Safe Human-Robot Interaction

Przemyslaw A. Lasota
Massachusetts Institute of Technology,
USA
plasota@mit.edu

Terrence Fong
NASA Ames Research Center,
USA
terry.fong@nasa.gov

Julie A. Shah
Massachusetts Institute of Technology,
USA
julie_a_shah@csail.mit.edu

Contents

Abstract

Ensuring human safety is one of the most important considerations within the field of human-robot interaction (HRI). This does not simply involve preventing collisions between humans and robots operating within a shared space; we must consider all possible ways in which harm could come to a person, ranging from physical contact to adverse psychological effects resulting from unpleasant or dangerous interaction. In this work, we define what safe HRI entails and present a survey of potential methods of ensuring safety during HRI. We classify this collection of work into four major categories: safety through control, motion planning, prediction, and consideration of psychological factors. We discuss recent work in each major category, identify various sub-categories and discuss how these methods can be utilized to improve HRI safety. We then discuss gaps in the current literature and suggest future directions for additional work. By creating an organized categorization of the field, we hope to support future research and the development of new technologies for safe HRI, as well as facilitate the use of these techniques by researchers within the HRI community.

P. A. Lasota, T. Fong and J. A. Shah. *A Survey of Methods for Safe Human-Robot Interaction*. Foundations and Trends® in Robotics, vol. 5, no. 4, pp. 261–349, 2014.
DOI: 10.1561/2300000052.

1

Introduction

Human-robot interaction — collaboration, communication, and cooperation between humans and robots — is a rapidly growing area of robotics research. From introducing robotic co-workers into factories (Unhelkar et al., 2014; Gleeson et al., 2013; Knight, 2013), to providing in-home robot helpers (Graf et al., 2004), to developing robotic assistants for astronauts on-board the International Space Station (ISS) (Fong et al., 2013; Diftler et al., 2011; Bualat et al., 2015), there are a wide variety of beneficial applications for HRI. Whether this interaction involves an industrial robot, mobile manipulator, free-flyer, or even a self-driving car or wheelchair, one should always approach the development of HRI platforms and technologies from a safety-focused perspective. The successful advancement of HRI depends upon safety being a top priority and an integral component of any HRI application. In order to understand how to tackle the challenging problem of ensuring safety in HRI, it is necessary to clearly define what safe HRI entails and what has been accomplished thus far in terms of standardizing safety metrics and methods, and survey the current literature to identify areas that warrant further research and development.

1.1 Defining Safety in HRI

In order to ensure safe HRI, it is necessary to first understand what constitutes safety and its various components. In 1942, science fiction writer Isaac Asimov proposed three "Laws of Robotics," the first of which states: "A robot may not injure a human being or, through inaction, allow a human being to come to harm" (Asimov, 1942). Inspired by Asimov's definition, we can identify two distinct ways in which a robot could inflict harm on a human being.

The first is through direct physical contact. In simple terms, in order for HRI to be safe, no unintentional or unwanted contact can occur between the human and robot. Furthermore, if physical contact is required for a given task (or strict prevention of physical contact is neither possible nor practical) the forces exerted upon the human must remain below thresholds for physical discomfort or injury. We define this form of safety in HRI as **physical safety**.

Preventing physical harm alone, however, does not necessarily translate to stress-free and comfortable interaction. Consider, for example, a hypothetical manufacturing scenario in which a robot uses a sharp cutting implement to perform a task in proximity to human workers, but is programmed to stop if a human gets too close. While direct physical harm is prevented through careful programming, this type of interaction can be stressful for humans. Importantly, psychological discomfort or stress can also be induced by a robot's appearance, embodiment, gaze, speech, posture, and other attributes (Mumm and Mutlu, 2011; Butler and Agah, 2001).

Stress can have serious negative effects on health (McEwen, 1993), which makes stressful HRI a potential source of harm. Furthermore, psychological discomfort caused by any of the other aforementioned factors, as well as robotic violation of social conventions and norms during interaction, can also have serious negative effects on humans over time. We define the prevention of this type of indirect, psychological harm as maintaining **psychological safety**. It is important to note that psychological harm, in contrast with physical harm, is not limited to proximal interaction, as it can also be sustained through distal interaction via a remote interface.

As HRI can be applied in a multitude of domains, we apply a broad definition of the term “robot” in the context of this work. Although the individual works described in this survey are generally presented in the context of interaction with one type of robot in a specific domain, the methods for safety in HRI we present in the following sections are domain independent and relevant to a wide array of robot types, such as manipulator arms, drones, personal robots, and self-driving cars.

1.2 Safety Standards and Criteria

The development of guidelines and requirements in the form of international safety standards represents an important effort toward ensuring safety during human-robot interaction. The International Organization for Standardization (ISO) has been working toward releasing documents that specify how best to maintain safety during interaction between humans and industrial robots. The first step in this process was the release of the ISO 10218 document entitled “Robots and robotic devices – Safety requirements for industrial robots,” which is composed of two parts: “Robots” and “Robot systems and integration” (International Organization for Standardization, 2011a,b). The ISO 10218 outlines some potential methods of safe collaborative manipulation — for example, *speed and separation monitoring* and *power and force limiting* — as well as relevant safety requirements.

The technical specification accompanying this document is the ISO/TS 15066 (entitled “Robots and robotic devices – Collaborative robots”) (International Organization for Standardization, 2016). This technical specification provides additional information and details about how to achieve the requirements established by ISO 10218. It includes quantitative biomechanical limits, such as allowable peak forces or pressures for various parts of the body, as well as equations for speed and separation monitoring. In support of the development of the ISO technical specification, organizations including the National Institute of Standards and Technology (NIST) collaborated with ISO to develop protocols and metrics that would allow for characterization of the effectiveness of a robot’s safety methods (National Institute of Standards and Technology, 2013).

The safety criteria mentioned above were developed in part through study of human-robot collisions. Recent experiments have incorporated collisions between robots and instrumented crash-test dummies, both in simulation (Oberer and Schraft, 2007) and using actual physical hardware (Haddadin et al., 2007, 2009). Other research has incorporated crash tests involving simulated human tissue, such as abdominal samples collected from pigs (Haddadin et al., 2012). Work with actual human-robot collisions has also been conducted to classify pain (Povse et al., 2010) and injury thresholds (Fraunhofer IFF, 2013), as well as to investigate the effectiveness of control strategies (Haddadin et al., 2008). Various injury prevention criteria for HRI have resulted from these works (Jung-Jun Park and Jae-Bok Song, 2009; Oberer and Schraft, 2007; Haddadin et al., 2012). Importantly, the findings are discussed in relation to the ISO standard regulations, providing feedback for their further refinement and improvement. (Haddadin (2013) have presented a detailed discussion of the limitations of the current standards and proposed improvements.) By combining the efforts of academic and industrial research groups and standardization organizations, more suitable and relevant standards and metrics can be developed and introduced in subsequent revisions of the ISO standards.

While the development of the aforementioned international safety standards represents a crucial first step toward improving HRI safety, it is important to note that these standards are being developed specifically for industrial applications. Although many of the principles would likely transfer to other types of robots and applications, the standards' scope is too narrow to fully address other uses, such as robotic tour guides or assistants for the elderly. We therefore must look beyond these industrial standards in order to identify all the pertinent aspects of safe HRI and the various possible safety methods that could be employed to address them.

1.3 Goals and Scope

The main goal of this work is to organize and summarize the large body of research related to facilitation of safe human-robot interaction. This survey describes the strategies and methods that have been developed thus far, organizes them into subcategories, characterizes relationships between the strategies, and identifies potential gaps in the existing knowledge that warrant further research.

1.3.1 Method

As there is a vast amount of work that could be applied to safe HRI, it was imperative to select a cohesive and meaningful subset of research. We conducted a survey to identify the various *methods* that could be utilized to make HRI safe. This is in contrast to other work, such as that of Vasic and Billard (2013), who partially outlined these possibilities but organized the paper according to application and focused on other aspects of safety, such as potential sources of danger and liability.

Also, we chose to focus our survey on recent research. A survey on safety in HRI by Pervez and Ryu (2008) covered much of the earlier work conducted within the field; this review mostly discusses research that had been published since that survey. Additionally, our survey focuses on the safety aspects of proximal HRI, and so we do not consider, for example, safety concerns during remote operation. Furthermore, we chose to focus this survey on interaction with robots acting as independent entities, and so we do not consider the regime of interaction with wearable robots, such as exoskeletons or orthotics. (We direct the reader interested in the latter topic to recent works in both industrial and medical applications (Kolakowsky-Hayner et al., 2013; O'Sullivan et al., 2015; Zeilig et al., 2012).) This survey also does not focus on the psychological safety aspects of interacting with social robots and the potential impact such robots can have when emulating human personality traits or social behaviors. (The reader interested in these aspects should consult works relating social psychology to robotics, such as papers by Young et al. (2008) and Fong et al. (2003).)

For the present work, we chose not to focus on robot hardware development as a potential method of ensuring safety in HRI. In recent years, robotics manufacturers have become increasingly involved in the development of robots designed specifically for proximal HRI. (Examples of such robots include the ABB YuMi (ABB, 2015), the RethinkRobotics Baxter and Sawyer (Robotics, 2015a,b), and the KUKA LBR (KUKA, 2015).) There has also been a significant amount of work in hardware development for safe HRI within the academic community, and the technologies used by these new robots are often a product of this research. This includes work on new actuators designed to be human-safe, such as series elastic actuators (SEA) (Pratt and Williamson, 1995), variable impedance actuators (VIA) (Vanderborght et al., 2013), distributed macro-mini actuation (Zinn et al., 2004), or external hardware, such as robot skins (Hoshi and Shinoda, 2006). (We direct the reader interested in compliant actuator designs to the review by Ham et al. (2009).)

Defining the scope of our work as outlined above, our selection process focused on papers published between 2008 and 2015 from the conference proceedings of the ACM/IEEE International Conference on Human-robot Interaction (HRI), IEEE International Conference on Robotics and Automation (ICRA), Robotics: Science and Systems (RSS), the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), the IEEE RAS/EMBS International Conference on Biomedical Robotics and Biomechatronics (BioRob), and the International Conference on Advanced Robotics (ICAR), as well as journal articles published in the *International Journal of Robotics Research* (IJRR), the *Journal of Mechanical Science and Technology* (JMST), the *IEEE Transactions on Robotics* (T-RO), the *IEEE Transactions on Automation Science and Engineering* (T-ASE), and the *Journal of Robotic Systems*.

We first grouped papers according to theme; common keywords among papers within each theme were then used as further search criteria. We focused our final selection on publications with higher impact factors and according to the selectivity of the publication venue. We relaxed these constraints if a topic associated with a keyword was underrepresented or the work was published within the last 3 years. We also recursively investigated works cited by the collected papers to identify additional potential sources. The resulting collection was then organized into the following main themes: safety through control, planning, prediction, and consideration of psychological factors, as depicted in Figure 1.1.

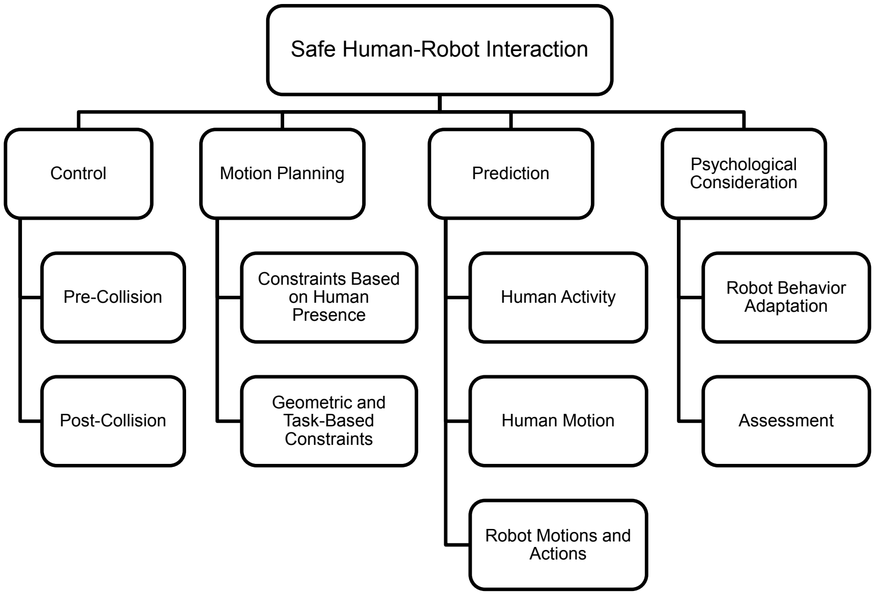
Safe Human-Robot Interaction
Control
Motion Planning
Prediction
Psychological Consideration
Pre-Collision
Post-Collision
Constraints Based on Human Presence
Geometric and Task-Based Constraints
Human Activity
Human Motion
Robot Motions and Actions
Robot Behavior Adaptation
Assessment

Figure 1.1: Diagram depicting the major methods of providing safety in HRI.

1.3.2 Organization

The remainder of this monograph is divided into sections based on the four main aspects of safety in HRI depicted in Figure 1.1. Each section describes, in detail, a selection of recent related works, synthesizes these works into various sub-topics, and outlines the relationships between them.

In Section 2: Safety Through Control, we describe pre- and post-collision control methods for providing safe HRI. The former category deals with control methods prior to contact between a human and robot. This involves limiting key parameters, such as velocity or energy, or preventing collisions from occurring through the use of methods including defining safety regions, tracking separation distance, and guiding robot motion away from humans. The post-collision sub-category involves techniques such as minimizing injury by switching between various control methods when a collision is detected, distinguishing between intentional and non-intentional contact, and allowing for safe physical contact if necessary for effective collaboration.

In Section 3: Safety Through Motion Planning, we highlight work focused on planning safer robot paths and motions in order to avoid potential collisions. By taking various human-related parameters such as separation distance or gaze direction directly into account when forming motion plans, a robot is able to choose safer and more efficient paths and motions.

In Section 4: Safety Through Prediction, we discuss the various ways that human and robot behavior prediction can allow for safer HRI. This involves predicting human actions and motions through a variety of methods, including sequence matching, probabilistic plan recognition, and motion characteristic analysis. Also, as HRI is inherently a two-way interaction, it is also important to consider the predictability of the robot in order to allow the human to anticipate the robot's motions and actions.

In Section 5: Safety Through Consideration of Psychological Factors, we focus on methods of assessing and maintaining psychological safety during HRI. As mentioned in Section 1.1, psychological safety maintenance involves ensuring that interaction remains stress-free and comfortable. Work in this field has included the development of metrics through physiological sensing, questionnaires, and behavioral metrics and identifying which factors — such as a robot's size and speed or a human's prior experience with robots — can affect perceived safety and comfort.

Finally, in Section 6: we discuss possible future directions for research that would benefit the field of HRI safety. We draw upon lessons learned from prior work, identify gaps in various research subcategories, and offer specific suggestions for what could be investigated further in order to address these gaps.

2

Safety Through Control

One common method for achieving safety during human-robot interaction is through low-level control of robot motion. This type of safety provision is often the simplest method of enabling safe human-robot coexistence, as it does not require complex prediction models or planners — nor, in some cases, does it even require sensing to monitor the human. Nonetheless, implementation of these solutions can be quite complex, as this method often includes time-critical constraints that require rapid execution.

Control methods for improving safety are divisible into two main categories: pre- and post-collision. Pre-collision control methods are implemented before human-robot collision occurs, either by ensuring collision does not occur in the first place or by bounding robot parameters such as velocity or energy. If unexpected or unpreventable contact occurs, post-collision control methods are designed to quickly detect the collision and minimize harm to both the human and robot. (Note that in this context, "collision" is not limited to blunt impacts, but can also include other harmful forms of contact, such as shearing, cutting, or puncturing.)

2.1 Pre-Collision Methods

Pre-collision control methods, sometimes referred to as "prevention" methods, are techniques intended to ensure safety during HRI by monitoring either the human, the robot, or both and modifying robot control parameters prior to incidence of collision or contact. The various techniques and methods designed to provide safety through pre-collision control discussed in this section are depicted in Figure 2.1 below.

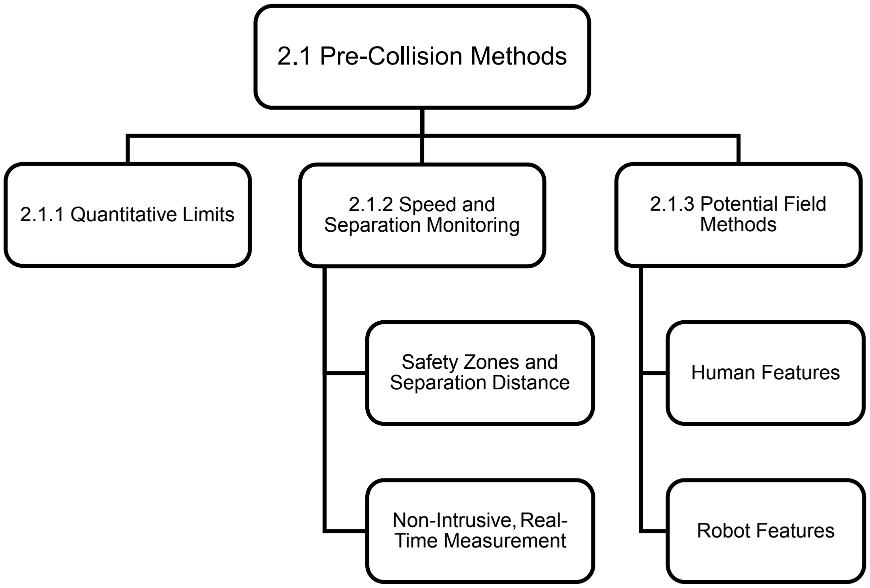
2.1 Pre-Collision Methods
2.1.1 Quantitative Limits
2.1.2 Speed and Separation Monitoring
2.1.3 Potential Field Methods
Safety Zones and Separation Distance
Non-Intrusive, Real-Time Measurement
Human Features
Robot Features

Figure 2.1: Diagram depicting the pre-collision control methods discussed in Section 2.1.

2.1.1 Quantitative Limits

One major subset of this category is focused on providing quantitative guarantees that a robot cannot pose any threat to a human, even in the event of a collision. This can be achieved by limiting a variety of parameters, such as the robot's joint velocity, energy, or potential exertion of force. Broquere et al. (2008), for example, developed a trajectory planner that limits jerk, acceleration, and velocity. The ability to compute new trajectories in real time is critical for applying such a planner in a dynamic HRI setting. The planner developed by Broquere et al. meets this need, as it constructs trajectories using polygonal chains of cubic functions for which the parameters are computed directly, allowing for real-time control.

In the approach taken by Laffranchi et al. (2009), real-time adjustment is even more critical: Instead of planning trajectories that might require adjustment due to changes in the positions of humans or other objects in the environment, their method focused on real-time tracking and limiting of the total amount of energy stored within the system — namely, the sum of kinetic, gravitational potential, and elastic potential energies. In this work, the controller was implemented on a prototype single-joint series elastic actuator and tested in two cases: accidental collision and free motion. The actuator was commanded to follow a sinusoidal path in both cases, but a foam block was placed in the actuator's path during the former case. Laffranchi et al found that the energy of the system remained below a predefined threshold through an online modification of the reference value of the position controller. Heinzmann and Zelinsky (2003), on the other hand, developed a control approach that limits the potential force of impact with static obstacles by imposing a safety envelope on the torque commands of a position control algorithm. This method successfully limited impact forces, regardless of where on the robot the collision occurred.

Finally, Haddadin et al. (2012) took the unique approach of embedding injury knowledge into robot control by studying the relationships between robot mass, velocity, and impact geometry with injury. As the authors wrote, attempting to form a direct relation between these various *input* parameters and injury is different from other approaches that rely upon deriving relationships with robot collision *outputs*, such as exerted forces or stresses. In this work, the authors identified impact geometry primitives and performed drop tests on abdominal samples from pigs while varying mass and speed, and used the international medical classification system developed by the AO Foundation (AO,

2015) to analyze the injuries. Risk curves for each impact geometry primitive were derived from the results from these tests, establishing a relationship between impact speed and impact geometry, mass, and the impacted body part. The authors then used these curves to scale the velocity of the robot to ensure that injury above a certain threshold could not occur in the event of an unexpected collision.

2.1.2 Speed and Separation Monitoring

Slowing down or stopping the robot through the use of safety zones or distance of separation is another method of preventing collision through control. The robotics and automation company ABB has developed SafeMove, a system that utilizes programmable, complex safety zones that can control robot speed (Kock et al., 2006). This system allows for safer interaction between a human and an industrial robot by using external sensing to track the presence of humans or objects within safety zones and adjusting the robot's speed to the zones' predefined limits. In contrast with static, predefined safety zones, Vogel et al. (2013) developed a system that incorporates dynamically changing zones based on robot joint positions and velocities and displays these zones on the surface around the robot via a projector. The control system detects when this virtual safety zone is entered, and stops the robot as needed.

Lasota et al. (2014) developed a safety system for close-proximity interaction with standard industrial robots that leverages accurate sensing of a human's location and the robot's current configuration to rapidly calculate the distance of separation between the human and robot. This measurement is then used to gradually decrease the robot's speed according to a tunable function that can be adjusted via task-dependent parameters. This approach eliminates the need for predefining conservative safety zones; however, while scaling the robot's velocity as a function of separation distance can be an effective method of improving safety in HRI, slowing the robot can often lead to significant decreases in the productivity of human-robot collaboration.

To address this, Zanchettin et al. (2015) developed a velocity scaling approach that takes advantage of redundant degrees of freedom, with the goal of maintaining safety while retaining productivity. In this work, a safety region is calculated based on the robot's velocity and braking distance, as well as a *clearance parameter* that takes uncertainties in measurement and modeling into account. The collision avoidance, calculated at the joint space level to allow for real-time deployment, uses redundant degrees of freedom to move the robot's joints away from the human while still maintaining the correct end effector position. This enables the robot to continue to perform its task while maintaining both a greater distance of separation from the human and a higher speed.

One significant challenge of deploying systems such as the ones cited above is providing nonintrusive methods of accurate human localization. One viable method presented by Rybski et al. (2012) uses sensor fusion techniques to combine data collected from stereo and range cameras in order to monitor an industrial workcell. The system detects people and robots within the environment and generates dynamically changing "danger zones" based on the position and trajectory of the robot. In contrast, Flacco et al. (2012) utilized depth sensors to estimate the distance between the robot and both static and moving obstacles. This real-time distance measurement, as well as an estimate of obstacle velocity, was then used with a controller based on repulsive force vectors as a collision prevention technique.

To reduce the possibility of occlusion and to accommodate unstructured tasks for which the optimal sensing locations are not known a priori, Buizza Avanzini et al. (2014) utilized an on-board sensing approach. The authors developed a distributed distance sensor and an optimization strategy for placement of sensor nodes on a robot's body. The distance sensor was integrated into a framework utilizing the *danger field* criterion: a scalar field based on the robot's configuration and velocity (Lacevic and Rocco, 2010). In the developed framework, the robot attempts to maintain task consistency by hierarchically abandoning tasks according to their specified priority: For example, the robot may maintain position but change its orientation as the danger metric increases, then finally abandon its position if the danger metric increases beyond a certain threshold.

2.1.3 Potential Field Methods

Another popular approach to collision prevention via robot control is calculation of danger criteria and fields, such as with the *potential field* approach developed by Khatib (1986). This method allows for more complex safety behaviors by defining a field of repulsive vectors that guide the robot's motion, modifying its trajectory in response to dynamically changing environmental factors. One recent work that used this control approach specifically for HRI safety is that of Calinon et al. (2010), in which the controller utilized a risk criterion based on the distance from the robot to the human's head, as well as the human's gaze direction, to safely guide the robot's motion along trajectories derived from kinesthetic teaching.

The potential field approach is also often deployed as a component of integrated safety frameworks. One such framework by Kulić and Croft (2007) incorporated a safe control module that considers safety factors such as separation distance and velocity to generate a danger index to be used by a potential field controller. Furthermore, the estimated affective state of the user, inferred from skin conductance and heart rate measurement, was also integrated into this danger index. A framework developed by De Luca and Flacco (2012) differs from this approach in that it utilizes two unique collision avoidance methods: one for the robot's end effector, and one for the other parts of the robot. This was also done in the work by Flacco et al. (2012) mentioned earlier, in which repulsive vectors were based on separation distance measurements.

Haddadin et al. (2010b) also developed a collision avoidance technique based on the potential field method, but their framework accommodates not only the virtual forces caused by proximity to the robot, but also actual physical contact. The algorithm, designed to have sufficiently low complexity to run in real time, is based on local reactive motion planning along with velocity scaling, a function not only of distance but also direction of approach. The resulting system is capable of producing smooth paths that avoid sudden accelerations and are thus more physically interpretable by humans.

In another approach, Polverini et al. (2014) developed the *kineostatic safety field*, a safety assessment at the kinematic level and an extension of the *danger field* concept by Lacevic and Rocco. The key improvement of the safety field upon the danger field is that it takes into account the relative motion between the source of danger and where the field is computed. Furthermore, the safety field is also dependent upon the shape and size of the source of danger. The authors validated the safety field concept by utilizing it as part of a real-time controller on an early version of the two-armed YuMi robot (ABB, 2015), and showed that the field can be utilized for preventing both self-collision and human-robot collision.

In Section 2.1, we presented three pre-collision control techniques for safety in HRI. First, we discussed how the various methods of calculating and limiting velocities, potential impact forces, and energy allow systems to provide quantitative guarantees about the robot's ability to inflict harm in the event of unwanted collision. Although such techniques can result in overly conservative motions in the absence of an imminent collision threat, by imposing global limits on the robot's motions, pre-collision safety methods can provide such guarantees without having to rely upon accurate and robust detection and tracking of colocated humans.

Next, we highlighted methods that gradually slow a robot's motion based on safety zones or separation distance from the human. These techniques allow for greater flexibility than strictly limiting the robot using parameters such as energy or velocity, but they also require a low-latency implementation with robust tracking of the human within the given space. Furthermore, nonintrusive methods of human localization and distance measurement are critical for real-world deployment.

Finally, we introduced techniques based on the potential field approach, which allows for implementation of more complex collision-avoidance behaviors beyond simply adjusting the robot's speed. The efficacy of such a method, however, is directly linked to the strategy used to construct the potential field. This has led to a variety of implementations involving not only separation distance, but additional factors such as the direction of approach, affective state, and human gaze direction.

2.2 Post-Collision Methods

Using a variety of control strategies to prevent collisions can also be an effective method of improving safety during HRI. However, depending upon myriad factors including the type of robot, sensing system, and assigned task, strict collision prevention is not always possible or practical — in fact, some human-robot collaborative tasks may require a certain level of physical contact. As a result, another body of research has focused on development of control strategies that further ensure safety through detection of and appropriate reaction to human-robot collisions. As mentioned previously, this includes not only blunt impacts but other harmful forms of contact, such as shearing, cutting, or puncturing. Figure 2.2 depicts the hierarchy of topics involved in utilizing post-collision control as a safety method for HRI.

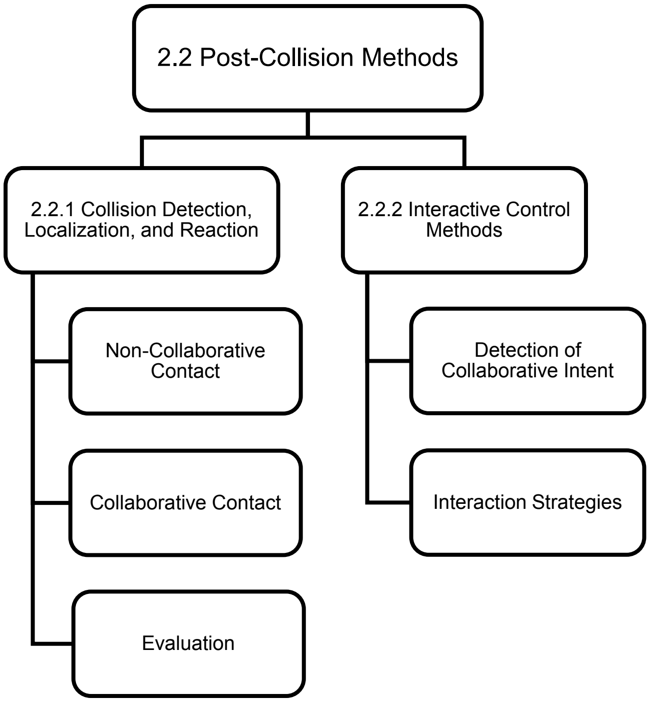
2.2 Post-Collision Methods
2.2.1 Collision Detection, Localization, and Reaction
2.2.2 Interactive Control Methods
Non-Collaborative Contact
Collaborative Contact
Evaluation
Detection of Collaborative Intent
Interaction Strategies

Figure 2.2: Diagram depicting the post-collision safety methods discussed in Section 2.2.

2.2.1 Collision Detection, Localization, and Reaction:

The first step toward utilizing post-collision control methods for HRI safety is detecting whether a collision has occurred. As the use of external sensing is often impractical, much of the work in detection and localization of human-robot collisions has focused on methods that incorporate on-board sensing. For example, De Luca et al. (2006) presented a collision detection system requiring only proprioceptive measurement. This system utilizes a collision detection signal that can be calculated using only the joint positions, velocities, and commanded torques, and incorporates a measure of energy defined as the sum of kinetic and gravitational potential energies. Furthermore, the system utilizes a collision identification signal calculated from the same quantities as the detection signal to provide information about which links experienced impact and from what direction, allowing the system to move the robot away from the collision site after impact detection.

Similar to the work by De Luca et al., Geravand et al. (2013) developed a detection and reaction system that does not require torque sensing. One key advantage of this system is that it does not rely upon knowledge of joint velocities, which often requires numerical integration that introduces noise; rather, it makes use of motor current measurements. This system, designed for industrial robots with a closed-control architecture, also does not require a priori knowledge of the dynamic model. It is not only capable of detecting collision, but also whether the collision was intentional or non-intentional, in order to switch the robot to a "collaborative mode" in which the robot accepts redirection from the human as needed. This switch is performed through parallel use of high- and low-pass filters on the motor currents, with the assumption that non-intentional, hard impacts generate a high-frequency signal and intentional, soft impacts generate a low-frequency signal. The system analyzes filtered signals and compares them with time-varying thresholds: a signal exceeding a threshold after being run through the high-pass filter indicates non-intentional impact, while intentional contact is identified if at least one of the low-pass-filtered signals exceeds a threshold and no thresholds are exceeded by signals run through the high-pass filter.

Another system developed by Golz et al. (2015) is also capable of discriminating between intentional and unintentional robot collisions utilizing not only machine learning but also a model of physical contact. This model, along with insights gained through observation of real impact data, were used to derive a set of features for classification with a non-linear support vector machine (SVM). The authors were able to show, both through simulation and results from physical experiments, that the classifier is capable of accurately discriminating between intentional and unintentional collisions online.

De Luca et al. (2009) developed a collision detection system for a prototype variable stiffness actuator that, similarly to the system developed by Geravand et al., does not require torque sensing. Their momentum-based system is combined with an active reaction strategy that simultaneously moves the robot arm and reduces its stiffness to allow the arm to gently bounce away from the collision and come to a stop.

In addition to developing collision detection and reaction strategies, several researchers have also quantitatively assessed the effectiveness of post-collision control strategies. For example, Haddadin et al. (2008) tested a collision detection and reaction scheme in a controlled experiment involving specialized hardware setups, including instrumented crash-test dummies. Impacts with actual humans at the chest and upper arm were performed and analyzed as well. The researchers presented results for four post-collision control strategies indicating how their post-impact force time-series profiles differed, and found that the implemented system was successful at maintaining forces below thresholds for harm. The same control strategies were tested in the work by De Luca et al. (2006) mentioned previously, in which the authors compared residual torque time-series during collision with a balloon.

Vick et al. (2013) presented a post-collision safety system for a standard industrial robot that uses estimations of external forces to limit torques and prevent exertion of force beyond a specified threshold. The researchers evaluated the effectiveness of this system in two experiments: one involving a human pushing against a stationary robot, and another involving human interference with a robot in motion. In the former, a participant applied force to the robot as it attempted to hold its position. Once the applied force surpassed a preset threshold of 150N, the robot would move away from the source in order to limit the contact force to the preset value. In the latter experiment, the robot attempted to perform a sinusoidal motion through position control while a human interfered with its motion. If the force exerted on the human surpassed the given threshold, the robot would modify its path to reduce the force. The authors were able to show through quantitative assessment that the predefined limit was successfully maintained.

Haddadin et al. (2010a) studied soft-tissue injury (such as abrasions, contusions, lacerations, and punctures) caused by robots holding sharp tools. The authors performed experiments in which robots performed stabbing and cutting motions on silicone, pig tissue, and human volunteers, and evaluated three collision reaction strategies in order to determine their effectiveness at reducing exerted forces and penetration depth. The results indicated significant potential for reducing soft-tissue injuries during HRI if appropriate control responses are deployed, even with robots holding sharp objects and moving at speeds as high as 0.75 m/s.

2.2.2 Interactive Control Methods

As mentioned in the previous section, determining whether contact is intentional or unintentional is often a goal of collision detection systems. Upon detection of intentional contact, specialized safety measures and methodologies that differ from the “detection and reaction” paradigm are required: Instead of simply moving away from the collision or switching control methods in order to minimize harm, the robot must now reason about the human’s collaborative intent and how best to support him or her during the interaction.

In addition to the works by Geravand et al. (2013) and Golz et al. (2015) mentioned earlier, several other researchers have developed systems that allow for safe collaborative control modes. One framework by De Luca and Flacco (2012) infers whether the user wants to enter a collaborative mode based on his or her gesture and speech. This framework designates specific parts of the human body that the robot is allowed to make contact with during collaboration (e.g., the hands), and those with which the robot is not allowed to make contact (e.g., the head). Once collaboration is initiated, contact forces with the allowed human contact points are estimated and the robot is controlled such that predefined thresholds are not exceeded, while restricted contact points are avoided altogether.

Erden and Tomiyama (2010) developed an interactive control scheme for back-drivable robots that does not require a dynamics model or joint velocities — as with the residual method used by De Luca et al. — nor measurement of joint torques or motor current. In this approach, a human moves a robotic manipulator via continuous contact, and the system determines his or her intent through calculation of control effort based on conservation of momentum. Namely, in a back-drivable, gravity-compensated robot, the total momentum added to the system by the user and the additional momentum required by the robot to return to the system to a stop is zero. Through knowledge of the momentum delivered by the controller, the momentum and force exerted by the human can be estimated. Once the interaction mode is active, the user can then guide the robot within a gravity-compensated mode[1].

As we discussed in Section 2.2, the foundation of any effective post-collision control approach to safe HRI is the ability to accurately detect and localize a collision and then appropriately react to it. Each of the methods presented above requires different robot sensors and has unique benefits and drawbacks, making optimal deployment of these and other, similar methods dependent upon the specific robot being used and the task being performed. Furthermore, thorough evaluations are imperative in order to understand the efficacy of the collision reaction strategies. The works listed above have made significant headway toward this goal, testing post-collision methods with crash-test dummies and even soft-tissue samples from animals.

[1] In general, for continuous interaction such as that presented in work by Geravand et al. (2013), De Luca and Flacco (2012), and Erden and Tomiyama (2010), some form of compliance control is needed. To learn more about recent advancements and applications of compliance control in the field of HRI, we direct the reader to a recent survey by Khan et al. (2014).

As interaction scenarios become increasingly complex, the ability to detect whether human-robot contact is intentional or accidental is key to selecting an appropriate robot behavior to ensure safe interaction. As seen from the related works above, collaborative contact requires special considerations, such as where on the human body the contact can occur, as well as the monitoring and limiting of forces during the interaction.

2.3 Summary

In this section, we discussed a variety of control-based methods for facilitating safe HRI. First, we outlined pre-collision methods, which attempt to prevent unwanted contact in three distinct ways: by enforcing quantitative limits on parameters such as speed or energy, by monitoring the physical distance between the human and robot and adjusting robot speed accordingly, and by using the potential field approach to guide the robot away from the human. We addressed a variety of trade-offs between these approaches, and noted that the three methods allow for progressively more complex safety behaviors, but at the cost of more-complex implementation.

As strict prevention of collisions is not always possible, the other class of control-based methods of safety in HRI discussed here is focused on how to minimize harm once collision occurs. These post-collision methods function by detecting a collision, localizing where on the robot contact was made, and deciding how to react. Furthermore, we addressed collaborative contact with the robot, along with the implications for how control strategy changes in such a scenario with regard to maintaining safety.

Overall, control-based methods have been shown to be effective tools for safe HRI. In general, these methods do not require complex models of the environment and, in some cases, require limited or no tracking of the human. This quality improves robustness, as these approaches do not need to rely upon accurate tracking or potentially faulty models.

However, control-based techniques tend to be purely reactive, and as such, can prove to be insufficient for maintaining safe HRI. By taking advantage of knowledge about the environment and task at hand and formulating appropriate models, more proactive approaches to maintaining safety involving planning and prediction can be realized. We discuss these topics in Sections 3 and 4, respectively.

3

Safety Through Motion Planning

Providing safety through real-time control can prevent or mitigate unwanted collisions between humans and robots; however, such methods can be insufficient with regard to both safety and efficiency in many applications. Results from an experiment by Lasota and Shah (2015) assessing close-proximity human-robot collaboration indicated that simply preventing collisions as they are about to occur can lead to inefficient human-robot interaction and negatively impact perceived safety and comfort. (For a dedicated discussion of methods useful for ensuring HRI does not cause psychological discomfort, see Section 5.)

In this experiment, participants performed a collaborative task with a robot operating in two distinct modes: a standard mode in which the robot determined the quickest path to its goals and employed a pre-collision safety system based on separation distance to slow and stop its motion (Lasota et al., 2014), and an adaptive mode in which the robot used human-aware motion planning to avoid portions of the shared workspace that it expected the human to occupy.

The researchers found that the human-aware motion planner led to better team fluency, as measured by quantitative metrics such as task execution time and the amount of concurrent motion. The motion planner was also associated with greater satisfaction with the robot as a teammate and higher perceived levels of safety and comfort among participants, as evaluated through questionnaire responses. Although the control-based system maintained physical safety in both modes, the lower degree of perceived safety observed during operation in the standard mode could have a significant negative impact on psychological safety.

The fact that, in certain scenarios, collision prevention and alleviation through low-level control has been shown to lead to significantly poorer safety and efficiency compared with human-aware motion planning provides significant motivation for utilizing motion planning as a safety measure during HRI. This would involve the development of motion planners that directly consider human presence and movement when computing robot paths and motions, as well as motion planners capable of reasoning on both geometric and task-based constraints and supporting rapid, real-time replanning. These and other related topics covered in this section are depicted in Figure 3.1.

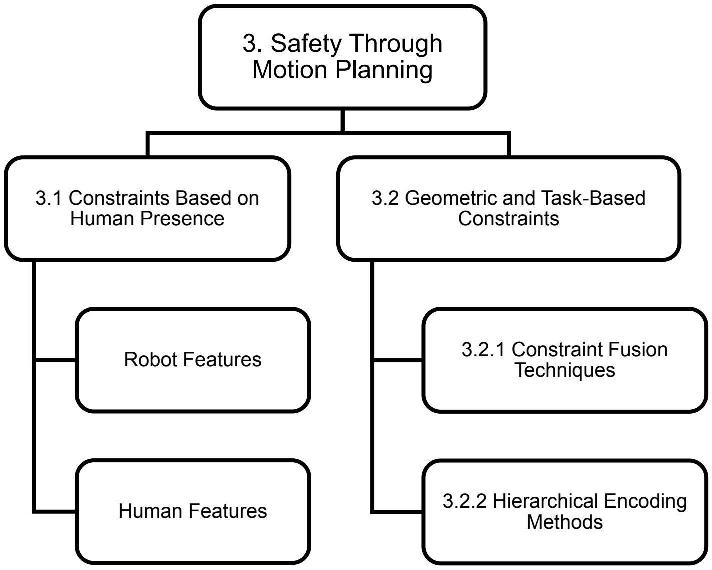
3. Safety Through Motion Planning
3.1 Constraints Based on Human Presence
3.2 Geometric and Task-Based Constraints
Robot Features
Human Features
3.2.1 Constraint Fusion Techniques
3.2.2 Hierarchical Encoding Methods

Figure 3.1: Diagram depicting the constraints, features, and techniques discussed in Section 3.

3.1 Constraints Based on Human Presence

The primary method by which robot motion planning can serve as a tool for HRI is through direct consideration of constraints related to the presence of human agents, such as distance of separation, human gaze direction, and robot motion legibility. This can be achieved by directly incorporating parameters such as these into the criteria and cost functions used by motion planning algorithms. Doing so enables motion planners to specifically consider how best to plan around the presence of humans, as opposed to treating humans as generic obstacles within an environment. Direct consideration of human-based constraints makes these motion planners very useful for improving safety in HRI through their ability to proactively avoid motion conflicts and produce comfortable and socially acceptable motions. Furthermore, in contrast with the threshold-based control approaches outlined in Section 2.1, parameters can be minimized over the course of motion rather than only when potentially dangerous motions approach safety thresholds.

A variety of planners and frameworks that consider human-based constraints on motion planning have been developed over the past decade. One motion planner by Kulić and Croft (2005), applied specifically to robotic manipulation during a handover task, minimized a danger criterion based on robot inertia and the distance between the human and the center of mass of each robot link. The safety framework by Kulić and Croft (2007) mentioned in Section 2.1 incorporates this danger criterion formulation and motion planner into a "long-term" component, which evaluates the safety of a proposed robot path prior to its execution. The pre-collision control methods described previously, on the other hand, are utilized within a "short-term" component, which considers safety methods that respond to imminent threats once the robot begins to carry out the planned path. By minimizing the inertia of the robot throughout the motion path, the planner ensures that the robot is already in a safe configuration in the event of an unanticipated collision.

This concept is quite similar to the idea behind some of the control-based safety systems described in Section 2.1 that attempt to keep parameters such as velocity or energy below predefined thresholds. The key difference, however, is that these systems also consider planning motions that minimize a parameter such as inertia, as opposed to imposing a limit through low-level control methods that react as the robot approaches or surpasses a threshold. In this way, the relevant safety parameters to be minimized can be kept as low as possible throughout robot motion — and, depending upon how conservatively the system defines the cost function, can remain far below predefined safety thresholds.

A different framework, developed by Sisbot and Alami (2012), focuses on designing not only safe but comfortable and socially acceptable robot motions. This is achieved by considering human kinematics, vision field, posture, and preferences, as well as the legibility of the robot's actions. In a handover task performed with this framework, the system chooses robot paths such that the human can easily grasp the object being handed over, while the robot both maintains a safe distance and moves in a manner that is visible to the human. The action is made legible by directing the robot's gaze toward the object being handed over at the appropriate time. This work, which focuses on close-proximity manipulation, is similar to a prior investigation of co-navigation conducted by the authors (Sisbot et al., 2007).

Another framework by Sisbot et al. (2010) combines various aspects of their prior work and incorporates considerations for making motion comfortable by limiting jerk and acceleration. Dehais et al. (2011) provided an experimental evaluation of the above motion planner using human subjects. Results from physiological sensing based on galvanic skin response, deltoid muscle activity, and ocular activity, combined with questionnaire responses, indicated significant differences between the considered motions. The evaluated motion types each involved different combinations of legibility, safety, and comfort based on robot speed and whether grasp detection and the Human-Aware Manipulation Planner developed by Sisbot et al. (2007) were used or not.

As indicated by questionnaire responses, participants preferred the condition that involved the human-aware planner and grasp detection. Due to the study design not being full factorial, the direct effects of the different motion components on physiological measures could not be determined, but the results did suggest that these measures could serve to discriminate between the motions. A motion planner developed by Mainprice et al. (2011) considers similar HRI constraints — specifically, human vision field, separation distance, and reachability — to drive a cost-based, random-sampling search in order to plan safe robot motions within cluttered environments. This planner incorporates the T-RRT algorithm and local path optimizations to generate paths.

Approaching the problem from a different perspective, Morales et al. (2015) evaluated safe motion planning for an autonomous vehicle with a human passenger. The authors developed the Human-Comfortable Path Planner (HCoPP), which takes human preferences into account, such as how far from a wall people prefer to travel when moving down a corridor, as well as visibility around corners when approaching a turn. The framework utilizes a three-layer cost map to integrate the constraints responsible for balancing between optimizing path length, as well as path comfort based on position and visibility. The authors assessed the effectiveness of the motion planner via a user study in which the HCoPP yielded significantly more pleasant and comfortable paths compared with a baseline motion planner, as assessed through questionnaire responses.

3.2 Geometric and Task-Based Constraints

In order for robot motion planning to be an effective method of improving safety in HRI, the motion planner must be capable of rapid replanning and of taking both geometric and task-based constraints into consideration. The necessity for rapid replanning is due to the inherent uncertainty resulting from the presence of human agents within a workspace. People not only display significant intra- and interpersonal variability and low repeatability of motions (Stergiou, 2004; Chaffin and Faraway, 2000), but can also unexpectedly change their preferences while performing tasks in terms of which actions are taken or the sequence in which actions are performed. Even the best motion and behavior models cannot account for all human-induced uncertainty, making it important for a robot to have the ability to replan motions quickly should the need arise.

While task models and representations may not be able to capture all potential task variations with perfect accuracy, combining task-based and geometric constraints in robot motion planning for HRI safety has several key benefits. Depending upon the level of integration, this method can also be viewed as the combination of task and motion planning into a single framework. While safety regions and buffers can be encoded with geometric constraints, the inclusion of task-based constraints allows for consideration of additional information that could serve to guide a robot, such as where a human is likely to move to or reach for based on prior actions. This is especially true in highly structured domains, where prior observations could be utilized to make predictions about future events.

Another key benefit of combining these two constraint types is that it allows for significantly faster computation through synergistic search-space pruning. In other words, by combining task-based and geometric constraints, it becomes possible to identify configurations that cannot be part of the solution and would not have been identified if task-based and geometric constraints were considered separately.

3.2.1 Constraint Fusion Techniques

Several different implementations of planners that utilize geometric and task-based constraints can be found in literature, and each implementation combines these constraint classes in different ways and employs various specific sub-planners to make use of the particular qualities of the chosen planning methods. Erdem et al. (2011), for example, introduced a framework that combines high-level, causality-based representations with low-level geometric reasoning. For causality-based reasoning, the framework utilizes the Causal Calculator (CCalc) (McCain, 1997) as its action domain description to encode effects and preconditions of actions, and the Virtual Reality Modeling Language (VRML) to encode a description of geometric models (Bell et al., 1995). Once the user defines the planning problem and the framework generates an initial plan, the low-level geometric reasoning is performed using a motion planner based on the Rapidly-Exploring Random Tree (RRT) algorithm. In this framework, selected geometric models are embedded directly into the high-level representation, allowing geometric constraints to guide the high-level planning process. Furthermore, if no kinematically feasible solution exists due to unmodeled geometric constraints, the motion planner can modify the high-level description in order to guide the search. In this manner, the geometric planner guides the causality-based planner at the representation level.

The method developed by Plaku and Hager (2010) also incorporates a combination of sampling-based motion planning and symbolic action planning, but integrates these by maintaining a single search tree that is iteratively extended until a dynamically feasible and collision-free path is identified. The algorithm utilizes a high-level specification based on the Stanford Research Institute Problem Solver (STRIPS) (Fikes and Nilsson, 1971) for its symbolic action planner, which guides the search of the motion planner to avoid focusing on regions unlikely to yield solutions by maintaining a heuristic-based estimate of the utility of exploring various actions. As the motion planner explores within the search space, it also provides feedback by updating action utilities. This interaction between the motion and action planners allows the search to progress quickly by focusing on search space regions that are more likely to yield solutions.

Cambon et al. (2009) combined task, motion, and manipulation planning through a process of defining and exploring the configuration spaces of robots and other objects within the environment. The authors also utilized a STRIPS-like formulation to evaluate symbolic and geometric constraints. In this method, identification and computation of constraints is performed incrementally, with the planner iterating between attempting to find a solution using the currently available knowledge and searching further within the different configuration spaces under consideration. One benefit of this approach is that it makes no assumptions about what methods are used to explore the symbolic and configuration spaces or how the geometry is represented, resulting in flexible implementation.

3.2.2 Hierarchical Encoding Methods

As mentioned previously, rapid computation and replanning are imperative when applying motion planners as a method for ensuring safe HRI, due to the uncertainty introduced by the variability inherent in human motion and preferences. Reduced computation time and quicker replanning in combined task and motion planning can be achieved by utilizing hierarchical encoding schemes.

Work by Wolfe et al. (2010) incorporated vertically integrated hierarchical task networks within a combined action and motion planner. In this framework, external solvers generate primitive actions, such as arm and base movements, at the bottom of the hierarchy. The authors further improved the speed of the developed system through implementation of the State-Abstracted Hierarchical Task Network (SAHTN) algorithm, which receives and reuses information about the relevance of state variables to particular subtasks.

Kaelbling and Lozano-Perez (2011) also incorporated a hierarchical representation, but with a focus on short-term planning. This can be a particularly useful technique for HRI, as the nondeterministic nature of many real-world environments and applications can render long-term plans invalid before they are fully executed. The presented motion and task planning method, given appropriate domain-dependent choices when developing its structure, allows for the pruning of large portions of the search space, as well as rapid computation.

3.3 Summary

In this section of the monograph, we discussed the manner in which motion planning can serve as an effective tool for ensuring safe HRI by encoding both physical and psychological safety into the motion planners' cost functions. This technique allows for a more proactive approach to ensuring safety compared with the control-based safety methods discussed in Section 2. Importantly, motion planners enhanced with consideration for safety can be used for both manipulation and navigation planning and can be applied to virtually any robotic platform, indicating the versatility of this method of ensuring safety.

Due to the dynamic nature of any environment occupied by people, however, such planners must be able to rapidly recompute new paths and motions. As described in the preceding sections, this is often achieved by combining task and motion planning to aid in efficient traversal of a complex search space, and by utilizing hierarchical encodings of the constraints.

There are practical limits, however, to constantly replanning based solely on the current configuration of a rapidly changing world state. Consequently, it is beneficial to reason not only on the current state, but also on predictions of future tasks and motions, which is the topic of the following section.

4

Safety Through Prediction

In some HRI situations, it is reasonable to assume that the environment is quasi-static and simply rely upon replanning motions quickly when the movement of human and robotic agents conflicts with the initial plan. However, this approach is not appropriate for maintaining safety within more dynamic environments. In this context, motion plans based on a quasi-static assumption quickly become obsolete, making reliance on replanning impractical — particularly if humans and robots are working in close proximity to one another, as there may not be sufficient time to replan. As a result, the ability to anticipate the actions and movements of members of a human-robot team is critically important for providing safety within dynamic HRI environments. Furthermore, this ability must extend to all team members, with both robots and humans able to predict one another's actions and motions.

4.1 Human Activity Prediction

By predicting which action a person might take next, robot motion and activity planners such as those described in Section 3 can identify actions and paths that will result in safe and efficient interaction.

This selection process can involve adjusting the action the robot will take, the timing of when that action should be performed, or the motions the robot will perform in order to achieve the action. Work in this field has incorporated a wide variety of methods, with researchers often formulating probabilistic models and frameworks according to which prediction could be made. Among these approaches, some reason directly on low-level features derived from cameras, depth sensors, and motion capture systems, while others reason on abstract representations of actions or task steps in order to predict future activities. Beyond predicting the next human action, other work in this field focuses on predicting the timing of the actions, which allows a robot to not only decide what action to take in order to maintain safety, but also when it would be best to take that action. These topics are summarized in Figure 4.1.

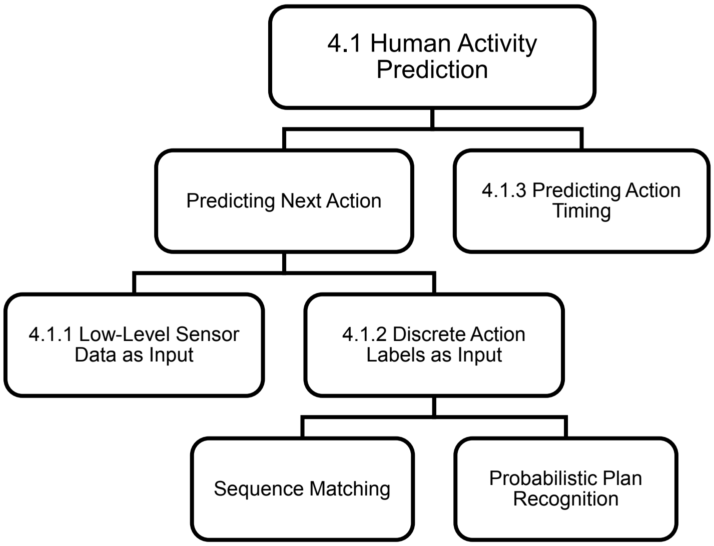
4.1 Human Activity Prediction
Predicting Next Action
4.1.3 Predicting Action Timing
4.1.1 Low-Level Sensor Data as Input
4.1.2 Discrete Action Labels as Input
Sequence Matching
Probabilistic Plan Recognition

Figure 4.1: Diagram depicting the methods for human activity prediction discussed in Section 4.1.

4.1.1 Low-Level Sensor Data as Input

Work in the field of activity prediction that utilizes low-level representations of sensor data as input has included a wide spectrum of prediction techniques, ranging from detection of activity processes (such as walking or pouring a liquid) to action goals (such as what item a person is about to grasp). With regard to activity processes, Ryoo (2011) developed an early prediction approach that utilizes an integral histogram of spatio-temporal features derived from RGB videos. The integral histogram, an encoding of how histogram distributions change over time, is constructed from histograms of *visual words* present in successive frames of the video input. These visual words are determined by analyzing points of interest in the frame where salient motion is occurring and computing a description of these points by summarizing gradients in the corresponding portion of the frame as time progresses. The method then uses a clustering algorithm to detect groups of similar features.

In an approach that utilizes higher-level structure and relationships between the current pose and the surroundings of the human, Koppula and Saxena (2013b) incorporate RGB-D input and object affordances to make predictions about future human actions. In their work, a conditional random field (CRF), which contains nodes for sub-activities and objects and edges defining spatio-temporal relations, is augmented with anticipated temporal segments to form an anticipatory temporal conditional random field (ATCRF). The system then calculates a distribution over possible futures defined by many ATCRFs and uses it for prediction. Related work by (Koppula and Saxena, 2013a) uses a similar conditional random field formulation, but rather than only projecting possible temporal graph segments into the future, the proposed method reasons on possible graph structures for the past as well. The system samples some spatio-temporal structures close to the ground truth, then makes randomized split and merge moves to explore the segmentation space. A spatio-temporal structure is not required a priori, as the algorithm considers multiple possible segmentations to be used for prediction.

Another method by Azorin-Lopez et al. (2014), which uses image data for activity prediction and detection, eliminates the need for information about timing or the sequence of events by representing interaction data from RGB input with a normalized activity description vector (ADV). This approach splits the image into a grid and calculates data on movements within each cell of the grid. The compact ADV representation can then be used with various standard classification methods.

As it is not always possible to have an external camera viewing the scene, a unique approach by Ryoo et al. (2015) includes robot-centric, first-person video with possible ego-motion as input for predicting human activities. This method, similar to the work by Ryoo, utilizes visual features derived from RGB video to form integral histograms that represent activities. This method focuses on early prediction of activities through detection of onset activities: subtle sub-activities that occur just prior to the main activity being predicted. This process involves learning onset signatures, which are characterizations of onset activities, and then utilizing these signatures and prior event history to classify actions early in their execution.

While the ability to predict activity processes can provide task-level information to task and motion planners, in the context of ensuring safety in HRI, the ability to predict action goals (such as where a person might reach toward) is also very important. This is especially true during close-proximity collaboration, where simply slowing down and stopping the robot via control-based methods (as discussed in Section 2) can lead to constant motion conflict and many stressful near-collision situations. Mainprice and Berenson (2013) developed a framework that utilizes labeled demonstrations of reaching motions to generate models for prediction of workspace occupancy. In this framework, separate Gaussian mixture models (GMMs) are trained for each goal position for a particular task and Gaussian mixture regressions (GMRs) are used to generate representative reaching motions. Then, based on observation of the initial segment of a new reaching motion and the computed GMMs and GMRs, the framework calculates the likelihood of occupancy of each voxel within a simulated shared workspace. The robot then selects actions and paths that minimize incursion into the regions of the workspace expected to be occupied by humans.

Recent work by Pérez-D'Arpino and Shah (2015) also focused on predicting reaching locations based on human demonstrations, but with a time-series analysis that utilizes multivariate Gaussian distributions over the tracked degrees of freedom of the human arm defined for each time step of the motion. The system uses the learned models to perform Bayesian classification on the initial stages of motion in order to predict where a person will reach toward and to select robot actions that minimize interference. In contrast to the GMM formulation used by Mainprice and Berenson, the models take the sequence of points along the motion trajectory into consideration, allowing for better discriminability and higher classification confidence very early in the process of the human's motion. PÃľrez-D'Arpino and Shah also explore how task-level prediction, incorporated into the framework as a prior in the Bayesian formulation, could affect prediction results as a function of the task-level prediction's accuracy.

4.1.2 Discrete Action Labels as Input

Rather than reasoning on low-level sensor data directly, a second class of human action prediction techniques utilizes task models to reason on what actions have been taken, and then uses this data to inform prediction. In these works, the purpose of the sensing platform is action detection.

Dominey et al. (2008) presented a straightforward method of reasoning on actions performed, which incorporated an *interaction history* to facilitate anticipatory robot behavior. The system, deployed for a collaborative assembly task, compares current action sequences to previously observed sequences in order to determine whether the current interaction is an instance of a previously observed task sequence. As the robot correctly anticipates specific tasks multiple times, the system adjusts a confidence parameter that modifies the anticipatory behavior. When the robot anticipates an action for the first time, the system predicts what the user will say when requesting the robot's help. Eventually, as the system builds confidence in its prediction, the robot begins to take initiative and perform the predicted action without having received an explicit request to do so.

Other works have considered probabilistic formulations for human action prediction. A framework developed by Hoffman and Breazeal (2007) utilizes a cost-based Markov process to anticipate human actions and select actions based on the robot's confidence in the validity of the prediction and risk. In this context, a "risky" robot action is one that could result in a significant waste of time in the event that it was not the proper action to take under the given circumstances. Results from a comparison of the proposed framework with a reactive agent in a simulated HRI scenario indicated that anticipation improved best-case task efficiency based on metrics such as human idle time and the amount of concurrent motion. Furthermore, anticipatory actions yielded qualitative improvements in the degree of human satisfaction with the robot as a teammate and perception of the robot's contribution to the team's success.

Another method of encoding a human-robot collaborative task with a probabilistic framework and utilizing the results to anticipate human actions was explored by Nikolaidis et al. (2013). In this work, the collaborative task was encoded as a Markov decision process (MDP), and results from human subject experiments indicated that observations of changes to the entropy rate of the derived Markov chain could be utilized to encode the uncertainty of the robot about what action a human team member will perform next.

An extension of the MDP, the mixed observability Markov decision process (MOMDP), was later utilized by Nikolaidis et al. (2015). In this work, the system learns user models automatically from joint-action demonstrations by clustering action sequences and learning reward functions for each cluster through inverse reinforcement learning. The robot then uses these models to predict user types in order to then predict users' actions and execute appropriate anticipatory actions. Results from this work indicated superior performance using the model compared with manually controlling the robot via voice commands, both with regard to quantitative metrics such as task execution time and human idle time and subjective measures evaluated through questionnaire responses.

While a majority of the previously mentioned works focus on short-term prediction of brief actions, Li and Fu (2014) developed a framework for predicting actions of longer duration composed of constituent actions with complex temporal compositions. In this framework, action causality is addressed via implementation of a probabilistic suffix tree, a variable order Markov model formulated to represent various order Markov dependencies between constituent actions. The authors also developed a method of extracting context cues derived from sequential pattern mining through analysis of object-action co-occurrences. This framework also incorporates a predictive accumulative function (PAF), which determines the predictability of activities by learning from data.

4.1.3 Predicting Action Timing

The timing of an action is another important aspect of human action anticipation. Work by Hawkins et al. (2013) utilized a probabilistic graphical model in which the start and end time of actions are treated as probabilistic variables conditioned on earlier events. This framework also considers the uncertainty in sensing executed actions and variability in timing among individual humans when making its predictions. The described system was implemented during an assembly task in which the robot attempted to minimize idle time by anticipating human actions and providing correct components at the most optimal time.

The authors later extended this work (Hawkins et al., 2014) to consider not only uncertainty in timing and action detection, but also more complex task descriptions containing ambiguous task ordering. Similar to their work mentioned above, this framework incorporated a probabilistic graphical model and minimized a cost function based on idle time. The results indicated that proper adjustment of parameters describing confidence in the action detectors improved team performance, as excessive confidence in the action detection caused the robot to take potentially costly actions. By utilizing the proposed framework with properly adjusted confidence parameters, the authors were able to show that the system could achieve robust anticipation even with noisy action detection.

While predicting the timing of actions can improve task performance, findings by Huang et al. (2015) indicated that this improvement can come at a cost to user experience. In their experiment, the researchers evaluated the effects of various handover strategies during a human-robot collaboration task. By observing human teams, the researchers derived techniques intended to better synchronize hand-overs during the task based on the receiving person's task demands and current state. They found that the adaptive method, which incorporates prediction of the user's state and the above-mentioned techniques, resulted in a balance between user experience and performance compared with proactive and reactive baselines, which yielded poorer user experience and performance, respectively.

In Section 4.1, we discussed how low-level sensor input can be utilized to reason on changes in sequences of image frames, relationships between objects in a given scene, and human trajectories in order to predict both activity processes and action goals. Additionally, human actions can also be predicted by making use of the Markovian relationship between constituent actions of a sequence, and utilizing formulations such as MDPs or MOMDPs. Finally, we highlighted that probabilistic methods for predicting action timing can aid in maintaining safe human-robot collaboration, but that practitioners should carefully consider the way in which integrating such predictions can affect user experience.

4.2 Human Motion Prediction

Although the ability to predict human actions can be useful for generating safe robot motions and action plans, knowledge about what activity will be performed or the end location a person is reaching or walking toward does not provide information about which specific portion of a shared human-robot workspace the human will occupy during the execution of that predicted action. This additional information could be leveraged to ensure safe robot motion by enabling the robot to reason not only on the expected start and end locations, but on the entire expected human motion.

The basis of human motion prediction can be divided into two distinct categories: goal intent and motion characteristics. For the former, action prediction can often serve to inform motion-level prediction by inferring humans' goals, which, in combination with an appropriate motion model, can be used to anticipate how a human will move as he or she walks or reaches toward the predicted goal. In the latter category, motion prediction is not linked to predicted goals or actions, but instead utilizes techniques such as analysis of biomechanical predictors or reasoning on features of natural motion. These methodologies for human motion prediction are summarized in Figure 4.2.

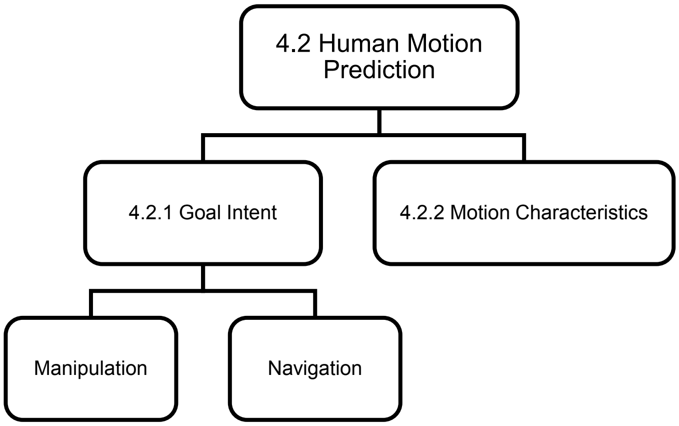

Figure 4.2: Diagram depicting the methods for human motion prediction discussed in Section 4.2.

4.2.1 Goal Intent

When predicting human motion based on goal intent, the system infers a human's goal and predicts the path or trajectory he or she is likely to take in order to reach that goal. This can include both smaller manipulation movements and larger movements involving ambulatory motion.

As goal prediction is necessary for this approach, several of the action prediction methods described in Section 4.1 could be used for this purpose. Some of the approaches mentioned earlier also predict the motion people could take while performing the predicted manipulation actions. In the paper by Mainprice and Berenson (2013), the system uses regressed motions derived via the GMR to compute swept volumes that define human workspace occupancy during execution of the predicted reach. In the work by Pérez-D'Arpino and Shah (2015), on the other hand, prediction during a reaching motion is calculated using multivariate Gaussian representations. In contrast to computing a swept volume of the entire reaching motion, the models computed in this work describe the mean and variance in the human's hand position for each of the possible goal positions at each time step, yielding a prediction of human position as a function of time during the reach.

In the previously mentioned work by Koppula and Saxena (2013b), once the system predicts an action, it utilizes Bézier curves to define potential trajectories of the human hand while performing the action. An extension of this work by Jiang and Saxena (2014) allows for more-detailed prediction of human motion during action execution. In this work, a low-dimensional representation of a high-dimensional model of human motion is computed through a Gaussian process. The low-dimensional description of the motion is then incorporated as latent nodes in a CRF representation to form a model called a Gaussian process latent conditional random field. By learning a two-way mapping between the high- and low-level representations, this approach allows for computationally tractable prediction of high-dimensional motion while maintaining the ability to reason on relationships between people and objects in the given scene. In addition to being able to predict human motions with higher fidelity, the compact representation of motion allows for reasoning on the physical plausibility of actions, thus improving prediction.

Another manipulation motion prediction technique, developed by Mainprice et al. (2015), was specifically designed for reaching motions performed during known, collaborative tasks based on inverse optimal control. In this work, example data of two people performing a co-manipulation task were collected via a motion capture system. The gathered trajectories, along with feature functions encoding smoothness and distance relationships, were then used as inputs for the path integral inverse reinforcement learning (PIIRL) algorithm in order to produce the reaching motion cost functions. The approach uses these cost functions and a human kinematic model with 23 degrees of freedom to predict the human's motion by iteratively replanning the motion with the stochastic trajectory optimization for motion planning (STOMP) algorithm. The authors found that the PIIRL algorithm is capable of correctly recovering the cost functions for sampled motions created by manually setting weights. Furthermore, the developed framework can predict human motion more effectively than hand-tuned cost functions.

Goal-based human motion prediction can also be applied to larger movements involving ambulatory motion, such as walking. Elfring et al. (2014), for example, developed a two-step approach incorporating growing hidden Markov models (GHMMs) and the social forces method. During the learning step of this approach, the system uses GHMMs to learn typical human walking patterns from collected data, allowing for continuous updating of the model as new data are collected. In the second step, the system combines goals inferred from partial trajectories using the learned GHMMs with a motion model based on the concept of social forces from work by Luber et al. (2010) in order to predict the path a human will take toward their goal based on static obstacles, other people, and the physical constraints of the environment.

An earlier approach by Bennewitz et al. (2005) uses no explicit motion models; instead, learning is performed automatically by clustering human motions via the expectation maximization (EM) algorithm and modeling these clustered motions using hidden Markov models (HMMs). In this approach, natural *resting places*, or places where the human's motion slows or pauses, are identified from the data and used as potential goal locations.

Ziebart et al. (2009), on the other hand, developed a goal-based human walking trajectory prediction method that leverages the assumption that people move efficiently when navigating a space by modeling human motions using maximum entropy inverse optimal control. One key benefit of this approach is that the cost function it learns is a linear combination of features based on environmental obstacles, allowing it to generalize well in the event that objects within the environment are moved to new locations or removed from the scene altogether. Furthermore, the presented approach uses an incremental planner that enables real-time deployment by developing cost maps without taking prediction of motion into account, and then iteratively plans a robot trajectory while simulating human motion.

4.2.2 Motion Characteristics

Some human motion prediction methods do not rely upon estimates of goal locations, but make use of observations about how people move and plan natural paths. This class of techniques encompasses a variety of approaches, including discovery of likely motion progressions, use of biomechanical predictors, consideration of features of natural motion, and general unsupervised approaches for learning about how agents move within an environment.

Takano et al. (2011) use motion capture data to encode skeletal motion patterns as HMMs, which are then grouped via Ward clustering to form a hierarchical structure called the “motion symbol tree.” The system then learns sequences of motion symbols through the use of Ngrams, forming a directed graph — the “motion symbol graph” — that represents transitions between motion patterns, and thus causality among human behaviors. The motion symbol graph is then used in conjunction with current motion observations to predict future motion patterns, represented as skeletal motions.

Xiao et al. (2015) used previously observed human trajectories to train an SVM classifier that first decomposes the data into high-level classes, such as *wandering* or *stopping*. The framework then forms clusters within these classes using the partitioning around medoids (PAM) algorithm, with a modified distance function that allows for better clustering of similar, non-overlapping trajectories, especially with limited movement. These clusters are then used to extract prototypes, which are matched to observed partial trajectories to enable prediction. As the clustering is performed in an unsupervised fashion, one key benefit of this method is the ability to utilize new trajectories to further refine and adapt prediction.

An alternate approach by Unhelkar et al. (2015) utilizes biomechanical turn indicators for motion prediction. In this work, the authors used a motion capture system to collect data of people walking toward targets in a room. Results from statistical analysis showed that indicators such as head orientation and body velocity normalized by height can signal a human's intention to turn prior to physical performance of the turn itself. This allows for prediction of turning motions without necessarily predicting a goal location. The authors also applied these turn indicators within a goal-based prediction framework based on the previously mentioned work by Pérez-D'Arpino and Shah (2015) to show that these indicators provide a signal strong enough to be used for motion prediction. It is important to note that the biomechanical turn indicators validated in this study can be incorporated into other prediction frameworks as well.

The final step taken by the authors was to test the utility of the prediction in a dynamic environment via a closed-loop simulation in which a robot planned its motions according to predictions about human motion. They found that by reasoning on the predicted human motion, the robot was able to take paths that avoided motion conflicts; the resulting robot path plan would allow for safer and more comfortable co-navigation of the space.

While the above works consider the motion of the human independently of other agents in the environment, one interesting and useful insight is that the trajectory of the robot affects that of the human. Based on this concept, Kuderer et al. (2012) used demonstrations of human motion to develop a method of motion prediction that learns joint trajectories. By analyzing trajectories during human-human interaction, the researchers investigated which features of walking motion the system could use to learn how to characterize and predict typical human walking behavior. The model, based on the principle of maximum entropy, considers features including the amount of time needed to reach a goal, acceleration profiles, walking velocity, and collision avoidance behavior. This approach, based on the physical aspects of the trajectory (as opposed to a Markov model), considers how people reason over these features when planning joint trajectories. (Humans might, for example, prefer to maintain a steady walking speed, try to reach their goals quickly, or attempt to minimize proximity to obstacles and other agents.) The developed method was validated both in simulation and in a physical environment, indicating that human trajectory prediction based on consideration of joint trajectories can successfully guide a robot in a natural and socially compliant way.

Interaction with other agents in the environment is not the only context of a scene that can be utilized for prediction, however. Within the computer vision community, a method developed by Walker et al. (2014) utilizes an unsupervised, data-driven approach for visual prediction using a static image. In this work, the system learns relationships between subsections of the image to form a context-based Markov model, such that no prior assumptions about agents and activities are necessary. This work provides a visual prediction of the future appearance of a scene by projecting the position and appearance of subsections in that scene. While this approach was not specifically designed for human motion prediction, the unsupervised, context-based method developed here could easily be applied to this domain, as human movement within an environment is also influenced by the context of the scene.

While prediction of human actions can be useful for maintaining safety in HRI, in Section 4.2, we discussed how predicting likely human motions yields additional information that robot path and motion planners can leverage. In structured environments, robots can utilize prediction of likely goals in combination with various motion models to predict the path a human will take toward that goal. In applications lacking clear goal locations, or for which action prediction is difficult or impractical, prediction based on motion characteristics is appropriate, with relevant characteristics including biomechanical indicators of motion, features of natural motion, or patterns of motion within a scene.

4.3 Prediction of Robot Motions and Actions

Human-robot collaboration is a two-way interaction; as such, human agents' ability to predict the actions and movements of a robot is as essential as the ability of a robot to predict the behavior of humans. By taking the predictability of robotic teammates into account, researchers can develop both implicit and explicit methods of conveying robot intent, allowing human teammates to accurately predict the behavior of their robotic coworkers. These topics are summarized in Figure 4.3.

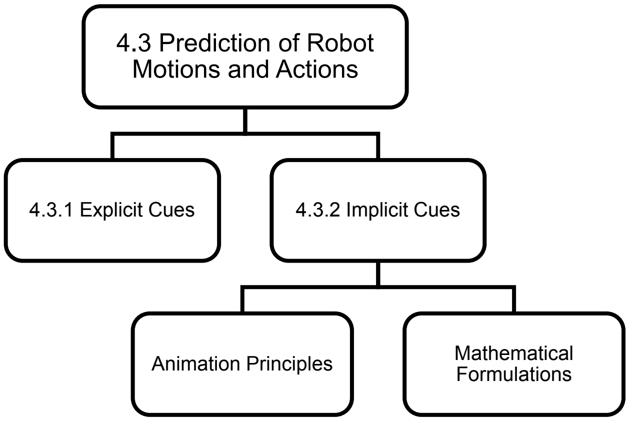
4.3 Prediction of Robot Motions and Actions
4.3.1 Explicit Cues
4.3.2 Implicit Cues
Animation Principles
Mathematical Formulations

Figure 4.3: Diagram depicting the types of cues that can aid in prediction of robot motion, as discussed in Section 4.3.

4.3.1 Explicit Cues

One method of making robot behavior more predictable to human teammates is explicit communication of intent, where the robot directly communicates its planned actions and motions through visual and auditory cues. In work by St. Clair and Mataric (2015), for example, verbal feedback from the robot was provided while coordinating a joint activity between a human and robot. Team coordination via robot task control and speech feedback was formulated as a planning problem, with the task environment represented as a Markov decision process. In this approach, the system observes a human's actions, infers his or her strategy, and reasons about the compatibility of possible robot actions with this strategy. The robot then provides verbal feedback to the human about the actions it will take (self-narrative feedback) and which actions the human should take (role-allocative feedback), and provides commentary on success and failure during task execution (empathetic feedback). In the proof-of-concept joint task, utilizing verbal communication to convey robot intent in addition to other forms of feedback led to faster task execution and more positive ratings of the robot as a teammate.

In addition to speech, Lallee et al. (2013) also studied the effects of communicating joint plans via the robot's gaze. In this work, a human and robot performed a joint task in which they moved a box to uncover an object, and then retrieved the object. The study analyzed the effects of a joint plan, directed gaze, and speech by assessing how the presence or absence of these elements influenced proper turn taking, motion conflicts, and the number of errors occurring during task execution. The authors found that use of a joint plan, gaze, and speech resulted in the best team performance; that directed gaze worked well in the presence of a joint plan; and that speech functioned well even without the use of directed gaze.

As it may not always be practical to use speech or gaze direction to convey robot intent, researchers have developed other methods of making robot motion more predictable. For example, Szafir et al. (2015) investigated the use of an LED array to communicate the motion of flying robots to humans within a shared workspace. After investigating and applying proper constraints on flying robot motion (such as maintaining constant altitude unless taking off or landing) based on human expectations, this study specifically evaluated the speed and accuracy of human prediction about the motion of a small robotic aerial vehicle using four distinct settings for the circular LED array.

The authors found that participants rated the robot as a better potential work partner when it incorporated any of the assessed LED-based communication settings, compared with the absence of any signaling method. Also, three of the four evaluated communication modes yielded better performance according to quantitative metrics of prediction response time and accuracy. The results also indicated that design trade-offs between occlusion, precision and generalizability must be taken into consideration when selecting a communication mode.

The projection system by Vogel et al. (2013) described in Section 2.1 offers another method of explicit visual feedback for robot motion prediction. By visually projecting the safety zone used to stop the robot and adjusting its extents via a commanded robot velocity, the projected shape of the safety zone could provide users with an explicit cue as to the direction in which the robot will increase its velocity. While this aspect of the system was not specifically studied in the paper by Vogel et al., it is plausible based on the results from preceding studies that the predictability of robot motion would increase with the use of the projection system.

4.3.2 Implicit Cues

While direct, explicit communication can be an effective method of conveying robot intent, implicit cues can also communicate future robot motions and actions. Using our understanding of human-human teaming as a basis, the benefits of implicit communication of intent are readily apparent, as one would not expect a coworker to always explicitly communicate where they are about to reach or move toward. Instead, people convey intent through subtle cues embedded in the ways they perform their motions.

Applying this concept to HRI, Takayama et al. (2011) investigated the use of animation principles to aid in the *readability* of robot intent. In this study, participants observed simulated robot scenarios and were asked to interpret the actions taken by the robot. The authors found that use of *forethought* — expressive movements, such as the robot changing its height to acknowledge the human, or directing its gaze at the human and then at an object about to be manipulated — resulted in a greater degree of confidence in participants' interpretation of the robot's actions. Forethought also led to participants rating the robot as more approachable and appealing. Participants also considered the robot to be smarter when it exhibited a reaction to the success or failure of the task it was performing.

While these animation principles improved the interpretation of activities upon completion in the study by Takayama et al., Szafir et al. (2014) demonstrated that animation principles can also improve the predictability of robot movement. In particular, Szafir et al. focused on implicit communication of motion intent for assistive small robotic aerial vehicles through animation-inspired modification of motion primitives. These motion primitives, such as *hover* or *cruise*, were modified via three manipulations: *arcing* (moving in arcs instead of straight lines), *anticipation* (moving slightly in reverse before moving forward), and *easing* (gradually slowing down from or accelerating to full speed).

The effectiveness of these motion primitive modifications was evaluated in two experiments: In the first, participants observed animations of virtual robots and were asked to anticipate the robots' motion intentions. The response time and accuracy of these predictions was measured to determine which combinations of motion primitive modifications resulted in optimal predictability. The researchers found that a combination of easing and anticipation, as defined above, demonstrated the best potential for increasing prediction accuracy, but that anticipation increased trajectory length and yielded slower response times due to the slight reverse movement, highlighting the trade-off between response time and accuracy when using this motion manipulation technique. In the second study, Szafir et al applied motion manipulations to actual robotic aerial vehicles, and evaluated the effects of these manipulations on high-level interactions. They found that, based on questionnaire responses, participants rated the robotic aerial vehicle higher in terms of usability, safety, and naturalness when it utilized the motion manipulations.

Dragan et al. (2013) also investigated how robot motion could implicitly convey intent, but approached this problem through mathematical definitions of predictability and legibility of robot motion. Their paper discussed how *legible* (intent-expressive) and *predictable* (expected) motions can differ substantially — and are, in fact, often contradictory. To evaluate the validity of the developed mathematical formulations, the authors performed experiments in which participants predicted what goal a robot or person was reaching toward when using predictable or legible motions. The predictability of the motions was evaluated via questionnaires and evaluation of participants' drawings of expected robot paths, while motion legibility was evaluated by measuring the length of time it took participants to feel confident in their predictions of which goal the robot was reaching toward. In general, trajectories that scored high based on the mathematical representation of legibility were more legible, but that trajectories with a high mathematical predictability were not always rated as more predictable by the participants. Dragan et al. attributed the latter to the fact that participants had a wide variety of expectations of how the robotic arm would move toward its goal, highlighting the importance of factoring human expectations into robot motion planning.

In later work, Dragan et al. (2015) analyzed the effects of legible and predictable motion on the quality of human-robot collaboration. In this study, participants were required to infer the intent of a robot in order to efficiently complete a collaborative task. The researchers compared three types of robot motion: *predictable*, *legible*, and *functional* (motion that reaches the goal and avoids obstacles without taking predictability or legibility into consideration). The results indicated that functional motion alone was not sufficient for fluent HRI, and that legible motion, while less efficient than functional motion in terms of trajectory length, led to more efficient interaction as measured by coordination time. Participants also reported via questionnaire responses that they preferred legible motion to functional motion.

Overall, results from the works discussed in Section 4.3 have indicated that both explicit and implicit cues can effectively convey a robot's intent. Techniques in the former category rely upon modalities such as speech, gaze direction, or light signaling to make the robot's intent known directly. Works within the latter category, on the other hand, attempt to make the robot predictable through more implicit means — either by modifying motions according to animation principles or mathematically formulating motion planning to optimize for legibility. If a robot utilizes the presented methods, making its behavior easier to predict, humans working with that robot can more effectively plan their own motions so that they remain safer.

4.4 Summary

In this part of the monograph, we demonstrated the utility of prediction as a method for safe human-robot interaction. We began by discussing works focused on predicting the most likely human actions and their timing, which can serve as inputs to robot path and motion planners such as those described in Section 3. The subsequent section discussed methods for predicting the actual path a person will take based on both the predicted goals of that person and characteristics of natural human motion.

By actively predicting the human's actions and motions, a robot can build upon the motion planning approach depicted in Section 3 and produce safe motions proactively, instead of relying on frequent replanning. As discussed earlier, this is especially useful for highly dynamic settings in which simply replanning each time the environment changes is impractical. This can be both due to the inability to produce new plans quickly enough, as well as paths based on constant replanning being unsatisfactory (in terms of both task efficiency and safety) given the potential for getting stuck in local minima. By incorporating prediction knowledge and planning in a proactive manner, it becomes possible to avoid these downfalls.

As we noted in the previous section, facilitating human prediction of the robot's behavior can also improve safety in HRI. Through the use of either explicit or implicit cues, the robot can make its intended goals and motions clearer to co-located humans, which in turn facilitates humans' ability to select actions and motions that maintain safety.

While prediction has been shown to be useful for ensuring safe HRI, it is important to note that the efficacy of this approach is directly related to the accuracy of the relevant predictors. (In the previously mentioned work by Pérez-D'Arpino and Shah (2015), for example, the authors showed that incorporating a low-confidence task-level predictor into motion-level prediction can actually deteriorate prediction performance.) Consequently, it is imperative that a practitioner choose an appropriate predictor for a given task and environment, and that he or she evaluate the efficacy of that predictor. Furthermore, a practitioner can also utilize control-based safety methods, such as those highlighted in Section 2, as a safeguard against incorrect predictions.

5

Safety Through Consideration of Psychological Factors

The maintenance of *physical safety* often dominates discussion of safety in HRI; however, ensuring *psychological safety* is also of critical importance, as we discussed in Section 1.1. Maintaining psychological safety involves ensuring that the human perceives interaction with the robot as safe, and that interaction does not lead to any psychological discomfort or stress as a result of the robot's motion, appearance, embodiment, gaze, speech, posture, social conduct, or any other attribute. Results from prior experiments have indicated that maintaining physical safety by simply preventing collisions as they are about to occur can lead to low levels of perceived safety and comfort among humans (Lasota and Shah, 2015). Therefore, maintenance of physical safety alone cannot ensure safe HRI.

One of the primary methods of ensuring psychological safety during human-robot interaction is appropriate adjustment of robot behavior. Such behavioral adjustments can be split into two categories: those based on robot features and those based on social considerations. Work within the former category involves adjustment of various parameters of the robot's motion, such as speed, acceleration, or proximity to the human, and also investigates how to properly adjust behavior based on the robot's appearance. Work in the latter category, on the other hand, is concerned with discovering which social rules observed in human-human interaction are important to follow during HRI, and the impact of factors such as culture and personality traits.

Although adapting robot behavior is important for maintaining psychological safety, it is also necessary to evaluate the effectiveness of these adjustments in a principled way. Toward this objective, researchers have developed three tools for assessing psychological safety: questionnaires, physiological metrics, and behavioral metrics. Each of these assessment methods possesses its own benefits and drawbacks, making it imperative to understand which approach should be used under what conditions.

All of these topics, including robot behavior adaptation and assessment, are discussed in this section of the monograph, as depicted in Figure 5.1.

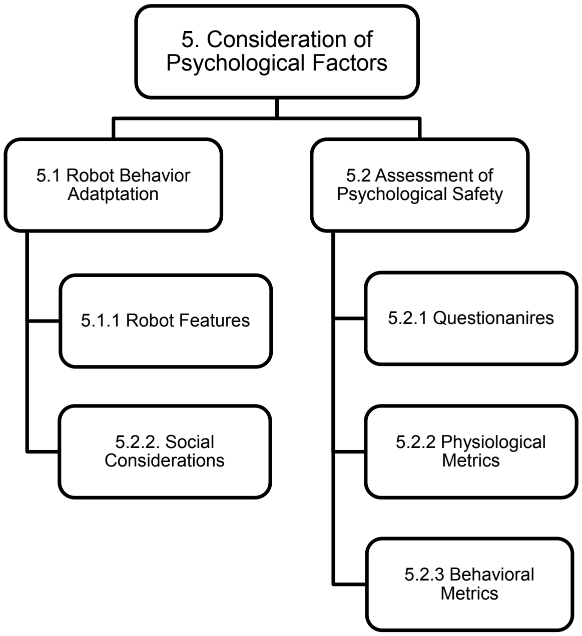
5. Consideration of Psychological Factors
5.1 Robot Behavior Adatptation
5.2 Assessment of Psychological Safety
5.1.1 Robot Features
5.2.2. Social Considerations
5.2.1 Questionanires
5.2.2 Physiological Metrics
5.2.3 Behavioral Metrics

Figure 5.1: Diagram depicting methods of assessing and potential factors influencing psychological safety, as discussed in Section 5.

5.1 Robot Behavior Adaptation

Although perceived safety and comfort were not the primary objectives of all the works mentioned in the preceding sections, a significant number take psychological safety into consideration by adjusting robot behavior specifically to make interaction more comfortable for humans. For example, the trajectory-forming strategy developed by Broquere et al. (2008) mentioned in Section 2.1 limits the jerk, acceleration, and velocity of the robot. The authors of that work noted that maintaining comfortable interaction is one of the major constraints on controllers developed for HRI.

In the work by Sisbot et al. (2010) outlined in Section 3.1, the authors combined several methods to support comfortable HRI and noted that physical safety alone is not sufficient for acceptable HRI, adding that the robot must also avoid any actions that might induce fear, surprise, or discomfort among humans. Authors expressed a similar sentiment in another paper focused on developing a safe and comfortable manipulation planner (Sisbot and Alami, 2012).

The motion planner developed by Mainprice et al. (2011), as outlined in Section 3.1, incorporates HRI constraints to generate motion that result in comfortable reactionary human motions, but focuses specifically on the physical comfort of the interaction — for example, how far the human must travel from a resting position to a proposed reach location, or how close he or she would be to exceeding biomechanical limits were they to do so.

While the aforementioned authors took psychological safety into account within their work, there also exists a body of research with the primary focus of evaluating how robot behaviors affect psychological safety factors, both in terms of studying and adjusting robot features, as well as understanding the impact of social factors.

5.1.1 Robot Features

Researchers across various fields of study have shared the concerns and sentiments regarding psychological safety expressed in the above works, leading to the study of the possible negative psychological effects of numerous aspects of robot behavior. Arai et al. (2010), for example, assessed the influence of robot speed, separation distance, and advance notice of motion on human operator stress. Through results obtained from physiological sensing and questionnaires, the authors found that all three of the above factors affected operator stress: specifically, operators exhibited more stress with shorter separation distances (1.0m) compared with longer distances (2.0m), faster robot speed (1000mm/s) compared with slower speeds (500mm/s and 250mm/s) and in the absence of advance notice of motion compared with receiving notice from the robot prior to movement.

Koay et al. (2006) investigated the influence of similar robot features, but focused on real-time reporting of discomfort via a hand-held device. Furthermore, the study evaluated human discomfort in live, unconstrained scenarios, rather than predefined interactions. Separation distance once again played a factor, with a majority of participants reporting discomfort at distances of less than 3m — with the greatest level of discomfort within 1-3 m. Participants also exhibited discomfort when the robot blocked their path or was on a collision course with them, particularly if the distance of separation was less than 3m.

Kulić and Croft (2006) also tracked participants' psychological state in real time in their work, but with the use of physiological sensing and a fuzzy inference engine. They compared two types of motion planners: a simple potential field planner and a safe planner that included a danger criterion, and performed assessments via physiological measurements and questionnaire responses. The authors found that robot speed had a significant effect on participants' levels of surprise, anxiety, and calm, and that use of the safe motion planner resulted in lower levels of anxiety and surprise and a greater degree of calm among participants.

In addition to evaluating the effects of robot speed and distance, Butler and Agah (2001) used questionnaires to study the effects of robot body design on participants' levels of perceived comfort. The authors evaluated these factors during execution of several robot behaviors, including physically approaching a person, avoiding contact with a person while passing them, and performing non-interactive tasks in proximity to a person. The researchers found that faster robot movement speeds made people uncomfortable, even in the absence of a collision; they also reported that a larger robot body led to lower comfort ratings in nearly all cases, suggesting that the robot's size and form factor can influence psychological safety.

Fischer et al. (2014) investigated human comfort during HRI in the context of acoustic communication. In their experiment, a robot attempted to pass a human participant from behind while either making no sound or emitting beeps of either rising or falling intonation. The experiment revealed significant differences between participants' comfort ratings under the three conditions, with participants reporting the highest comfort levels when the robot beeped with a rising intonation.

5.1.2 Social Considerations

Several researchers have focused on how a robot's social behavior affects psychological safety. This involves discovering which social behaviors observed during human-human interaction are important for HRI, how failing to follow social conventions during HRI impacts psychological safety, and how personality traits, culture, and demographic factors can affect these considerations. Learning this information allows for more appropriate adjustment of robot behaviors, ensuring safer interaction.

Joosse et al. (2013), for example, conducted human-subject experiments in which the agent type (human or robot) and approach speed (slow or fast) were varied during an invasion of a participant's personal space by a robot. The authors found that humans were highly sensitive to the degree to which robots obeyed social norms, and that people's attitudes and expectations toward robots were not necessarily the same as those they had toward other people; indeed, people reacted more strongly to robots invading their personal space than humans.

Feil-Seifer and Matarić (2011) explored one method of adherence to social norms through modification of robot behavior: using human demonstrations to build a model of proper following behavior for a robot leading a person to a goal. Relative distances between the human, robot, and goal in the observed human demonstrations were used to fit a GMM to formulate the model. The authors then used this model to enhance the robot's motion planner, which they then tested on various human following behaviors to show that the modified planner learned to adjust the robot's speed appropriately. In follow-up work, the authors evaluated the modified motion planner from a bystander's viewpoint by showing participants a two-dimensional, virtual representation of a leader and follower (Feil-Seifer and Matarić, 2012). The results indicated that when the modified motion planner was used, participants rated the behavior as "leading" more and "ignoring" and "avoiding" less, suggesting that the intended social behavior was more clearly conveyed with the modified planner.

In addition to proxemic distance, Kim and Mutlu (2014) investigated how "power distance" (the role of the supervisor vs. the subordinate) and "task distance" (cooperative vs. competitive performance) can affect user experience and comfort during HRI. In their first study, which involved manipulating power and proxemic distances, the authors found that user experience improved when a robot in a supervisory role was in closer physical proximity, while a robot acting as a subordinate was further away. The authors also reported that task performance worsened when the robot was physically near to the human, regardless of power distance. In a second study, the authors manipulated task and proxemic distances and found that user experience was better when competitive robots were physically close and cooperative robots were further away.

Takayama and Pantofaru (2009) wrote that robot behavior designers should also consider people's personalities and experiences. The authors determined that prior experience with robots, as well as pet ownership, reduced the physical distance people maintained with robots during interaction. Furthermore, they observed significant effects of personality traits on behavioral and attitudinal measures, as well as gender-based differences. The authors wrote that in order for HRI to be comfortable, the robot's proxemic behavior should be a function of each person's prior experience with robots, as well as his or her personality traits and gender.

In contrast, Mumm and Mutlu (2011) investigated how the robot's "personality," including likability and gaze behavior, affected physical and psychological distancing (i.e., how much information the human discloses to the robot). In the authors' experiment, the robot was made to appear either rude and selfish or pleasant and empathetic, while gaze direction was altered to be either inclined toward or averted from the participant's face. They found that participants who did not like the robot maintained greater physical and psychological distances when the robot gazed at them. Similarly to the results of the work by Takayama and Pantofaru, these findings indicate that proxemic behaviors should be a function of personal characteristics, such as gender and prior experience with robots, but also that robots must be liked by their human partners in order to enable comfortable close-proximity interaction.

While proxemic preferences are a function of personal characteristics, these preferences can shift over time and can even vary from culture to culture. In a 6-week study, Walters et al. (2011) assessed participants' proxemic preferences when interacting with robots in homelike environments. They found that the majority of adaptation of preferred interaction distances occurred within the first few sessions, then stabilized afterwards. They also wrote that robot malfunctions led to greater preferred distances, even if a safety concern was not reported, and that humans approached the robot more closely themselves than they allowed the robot to approach them when in a physically constrained area.

Research has shown that cultural differences also play a role in proxemic preferences. The seminal work by Hall (1966) indicated that people of different cultures have significantly different standards for maintaining personal space, and Joosse et al. (2014) empirically demonstrated that such standards also manifest for HRI. In that study, the authors observed that participants from different cultures (namely, the United States, China, and Argentina) maintained different standards for appropriate approach distances for a robot moving toward a small group of people.

As proxemic preferences and other social standards can vary among cultures and change over time, it is important for robots to be able to reason about these conventions and adjust their behavior appropriately. Frameworks such as the one developed by Kirby et al. (2009) can be quite useful for this purpose: in this work, social conventions are treated as constraints in a planning problem, which is solved to select socially acceptable paths for the robot. The framework allows for any number of conventions to be included, and the strengths of these conventions can be tuned appropriately to capture and enforce personal or cultural standards.

5.2 Assessment of Psychological Safety

In order to determine whether certain robot behaviors negatively affect psychological safety, and to ensure that the proposed behavior adaptations are capable of mitigating these negative effects, it is necessary to assess psychological safety in a principled manner. Assessments of psychological safety can be split into three main categories: questionnaires, physiological metrics, and behavioral metrics. These three distinct methods of assessment differ in key qualities, such as whether or not they can be collected in real-time, whether they are subject to interpretation or purely objective, and how intrusive or disruptive the measurement can be to the human. These techniques represent valuable tools for researchers to ensure that psychological safety is properly taken into consideration in HRI.

5.2.1 Questionnaires

Several validated questionnaires developed in prior work have proven useful for the assessment of psychological safety. For example, Bartneck et al. (2008) developed the “Godspeed” questionnaire, a carefully designed set of questions based on prior work in HRI that allows researchers to measure several key HRI metrics on a standardized, validated scale. Perceived safety is one of five major dimensions addressed by the questionnaire, along with anthropomorphism, animacy, likeability, and perceived intelligence.

In an attempt to quantify humans’ satisfaction during HRI, Nomura et al. (2006) developed the Negative Attitude toward Robots Scale (NARS). As its name suggests, this psychological scale measures the negative attitudes people hold toward robots. It is composed of three sub-scales assessing “situations of interaction with robots,” “the social influence of robots,” and “emotions in interaction with robots.”

More recently, Joosse et al. (2013) developed BEHAVE-II, a method that utilizes both subjective and objective metrics to assess human responses to robot behavior. This approach considers two types of responses: attitudinal and behavioral. For the former, the attitude of the person toward robots and their level of trust, as well as the robot’s human-likeness and attractiveness, are assessed via questionnaire responses. For the latter, people’s behavioral responses, such as facial expressions or the number of steps they took to move away from a robot, are analyzed. Together, these two metrics form an overall picture of a human’s attitude toward the robot during HRI.

In addition to developing and validating new questionnaires and metrics specifically designed to assess perceived safety and comfort, researchers have also incorporated such questionnaires into their experiments. Several works mentioned in the preceding sections included questions about perceived safety and comfort within their assessments, even if this was not the main topic of the research.

In the evaluation of the framework by Sisbot et al. presented in Section 3.1, Dehais et al. (2011) assessed human comfort and perceived safety through subjective questionnaire responses and physiological measurements. Similarly, Morales et al. (2015) evaluated participants’ senses of perceived comfort and pleasantness with regard to the motion plans of an autonomous wheelchair via questionnaire responses. In another paper mentioned in Section 4.3, Szafir et al. (2014) reported that small aerial vehicles utilizing motion manipulations aimed at conveying intent were rated as safer by participants than vehicles that did not incorporate motion manipulations.

Several of the papers mentioned in the previous section addressing behavioral adaptation also incorporated questionnaires to assess psychological safety. In the experiment by Butler and Agah (2001), for example, the authors used a Likert scale questionnaire to assess participants’ comfort as the robot’s speed, distance from the participant, and body design were manipulated under various conditions. In the acoustic communication experiment by Fischer et al. (2014), the authors used questionnaires to measure responses to different intonation contours. These questionnaires required participants to rate their feelings about the interaction according to seven adjectives: "angry," "comfortable," "cooperative," "relaxed," "uncomfortable," "warm," and "afraid." In the work by Kim and Mutlu (2014), the survey given to the participants contained questions not only about comfort, but also other impressions, such as whether interacting with the robot was "annoying" or "confusing," for example.

5.2.2 Physiological Metrics

While questionnaires can be a useful method of assessing psychological safety, this method also has certain drawbacks. Questionnaire items are generally open to interpretation, which can lead to unexpected biases and noise among responses. Furthermore, questionnaires can be impractical if perceived safety is to be tracked outside of controlled experiments, either for online behavior adjustment or logging for future analysis. As a result, researchers have also considered analysis of physiological metrics as a potential method of assessing psychological safety during HRI.

Kulić and Croft (2006) devised a study to determine whether physiological signals are suitable for online inference of the affective state of a human during HRI. As discussed in Section 5.1, the authors created a fuzzy inference engine for estimating the affective state in response to robot motions. The system utilizes various physiological sensors, and through analysis of the signals emitted by these sensors, the authors selected five features for study: heart rate, heart rate acceleration, skin conductance response, rate of change of skin conductance, and corrugator muscle response. Robot speed, motion type, and motion planning style were varied during the experiments in order to test the inference engine. Overall, the inference engine was found to work well for estimating affect, particularly when a participant's physiological response was greater.

In the previously mentioned work by Arai et al. (2010), the authors also evaluated human operator stress when working with robots via physiological sensing and subjective assessment. They studied the influence of robot speed on human operator stress, as well as the effects of separation distance and advance notice of robot motion, by observing the maximum amplitude and rate of spikes in the skin potential response (SPR) signal, as well as through responses to the semantic differential (SD) questionnaire, which assessed levels of fear, surprise, and discomfort. They found that all of the tested robot motion parameters influenced the SPR signal, indicating that robot speed, advance notice of robot motion, and separation distance are all important factors to consider when assessing psychological safety. Furthermore, while significant differences emerged for the SPR signal as a function of experiment conditions, the same was not always true for the SD scores, suggesting that some stresses encountered during HRI may not be consciously perceived by the human.

5.2.3 Behavioral Metrics

Another method of assessing psychological safety is through observation of human behavior in response to a robot. Similarly to physiological metrics, this method does not rely upon self-reported information, which is susceptible to variations in interpretation. One advantage of utilizing behavioral metrics over physiological signals is that physiological signals are often difficult to analyze, as a variety of emotions can affect them. (Skin conductance, for example, can increase when a person feels frightened or happy (Albert and Tullis, 2013, p. 177).) Carefully selected behavioral metrics can be a more direct proxy for perceived safety and comfort.

In the work by Takayama and Pantofaru (2009) mentioned earlier, for example, the authors conducted a study to evaluate the personal factors that can affect proxemic behaviors in HRI. This experiment incorporated various human-robot approach scenarios, with some but not all of the participants reporting prior experience with robots or pets. The robot's behavior varied such that it gazed at the participant's face during some scenarios, but not others. Interaction was evaluated according to behavioral metrics, including the average and minimum distance the person maintained from the robot during interaction, as well as attitudinal metrics of perceived safety as indicated by questionnaire responses. The researchers also took a number of personality traits and demographic measures (also derived from questionnaire responses) into consideration.

The previously mentioned work by Mumm and Mutlu (2011) also considered personal differences among humans while evaluating proxemic behaviors during HRI, grounding the work in the context of interpersonal distancing models for human-human interaction. In this study, the researchers manipulated the likeability and gaze behavior of the robot. The dependent behavioral measures of the study included metrics of both physical and psychological distancing. The former involved tracking how far the person chose to be from the robot, while the latter measured how much information a participant was willing to disclose to the robot.

One interesting example of making use of behavioral metrics of psychological safety is the work by Walters et al. (2011), discussed earlier. In their long-term proxemics study, the authors not only tracked how closely people preferred to approach the robot themselves, but also had the robot approach participants and gave them the ability to move the robot forward or backward to their preferred distance of separation.

5.3 Summary

In this section of the monograph, we discussed methods and tools for ensuring psychological safety during human-robot interaction. In contrast to the preceding sections, which addressed prevention of unwanted contact or collisions (physical safety), the methods presented in this section dealt with ensuring that interaction also feels safe and is not stressful.

We began by discussing how robot behavior adaptation can improve psychological safety. First, we summarized works that investigated the impact of features of the robot's motion, such as speed or the degree of distance maintained from the human, as well as the effect of the robot's physical appearance. Later, we highlighted the influence of social aspects on HRI, including how social conventions of human-human interaction translate into interaction with robots, the impact that violation of social standards by robots has on psychological safety, and the impact of personality traits, experience, and culture on these issues.

One of the key limitations of robot behavior adaptation is that many of the studied factors affecting psychological safety interact with one another in complex ways, making it difficult to provide concrete guidelines for parameters such as speed or the distance between the human and robot. The mentioned works allow for understanding of existing trends (e.g., the finding that increasing robot speed increases stress among human co-workers), but not necessarily concrete values for the parameters in question. The actual speed that a robot should be limited to, for example, can be a function of that robot's size and appearance, what object the robot is holding, and the human's level of prior experience with robots. While the works described here provide valuable insight into the influences of each of these factors, they are not necessarily independent.

One key limitation exists with regard to incorporating knowledge about personality traits or human experience with robots: in many HRI scenarios, obtaining such information is impractical. Take, for example, a robotic mall guide: while knowledge of a human's prior experience with robots would allow the guide to be physically closer to that human, the robot has no way of knowing how much experience with robots each person at the mall has. Consequently, this type of adaptation is limited to domains in which information about prior experience and personality traits can be obtained in advance.

After describing potential robot behavior adaptations, we discussed several potential methods for assessing psychological safety, which is necessary both for knowing when to adapt robot behavior and to ensure that such adaptations have the desired effect. Specifically, we addressed the development and use of questionnaires, utilizing physiological signals to estimate the psychological state of the human, and tracking human behavior in response to the robot.

Despite the limitations noted above, the research conducted on understanding the impact of the robot's motion and behavior on the psychological safety of nearby humans, as well as the various tools for assessing psychological safety in a principled manner, are invaluable for ensuring the overall safety of HRI.

6

Future Directions

The large body of work represented here suggests a substantial research effort related to safety during human-robot interaction. Interestingly, safety is not always explicitly mentioned as an application of these works; nonetheless, the inclusion of research relevant to safety in HRI across such a large variety of works highlights the importance of this topic. The presented techniques provide a substantial tool set that can be utilized to ensure safe HRI.

Safety in HRI remains an open problem, however, as many of the mentioned sub-fields are still relatively young. Therefore, our aim for this section is to outline potential directions for future research that would further advance safety in HRI.

6.1 Expanding Post-Collision Control Methods

A majority of the work concerned with improving safety through control has focused on collision prevention or limiting the velocity or energy stored within the system. Minimizing these parameters at all times, however, might be too conservative, causing a robot to become unnecessarily inefficient. Furthermore, in many potential close-proximity applications of HRI, strictly employing collision prevention may be unrealistic, as prevention methods primarily rely upon exteroceptive sensors, which are susceptible to inaccuracy due to issues such as occlusions, variable lighting conditions, or reflections. A momentary breakdown of perception algorithms due to such inaccuracy could prevent pre-collision systems from engaging.

The aforementioned drawbacks provide strong motivation for the use of post-collision control safety methods. Compared with pre-collision methods, however, there has been substantially less focus on this topic throughout the existing literature. The field would benefit from additional novel tests and methods assessing three areas: impacts at a wider array of contact locations not previously studied; collisions in nonstandard configurations; and intentional, collaborative contact between humans and robots.

With regard to the first area, an even more systematic approach than those used previously would further expand our knowledge of potential injury following human-robot collision. Much of the work conducted thus far has been related to impacts at and injury of the head, neck, and chest, but there has been less focus on impact at the back or legs. Given the possible dangers of a robot making unintentional contact with a person's back, understanding the potential for injury in human-robot collisions involving this part of the body is a critical research gap that must be addressed. New human-robot impact tests would also be useful for validation of prior results and the continuing development of safety standards, as formulation of tests and metrics for human-robot collisions poses one of the major challenges for standard development.

Additional work could also be performed with regard to impacts in nonstandard configurations. For example, existing tests may be inappropriate if a collision occurs within a pinch point of two robot links or if a person is pinned between a robot and a solid surface. Such configurations can also impact the rigidity of a collision, and rigid collisions have been identified as a particularly difficult technical challenge in post-collision control literature (Haddadin et al., 2007).

The knowledge gained from performing tests involving new contact points and nonstandard configurations would also be vital for development of new post-collision control methods. For example, the ideal approach for minimizing harm from impact with a person's chest may not be the optimal approach for minimizing harm from impact with a person's back. Studying how to detect whether a pinch point collision has transpired —and the best control response a robot should take in order to minimize harm in such a scenario — is also an open research problem. Developing a system capable of robustly detecting *imminent* rigid impacts, allowing for the control response to be engaged earlier, could be another topic for future research.

Ensuring safety during intentional human-robot contact is another concept that warrants further research. While some works have allowed for this interaction paradigm by, for example, detecting whether contact is intentional or by limiting exerted forces during contact, techniques that integrate detection of intentional contact and maintenance of safe interaction by limiting the amount of power transmitted to the person would be beneficial. In particular, it would be useful for newly developed methods to not rely upon engagement of a gravity-compensated mode, as is often done currently, as switching to such a mode limits the collaborative tasks that can be performed. If a person and robot are co-manipulating an object, for example, and the robot's elbow comes into contact with the person's side, it would be useful for the safety system to monitor the amount of power transferred to the person and continue the original guidance or motion until a safety threshold is reached. It is also important to monitor the power transfer through both the contact point and the object being manipulated.

While some work regarding this type of integration has already been conducted, it has mostly applied to lightweight, back-drivable robotic arms. It would be useful to extend this type of interaction to stiffer, heavy-lift industrial robots; allowing for safe collaborative interaction with such robots would enable a wide variety of useful applications within the manufacturing domain.

6.2 Extending Safe Motion Planning

One potential improvement in the realm of safe motion planning is the introduction of additional safety knowledge into planners' cost functions. While planning using parameters such as separation distance, robot inertia, or any of the other constraints mentioned in Section 3 could be quite useful, a substantial benefit could also be derived from encoding more complex safety knowledge, as each additional parameter that the robot can reason over when planning its motions would enable it to select safer paths. This would also allow the robot to utilize more efficient, direct paths when nearby humans are not in danger of being harmed.

The biggest challenge of incorporating new and more complex safety knowledge into motion planning is identifying what knowledge is significant when planning motions and determining the optimal, generalizable numerical representation of this knowledge such that it can be effectively incorporated into a cost function. Consider, for example, the high-level concept of team fluency. When an individual plans his or her motions around others during a task, he or she most likely takes the experience and knowledge of others sharing the task into account. He or she maintains a mental model of how well the team performs the given task — and, therefore how quickly and how closely to others he or she is able to move without interfering with one another's motion. It follows that incorporating this concept into robot motion planning could be useful for planning safe motions in proximity to people. The ability to encode team fluency and incorporate it into a motion planner's cost function is not a trivial task, however, and remains an open research problem.

Injury knowledge is another high-level concept that could prove useful for safe motion planning. As described in Section 2.1.1, this concept was incorporated into a pre-collision control technique by Haddadin et al. (2012), but not at the motion planning level. Expanding on the results from that work, discovering new relevant metrics of injury knowledge and embedding them into a motion planner would allow a robot to, for example, identify the type of tool it is holding, analyze the potential for human-robot impact while moving along a path, and adjust its path accordingly. In this way, the robot would perform more conservative motions when holding a dangerous object near humans than when it was positioned further away or holding a less-dangerous tool. Injury knowledge could also be applied to direct motion away from more vulnerable parts of the human body depending upon the tool being held by the robot (i.e., orienting a bright light source away from a person's eyes or keeping sharp objects away from people altogether during motion).

6.3 Improving Safety Through Prediction

Most of the human prediction methods mentioned in Section 4 are designed for very specific tasks and do not generalize well to other situations, which prevents them from being effectively deployed in real-world HRI applications. For example, if a task is well-structured and repetitive, treating it as a Markov process and utilizing probabilistic plan recognition may yield accurate predictions of how a person might move next. If a task does not possess such a structure, however, motion characteristics such as biomechanical predictors would likely generate better results. As it may not always be clear which prediction method would yield the desired performance for a particular task, fielding prediction approaches as a safety measure for HRI can be difficult. Furthermore, an incorrect prediction method could result in poor accuracy, which can have serious safety consequences.

Development of generalizable prediction methods would be a useful step toward making prediction a viable safety measure for a wider variety of tasks. How to best select among various types of predictors based on data collected from a given task and how to most effectively combine these predictors to produce accurate and robust predictions are open research questions.

There is also a limited amount of work that has combined prediction of larger body motion, such as walking, with that of more intricate motions, such as reaching. While only one type of predictor may be useful for maintaining safety in certain cases, prediction of both walking and reaching motions can be useful for planning safe robot motion during tasks involving a wide range of interaction distances.

Also, the majority of work within the realm of prediction has been focused on robot predictions of human behavior. As mentioned in Section 4.3, HRI is a two-way interaction; ensuring that robot behavior is predictable so that a person can contribute to the safety of the overall system by “planning” safe paths themselves is also quite important for safe interaction. Further research in this field could include evaluating the optimal motions for predictability based on the physical characteristics of the robot, developing new devices for effective communication of intent, or investigating the relationship between a human’s ability to predict robot behavior and whether he or she has prior experience working with robots.

6.4 Expanding Psychological Safety Considerations

Psychological safety is often overlooked when designing systems for HRI. While some research efforts have investigated maintenance of psychological safety, many have not considered this aspect of safety at all, or treated it as a secondary consideration. However, as robots are introduced as permanent residents of homes, coworkers in an office or factory setting, companions for children or the elderly, or into myriad other potential HRI applications, people will interact with robots more frequently and for extended periods of time. Consequently, negative psychological effects resulting from HRI are of substantial concern.

Physiological sensing enabling online adjustment of robot behavior is one area that warrants additional research. This capability would allow robots to reason about the stress or discomfort they induce in humans and create a feedback loop to reduce speed, maintain a greater distance from a person, or change communication modes, among other possibilities. Physiological signals are difficult to collect, however, as sensors are often large and intrusive. Effective physiological sensing for online feedback requires development of less-intrusive sensing methods. This could be accomplished through software, such as an algorithm that derives heart rate from video data (Balakrishnan et al., 2013), or hardware, such as a less-intrusive method of electroencephalography (EEG) sensors (Emotiv, 2014).

Once data is collected, understanding the resulting complex and noisy physiological signals poses another problem. Research into deriving the link between physiological signals and perceived safety and comfort through signal processing methods or other analysis is also necessary. Significant efforts to identify useful metrics from physiological signals have been made outside of the field of HRI: for example, the DARPA Augmented Cognition (AugCog) program focused on assessing cognitive activity and performance in order to adapt information systems (St. John et al., 2004). Although the results of such work may not be directly applicable to HRI safety, the insights gained can be useful for development of physiologically-based metrics of psychological safety. Tools designed for collection and analysis of physiological signals would facilitate integration of physiological measurements into HRI safety methods, allowing for maintenance of psychological safety across all HRI applications.

Also, greater consistency in evaluation of psychological factors would be beneficial for the field. Regardless of whether a given work specifically addresses the issue of safety, researchers should strive to ensure the algorithms and systems they develop not only improve interaction by some metric, but also do so without compromising psychological safety. By collecting psychological safety data, researchers can better understand the potential impact of their systems and are better equipped to refine their work such that it is less likely to result in harm. However, validated perceived safety and comfort metrics are necessary in order for researchers to be able to do this effectively. While some validated surveys for HRI do exist, such as the “Godspeed” or “BEHAVE-II” questionnaires mentioned in Section 5.2.1, the scope of these surveys is typically quite broad and extends far beyond assessing perceived levels of safety and comfort. The relatively small number of safety-relevant items included in surveys of this scope may not be sufficient for comprehensive evaluation.

6.5 Integration of Safety Methods

The four main methods of ensuring safety in HRI described throughout this work have various benefits and drawbacks individually. However, by combining these methods, it is possible to exploit their individual, complementary strengths to develop a more effective integrated safety system.

Several works mentioned earlier have integrated multiple prediction methods. In the paper by Lasota and Shah (2015), for example, using a human-aware motion planner in conjunction with a pre-collision control system that adjusts the robot's speed based on distance of separation, the system was able to prevent collisions and also exhibited improved psychological safety over use of the control-based component alone. Similarly, Kulić and Croft (2007) developed a combined system containing safety components for various time horizons that also incorporates both control- and planning-based methods. The authors were able to demonstrate smooth integration of these components through empirical evaluation.

While the above works represent some examples of the integration of multiple prediction techniques, the field would benefit from additional research into which methods to combine and how best to combine them. Integrating post-collision control methods into safety systems, for example, could provide an effective failsafe for planning- or prediction-based approaches.

Creating such new integrated safety systems poses certain challenges, however. First, it is necessary to determine a method for balancing the contributions of the various sub-components under circumstances in which they might suggest conflicting responses. (Which methods should take precedence, and what factors must be considered when making such a decision?) Second, it is difficult to evaluate the effectiveness of a complex, multi-tiered safety system. A standardized method of testing and comparing safety systems, which incorporates a wide variety of scenarios and test cases, is needed. Which scenarios and test cases are most important, and which measures of effectiveness should be considered most relevant, are both open questions.

7

Conclusion

In this work, we surveyed and categorized prior research addressing safety during human-robot interaction in order to identify and describe various potential methods of ensuring safe HRI. Toward this goal, we identified four main methods of providing safety: control, motion planning, prediction, and consideration of psychological factors.

Although significant strides have been made thus far, ensuring safe HRI remains an open problem. Novel, robust, and generalizable safety methods are required in order to enable safe incorporation of robots into homes, offices, factories, or any other setting.

Due to the commercial incentives for introducing robotic assistants onto factory floors, there has been a significant amount of interest in HRI within the manufacturing domain. Consequently, much of the work surveyed in this monograph focused on interaction with manipulator arms in the manufacturing setting. The presented techniques and methods, however, can also be adapted and applied to a variety of other types of robots and domains. Through this work, we hope to encourage and facilitate such adaptation, along with the development of new methods for safety in HRI as it expands into these new domains.

By applying and building upon the lessons learned from prior work, the research community will be able to make HRI increasingly safe over time, which will inherently decrease the risks associated with HRI. Mitigating this risk will, in turn, lead to a more rapid transition of HRI systems from research labs into homes, offices, and factories.

Acknowledgements

This work was supported by a NASA Space Technology Research Fellowship. The authors would like to thank Yunkyung Kim and David Feil-Seifer for their help with improving the “Safety Through Consideration of Psychological Factors” section.

References

AO Foundation - transforming surgery - changing lives. AO Foundation, `https://www.aofoundation.org`, 2015.

ABB. Yumi - creating an automated future together. ABB, `http://new.abb.com/products/robotics/yumi`, 2015.

W. Albert and T. Tullis. *Measuring the User Experience: Collecting, Analyzing, and Presenting Usability Metrics*. Interactive Technologies. Elsevier Science, 2013.

T. Arai, R. Kato, and M. Fujita. Assessment of operator stress induced by robot collaboration in assembly. *CIRP Annals - Manufacturing Technology*, 59(1):5–8, January 2010.

I. Asimov. Runaround. In *Astounding Science Fiction*. 1942.

J. Azorin-Lopez, M. Saval-Calvo, A. Fuster-Guillo, and A. Oliver-Albert. A predictive model for recognizing human behaviour based on trajectory representation. In *2014 International Joint Conference on Neural Networks (IJCNN)*, pages 1494–1501. IEEE, July 2014.

G. Balakrishnan, F. Durand, and J. Guttag. Detecting Pulse from Head Motions in Video. In *Proceedings of Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3430–3437. IEEE, June 2013.

C. Bartneck, D. Kulić, E. Croft, and S. Zoghbi. Measurement Instruments for the Anthropomorphism, Animacy, Likeability, Perceived Intelligence, and Perceived Safety of Robots. *International Journal of Social Robotics*, 1(1): 71–81, November 2008.

G. Bell, A. Parisi, and M. Pesce. The virtual reality modeling language. 1995.

M. Bennewitz, W. Burgard, G. Cielniak, and S. Thrun. Learning Motion Patterns of People for Compliant Robot Motion. *The International Journal of Robotics Research*, 24(1):31–48, January 2005.

X. Broquere, D. Sidobre, and I. Herrera-Aguilar. Soft motion trajectory planner for service manipulator robot. In *Proceedings of IROS*, pages 2808–2813. IEEE, September 2008.

M. Bualat, J. Barlow, T. Fong, C. Provencher, T. Smith, and A. Zuniga. Astrobee: Developing a free-flying robot for the international space station. In *AIAA SPACE 2015 Conference and Exposition*, page 4643, 2015.

G. Buizza Avanzini, N. M. Ceriani, A. M. Zanchettin, P. Rocco, and L. Bascetta. Safety Control of Industrial Robots Based on a Distributed Distance Sensor. *IEEE Transactions on Control Systems Technology*, 22 (6):2127–2140, November 2014.

J. T. Butler and A. Agah. Psychological Effects of Behavior Patterns of a Mobile Personal Robot. *Autonomous Robots*, 10(2):185–202, 2001.

S. Calinon, I. Sardellitti, and D. G. Caldwell. Learning-based control strategy for safe human-robot interaction exploiting task and robot redundancies. In *Proceedings of IROS*, pages 249–254. IEEE, October 2010.

S. Cambon, R. Alami, and F. Gravot. A Hybrid Approach to Intricate Motion, Manipulation and Task Planning. *The International Journal of Robotics Research*, 28(1):104–126, January 2009.

D. B. Chaffin and J. J. Faraway. Stature, age and gender effects on reach motion postures. *Human Factors*, pages 408–420, 2000.

A. De Luca and F. Flacco. Integrated control for pHRI: Collision avoidance, detection, reaction and collaboration. In *Proceedings of BioRob*, pages 288–295. IEEE, June 2012.

A. De Luca, A. Albu-Schäffer, S. Haddadin, and G. Hirzinger. Collision detection and safe reaction with the DLR-III lightweight manipulator arm. In *Proceedings of IROS*, pages 1623–1630, 2006.

A. De Luca, F. Flacco, A. Bicchi, and R. Schiavi. Nonlinear decoupled motion-stiffness control and collision detection/reaction for the VSA-II variable stiffness device. In *Proceedings of IROS*, pages 5487–5494. IEEE, October 2009.

F. Dehais, E. A. Sisbot, R. Alami, and M. Causse. Physiological and subjective evaluation of a human-robot object hand-over task. *Applied ergonomics*, 42 (6):785–91, November 2011.

M. Diftler, J. Mehling, M. Abdallah, N. Radford, L. Bridgwater, A. Sanders, R. Askew, D. Linn, J. Yamokoski, F. Permenter, B. Hargrave, R. Piatt, R. Savely, and R. Ambrose. Robonaut 2 - the first humanoid robot in space. In *Proceedings of ICRA*, pages 2178–2183, May 2011.

P. F. Dominey, G. Metta, F. Nori, and L. Natale. Anticipation and initiative in human-humanoid interaction. In *Proceedings of International Conference on Humanoid Robots (Humanoids)*, pages 693–699. IEEE-RAS, 2008.

A. D. Dragan, K. C. Lee, and S. S. Srinivasa. Legibility and predictability of robot motion. In *Proceedings of HRI*, pages 301–308, 2013.

A. D. Dragan, S. Bauman, J. Forlizzi, and S. S. Srinivasa. Effects of Robot Motion on Human-Robot Collaboration. In *Proceedings of HRI*, pages 51–58, 2015.

J. Elfring, R. van de Molengraft, and M. Steinbuch. Learning intentions for improved human motion prediction. *Robotics and Autonomous Systems*, 62 (4):591–602, April 2014.

Emotiv. Emotiv insight: The next generation brainwear. Emotiv, `https://emotiv.com/insight.php`, 2014.

E. Erdem, K. Haspalamutgil, C. Palaz, V. Patoglu, and T. Uras. Combining high-level causal reasoning with low-level geometric reasoning and motion planning for robotic manipulation. In *Proceedings of ICRA*, pages 4575–4581. IEEE, May 2011.

M. Erden and T. Tomiyama. Human-Intent Detection and Physically Interactive Control of a Robot Without Force Sensors. *IEEE Transactions on Robotics*, 26(2):370–382, April 2010.

D. Feil-Seifer and M. Matarić. People-aware navigation for goal-oriented behavior involving a human partner. In *Proceedings of International Conference on Development and Learning*. IEEE, 2011.

D. Feil-Seifer and M. Matarić. Distance-Based Computational Models for Facilitating Robot Interaction with Children. *Journal of Human-Robot Interaction*, 1(1), 2012.

R. E. Fikes and N. J. Nilsson. Strips: A new approach to the application of theorem proving to problem solving. *Artificial Intelligence*, 2(3-4):189–208, December 1971.

K. Fischer, L. C. Jensen, and L. Bodenhagen. To Beep or Not to Beep Is Not the Whole Question. In *Proceedings of International Conference on Social Robotics (ICSR)*. Springer International Publishing, 2014.

F. Flacco, T. Kroger, A. De Luca, and O. Khatib. A depth space approach to human-robot collision avoidance. In *Proceedings of ICRA*, pages 338–345. IEEE, May 2012.

T. Fong, I. Nourbakhsh, and K. Dautenhahn. A survey of socially interactive robots. *Robotics and Autonomous Systems*, 42(3-4):143–166, March 2003.

T. Fong, M. Micire, T. Morse, E. Park, C. Provencher, V. To, D. Wheeler, D. Mittman, R. J. Torres, and E. Smith. Smart spheres: a telerobotic free-flyer for intravehicular activities in space. In *Proceedings AIAA Space*, volume 13, 2013.

Fraunhofer IFF. Determination of verified thresholds for safe human-robot collaboration. Fraunhofer Institute for Factory Operation and Automation IFF, `http://iff.fraunhofer.de`, 2013.

M. Geravand, F. Flacco, and A. De Luca. Human-robot physical interaction and collaboration using an industrial robot with a closed control architecture. In *Proceedings of ICRA*, pages 4000–4007, 2013.

B. Gleeson, K. MacLean, A. Haddadi, E. Croft, and J. Alcazar. Gestures for industry: Intuitive human-robot communication from human observation. In *Proceedings of HRI*, pages 349–356. IEEE Press, 2013.

S. Golz, C. Osendorfer, and S. Haddadin. Using tactile sensation for learning contact knowledge: Discriminate collision from physical interaction. In *Proceedings of ICRA*, pages 3788–3794, 2015.

B. Graf, M. Hans, and R. D. Schraft. Care-o-bot ii–development of a next generation robotic home assistant. *Autonomous robots*, 16(2):193–205, 2004.

S. Haddadin, A. Albu-Schäffer, and G. Hirzinger. Soft-tissue injury in robotics. In *Proceedings of ICRA*, pages 3426–3433, May 2010a.

S. Haddadin, H. Urbanek, S. Parusel, D. Burschka, J. Rossmann, A. Albu-Schäffer, and G. Hirzinger. Real-time reactive motion generation based on variable attractor dynamics and shaped velocities. In *Proceedings of IROS*, pages 3109–3116. IEEE, October 2010b.

S. Haddadin, A. Khoury, T. Rokahr, S. Parusel, R. Burgkart, A. Bicchi, and A. Albu-Schäffer. On making robots understand safety: Embedding injury knowledge into control. *International Journal of Robotics Research*, 31(13): 1578–1602, November 2012.

S. Haddadin. *Towards safe robots: approaching Asimov's 1st law*, volume 90. Springer, 2013.

S. Haddadin, A. Albu-Schäffer, and G. Hirzinger. Safety evaluation of physical human-robot interaction via crash-testing. In *Proceedings of RSS*, pages 217–224, 2007.

S. Haddadin, A. Albu-Schäffer, A. De Luca, and G. Hirzinger. Collision detection and reaction: A contribution to safe physical Human-Robot Interaction. In *Proceedings of IROS*, pages 3356–3363. IEEE, September 2008.

S. Haddadin, A. Albu-Schäffer, and G. Hirzinger. Requirements for Safe Robots: Measurements, Analysis and New Insights. *International Journal of Robotics Research*, 28:1507–1527, 2009.

E. T. Hall. *The Hidden Dimension.* Doubleday, 1966.

V. R. Ham, T. G. Sugar, B. Vanderborght, K. W. Hollander, and D. Lefeber. Compliant actuator designs: Review of actuators with passive adjustable compliance/controllable stiffness for robotic applications. *IEEE Robotics and Automation Magazine*, 16:81–94, 2009.

K. P. Hawkins, N. Vo, S. Bansal, and A. F. Bobic. Probabilistic Human Action Prediction and Wait-sensitive Planning for Responsive Human-robot Collaboration. In *Proceedings of International Conference on Humanoid Robots (Humanoids)*. IEEE-RAS, 2013.

K. P. Hawkins, S. Bansal, N. N. Vo, and A. F. Bobick. Anticipating human actions for collaboration in the presence of task and sensor uncertainty. In *Proceedings of ICRA*, pages 2215–2222. IEEE, May 2014.

J. Heinzmann and A. Zelinsky. Quantitative Safety Guarantees for Physical Human-Robot Interaction. *International Journal of Robotics Research*, 22: 479–504, 2003.

G. Hoffman and C. Breazeal. Cost-based anticipatory action selection for human-robot fluency. In *IEEE Transactions on Robotics*, volume 23, pages 952–961, 2007.

T. Hoshi and H. Shinoda. Robot skin based on touch-area-sensitive tactile element. In *Proceedings of ICRA*, pages 3463–3468. IEEE, 2006.

C.-M. Huang, M. Cakmak, and B. Mutlu. Adaptive Coordination Strategies for Human-Robot Handovers, 2015.

International Organization for Standardization. ISO 10218-1:2011 Robots and robotic devices – Safety requirements for industrial robots – Part 1: Robots. International Organization for Standardization, `http://www.iso.org`, 2011a.

International Organization for Standardization. ISO 10218-2:2011 Robots and robotic devices – Safety requirements for industrial robots – Part 2: Robot systems and integration. International Organization for Standardization, `http://www.iso.org`, 2011b.

International Organization for Standardization. ISO 15066 Robots and robotic devices – Collaborative robots. International Organization for Standardization, `http://www.iso.org`, 2016.

Y. Jiang and A. Saxena. Modeling High-Dimensional Humans for Activity Anticipation using Gaussian Process Latent CRFs. In *Proceedings of RSS*, 2014.

M. Joosse, A. Sardar, M. Lohse, and V. Evers. BEHAVE-II: The Revised Set of Measures to Assess User's Attitudinal and Behavioral Responses to a Social Robot. *International Journal of Social Robotics*, 5(3):379–388, June 2013.

M. P. Joosse, R. W. Poppe, M. Lohse, and V. Evers. Cultural differences in how an engagement-seeking robot should approach a group of people. In *Proceedings of International Conference on Collaboration Across Boundaries: Culture, Distance & Technology (CABS)*, pages 121–130. ACM Press, August 2014.

Jung-Jun Park and Jae-Bok Song. Collision analysis and evaluation of collision safety for service robots working in human environments. In *Proceedings of International Conference on Advanced Robotics*, pages 1–6, 2009.

L. P. Kaelbling and T. Lozano-Perez. Hierarchical task and motion planning in the now. In *Proceedings of ICRA*, pages 1470–1477. IEEE, May 2011.

S. G. Khan, G. Herrmann, M. Al Grafi, T. Pipe, and C. Melhuish. Compliance Control and Human-Robot Interaction: Part 1 - Survey. *International Journal of Humanoid Robotics*, 11(03):1430001, September 2014.

O. Khatib. Real-Time Obstacle Avoidance for Manipulators and Mobile Robots. *International Journal of Robotics Research*, 5(1):90–98, March 1986.

Y. Kim and B. Mutlu. How social distance shapes human-robot interaction. *International Journal of Human Computer Studies*, 72(12):783–795, 2014.

R. Kirby, R. Simmons, and J. Forlizzi. COMPANION: A Constraint-Optimizing Method for Person-Acceptable Navigation. In *Proceedings of International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pages 607–612. IEEE, September 2009.

W. Knight. Smart robots can now work right next to auto workers. *MIT Technology Review*, 17, 2013.

K. L. Koay, K. Dautenhahn, S. N. Woods, and M. L. Walters. Empirical results from using a comfort level device in human-robot interaction studies. In *Proceedings of HRI*, page 194. ACM Press, March 2006.

S. Kock, J. Bredahl, P. J. Eriksson, M. Myhr, and K. Behnisch. Taming the robot Better safety without higher fences. *ABB Review*, pages 11–14, April 2006.

S. A. Kolakowsky-Hayner, J. Crew, S. Moran, and A. Shah. Safety and Feasibility of using the EksoTM Bionic Exoskeleton to Aid Ambulation after Spinal Cord Injury. *Journal of Spine*, 2013(2):S4–003, 2013.

H. Koppula and A. Saxena. Learning Spatio-Temporal Structure from RGB-D Videos for Human Activity Detection and Anticipation. In *Proceedings of International Conference on Machine Learning (ICML)*, pages 792–800, 2013a.

H. S. Koppula and A. Saxena. Anticipating Human Activities using Object Affordances for Reactive Robotic Response. In *Proceedings of RSS*, 2013b.

M. Kuderer, H. Kretzschmar, C. Sprunk, and W. Burgard. Feature-Based Prediction of Trajectories for Socially Compliant Navigation. In *Proceedings of RSS*, 2012.

KUKA. LBR IIWA 14 R820. KUKA, `http://www.kuka-robotics.com/en/products/industrial_robots/sensitiv/lbr_iiwa_14_r820/start.htm`, 2015.

D. Kulić and E. Croft. Physiological and subjective responses to articulated robot motion. *Robotica*, 25(01):13, August 2006.

D. Kulić and E. Croft. Pre-collision safety strategies for human-robot interaction. *Autonomous Robots*, 22:149–164, 2007.

D. Kulić and E. A. Croft. Safe planning for human-robot interaction. *Journal of Robotic Systems*, 22(7):383–396, July 2005.

B. Lacevic and P. Rocco. Kinetostatic danger field - A novel safety assessment for human-robot interaction. In *Proceedings of IROS*, pages 2169–2174, 2010.

M. Laffranchi, N. G. Tsagarakis, and D. G. Caldwell. Safe human robot interaction via energy regulation control. In *Proceedings of IROS*, pages 35–41. IEEE, October 2009.

S. Lallee, K. Hamann, J. Steinwender, F. Warneken, U. Martienz, H. Barron-Gonzales, U. Pattacini, I. Gori, M. Petit, G. Metta, P. Verschure, and P. Ford Dominey. Cooperative human robot interaction systems: IV. Communication of shared plans with Naïve humans using gaze and speech. In *Proceedings of IROS*, pages 129–136, 2013.

P. A. Lasota and J. A. Shah. Analyzing the Effects of Human-Aware Motion Planning on Close-Proximity Human-Robot Collaboration. *Human Factors*, 57(1):21–33, January 2015.

P. A. Lasota, G. F. Rossano, and J. A. Shah. Toward safe close-proximity human-robot interaction with standard industrial robots. In *Proceedings of CASE*, pages 339–344. IEEE, August 2014.

K. Li and Y. Fu. Prediction of Human Activity by Discovering Temporal Sequence Patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(8):1644–1657, August 2014.

M. Luber, J. A. Stork, G. D. Tipaldi, and K. O. Arras. People tracking with human motion predictions from social forces. In *Proceedings of ICRA*, pages 464–469. IEEE, May 2010.

J. Mainprice and D. Berenson. Human-robot collaborative manipulation planning using early prediction of human motion. In *Proceedings of IROS*, pages 299–306. IEEE, November 2013.

J. Mainprice, E. Akin Sisbot, L. Jaillet, J. Cortes, R. Alami, and T. Simeon. Planning human-aware motions using a sampling-based costmap planner. In *Proceedings of ICRA*, pages 5012–5017, 2011.

J. Mainprice, R. Hayne, and D. Berenson. Predicting human reaching motion in collaborative tasks using Inverse Optimal Control and iterative replanning. In *Proceedings of ICRA*, pages 885–892, 2015.

N. McCain. *Causality in Commonsense Reasoning about Actions*. PhD thesis, Computer Sciences Department, The University of Texas at Austin, 1997.

B. S. McEwen. Stress and the Individual. *Archives of Internal Medicine*, 153 (18):2093, September 1993.

Y. Morales, A. Watanabe, F. Ferreri, J. Even, T. Ikeda, K. Shinozawa, T. Miyashita, and N. Hagita. Including human factors for planning comfortable paths. In *Proceedings of ICRA*, pages 6153–6159, 2015.

J. Mumm and B. Mutlu. Human-robot proxemics: Physical and Psychological Distancing in Human-Robot Interaction. In *Proceedings of HRI*, page 331, 2011.

National Institute of Standards and Technology. Performance assessment framework for robotic systems. The National Institute of Standards and Technology (NIST), `http://www.nist.gov`, 2013.

S. Nikolaidis, P. Lasota, G. Rossano, C. Martinez, T. Fuhlbrigge, and J. Shah. Human-robot collaboration in manufacturing: Quantitative evaluation of predictable, convergent joint action. In *Proceedings of ISR*, pages 1–6. IEEE, October 2013.

S. Nikolaidis, R. Ramakrishnan, K. Gu, and J. Shah. Efficient model learning from joint-action demonstrations for human-robot collaborative tasks. In *Proceedings of HRI*, pages 189–196. ACM, 2015.

T. Nomura, T. Suzuki, T. Kanda, and K. Kato. Measurement of negative attitudes toward robots. *Interaction Studies*, 7(3):437–454, 2006.

S. Oberer and R. D. Schraft. Robot-Dummy Crash Tests for Robot Safety Assessment. In *Proceedings of ICRA*, pages 2934–2939, 2007.

L. O'Sullivan, R. Nugent, and J. van der Vorm. Standards for the Safety of Exoskeletons Used by Industrial Workers Performing Manual Handling Activities: A Contribution from the Robo-Mate Project to their Future Development. *Procedia Manufacturing*, 3:1418–1425, 2015.

C. Pérez-D'Arpino and J. A. Shah. Fast Target Prediction of Human Reaching Motion for Cooperative Human-Robot Manipulation Tasks using Time Series Classification. In *Proceedings of ICRA*, 2015.

A. Pervez and J. Ryu. Safe physical human robot interaction-past, present and future. *Journal of Mechanical Science and Technology*, 22:469–483, 2008.

E. Plaku and G. D. Hager. Sampling-Based Motion and Symbolic Action Planning with geometric and differential constraints. In *Proceedings of ICRA*, pages 5002–5008. IEEE, May 2010.

M. P. Polverini, A. M. Zanchettin, and P. Rocco. Real-time collision avoidance in human-robot interaction based on kinetostatic safety field. In *Proceedings of IROS*, pages 4136–4141. IEEE, September 2014.

B. Povse, D. Koritnik, T. Bajd, and M. Munih. Correlation between impact-energy density and pain intensity during robot-man collision. In *Proceedings of BioRob*, pages 179–183, 2010.

G. Pratt and M. Williamson. Series elastic actuators. In *Proceedings of IROS*, volume 1, pages 399–406. IEEE, 1995.

R. Robotics. Baxter with intera 3. RethinkRobotics, `http://www.rethinkrobotics.com/baxter/`, 2015a.

R. Robotics. Sawyer with intera 3. RethinkRobotics, `http://www.rethinkrobotics.com/sawyer-intera-3/`, 2015b.

P. Rybski, P. Anderson-Sprecher, D. Huber, C. Niessl, and R. Simmons. Sensor fusion for human safety in industrial workcells. In *Proceedings of IROS*, pages 3612–3619. IEEE, October 2012.

M. S. Ryoo. Human activity prediction: Early recognition of ongoing activities from streaming videos. In *Proceedings of International Conference on Computer Vision (ICCV)*, pages 1036–1043. IEEE, 2011.

M. S. Ryoo, T. J. Fuchs, L. Xia, J. K. Aggarwal, and L. Matthies. Robot-Centric Activity Prediction from First-Person Videos : What Will They Do to Me? In *Proceedings of HRI*, pages 295–302, 2015.

E. A. Sisbot and R. Alami. A Human-Aware Manipulation Planner. *IEEE Transactions on Robotics*, 28(5):1045–1057, October 2012.

E. A. Sisbot, L. F. Marin-Urias, R. Alami, and T. Simeon. A Human Aware Mobile Robot Motion Planner. *IEEE Transactions on Robotics*, 23(5):874–883, October 2007.

E. A. Sisbot, L. F. Marin-Urias, X. Broquère, D. Sidobre, and R. Alami. Synthesizing Robot Motions Adapted to Human Presence. *International Journal of Social Robotics*, 2(3):329–343, June 2010.

A. St. Clair and M. Mataric. How Robot Verbal Feedback Can Improve Team Performance in Human-Robot Task Collaborations. In *Proceedings of HRI*, pages 213–220. ACM Press, March 2015.

M. St. John, D. A. Kobus, J. G. Morrison, and D. Schmorrow. Overview of the darpa augmented cognition technical integration experiment. *International Journal of Human-Computer Interaction*, 17(2):131–149, 2004.

N. Stergiou. *Innovative analyses of human movement*. Human Kinetics, 1st edition, 2004.

D. Szafir, B. Mutlu, and T. Fong. Communication of Intent in Assistive Free Flyers. In *Proceedings of HRI*, pages 358–365, 2014.

D. Szafir, B. Mutlu, and T. Fong. Communicating Directionality in Flying Robots. In *Proceedings of HRI*, pages 19–26, 2015.

W. Takano, H. Imagawa, and Y. Nakamura. Prediction of human behaviors in the future through symbolic inference. In *Proceedings of ICRA*, pages 1970–1975, 2011.

L. Takayama and C. Pantofaru. Influences on proxemic behaviors in human-robot interaction. In *Proceedings of IROS*, pages 5495–5502, 2009.

L. Takayama, D. Dooley, and W. Ju. Expressing thought: Improving Robot Readability with Animation Principles. In *Proceedings of HRI*, page 69, 2011.

V. V. Unhelkar, J. Perez, J. C. Boerkoel, J. Bix, S. Bartscher, and J. A. Shah. Towards control and sensing for an autonomous mobile robotic assistant navigating assembly lines. In *Proceedings of ICRA*, pages 4161–4167. IEEE, 2014.

V. V. Unhelkar, P. Claudia, L. Stirling, and J. A. Shah. Human-Robot Co-Navigation using Anticipatory Indicators of Human Walking Motion. In *Proceedings of ICRA*, 2015.

B. Vanderborght, A. Albu-Schäeffer, A. Bicchi, E. Burdet, D. Caldwell, R. Carloni, M. Catalano, O. Eiberger, W. Friedl, G. Ganesh, M. Garabini, M. Grebenstein, G. Grioli, S. Haddadin, H. Hoppner, A. Jafari, M. Laffranchi, D. Lefeber, F. Petit, S. Stramigioli, N. Tsagarakis, M. Van Damme, R. Van Ham, L. Visser, and S. Wolf. Variable impedance actuators: A review. *Robotics and Autonomous Systems*, 61(12):1601–1614, December 2013.

M. Vasic and A. Billard. Safety issues in human-robot interactions. In *Proceedings of ICRA*, pages 197–204. IEEE, May 2013.

A. Vick, D. Surdilovic, and J. Krüger. Safe physical human-robot interaction with industrial dual-arm robots. In *Proceedings of International Workshop on Robot Motion and Control (RoMoCo)*, pages 264–269, 2013.

C. Vogel, C. Walter, and N. Elkmann. A projection-based sensor system for safe physical human-robot collaboration. In *Proceedings of IROS*, pages 5359–5364, 2013.

J. Walker, A. Gupta, and M. Hebert. Patch to the Future: Unsupervised Visual Prediction. In *Proceedings of Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3302–3309. IEEE, June 2014.

M. L. Walters, M. A. Oskoei, D. S. Syrdal, and K. Dautenhahn. A long-term Human-Robot Proxemic study. In *Proceedings of International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pages 137–142. IEEE, July 2011.

J. Wolfe, B. Marthi, and S. Russell. Combined task and motion planning for mobile manipulation. In *Proceedings of International Conference on Automated Planning and Scheduling (ICAPS)*, 2010.

S. Xiao, Z. Wang, and J. Folkesson. Unsupervised robot learning to predict person motion. In *Proceedings of ICRA*, pages 691–696, 2015.

J. E. Young, R. Hawkins, E. Sharlin, and T. Igarashi. Toward Acceptable Domestic Robots: Applying Insights from Social Psychology. *International Journal of Social Robotics*, 1(1):95–108, November 2008.

A. M. Zanchettin, N. M. Ceriani, P. Rocco, H. Ding, and B. Matthias. Safety in Human-Robot Collaborative Manufacturing Environments: Metrics and Control. *IEEE Transactions on Automation Science and Engineering*, pages 1–12, 2015.

G. Zeilig, H. Weingarden, M. Zwecker, I. Dudkiewicz, A. Bloch, and A. Esquenazi. Safety and tolerance of the ReWalk^{TM} exoskeleton suit for ambulation by people with complete spinal cord injury: A pilot study. *The Journal of Spinal Cord Medicine*, 35(2):96–101, March 2012.

B. D. Ziebart, N. Ratliff, G. Gallagher, C. Mertz, K. Peterson, J. A. Bagnell, M. Hebert, A. K. Dey, and S. Srinivasa. Planning-based prediction for pedestrians. In *Proceedings of IROS*, pages 3931–3936, 2009.

M. Zinn, O. Khatib, B. Roth, and J. Salisbury. Playing it safe [human-friendly robots]. *Robotics Automation Magazine, IEEE*, 11(2):12–21, June 2004.