A New Framework for Robot Privacy

Kevin L. Miller[1]

Abstract

The presence of robotic devices in our environment gives rise to unique privacy problems unlike those in other domains. Despite rapid advancement in the perception, movement, and learning capabilities of robots, issues in robot privacy remain without an effective research program. This research advances the conversation by proposing technological solutions aimed at the nexus between privacy as a legal and sociological concept and robot control in multi-actor environments. The following explores the system architecture and characteristics of a technical framework for making available, fusing and reconciling the privacy preference data of multiple actors across every contextual level (cultural, societal, group, locational, individual, and situational) and transforming them into concrete instructions usable by the robot as higher-level behavioral controls.

To that end, a taxonomic schema is described that can be accessed by robotic device makers to inform sensor collection, data collection, storage parameters and constraints, and the permissible range of movements, motions, and activities of a robot based on individualized, context- and role-sensitive privacy preference rules. A privacy preference enunciator device and associated transport mechanisms are introduced that allow individuals and the robots they encounter in ad hoc environments to exchange privacy preference data in accordance with the taxonomic schema. Privacy preference rule selection and comprehensive resolution protocols are developed that allow for the automated or interactive resolution of conflicts arising between individuals in multi-actor environments or ambiguous contexts. Accountability and audit mechanisms are discussed, as are trust and security models for mitigating secondary privacy harms.

1. Introduction & Motivation

In the near future, robots will increasingly populate the physical space in which humans live and work—perceiving our presence, observing and interpreting what we say and do, recording video, audio, and other sensor data, and physically interacting with us. Whether human or autonomous device, someone or something that does not appear to respect our personal boundaries is considered "creepy," which is that vague sensation of violation some people feel when their

[1] **Kevin L. Miller** is a shareholder at Labyrinth Law PLLC (www.labyrinthlaw.com), where he is an intellectual property, patent, and technology law attorney. His practice focuses on innovative technologies in software and computer engineering, cybersecurity, privacy, algorithm law, and other cutting-edge issues at the intersection of law and technology. Before becoming an attorney, he was a software engineer and architect for several major technology companies, including Microsoft, and an adjunct professor of computer science. He is the author of several articles on cybersecurity and privacy issues and a book on software development design techniques. He can be reached at kevin@labyrinthlaw.com.

personal space is encroached upon or when their behaviors are observed more closely than they expected. Our perception of creepiness often arises from the presence of devices in our environment that are controlled by others—devices we suspect may be making different choices about our privacy than we ourselves would make. The idea that a company, person, or the government might be able to use their devices to monitor their conversations or actions is unsettling to many people. So long as robotic devices are perceived to be generally misaligned with our ultimate privacy goals, the sensation of robots as creepy will persist and widespread adoption will be impacted. Public response to devices like Google Glass [1] [2] and Amazon Alexa [3] has largely borne this out.

The goal of this interdisciplinary research is to investigate technological approaches that lay the groundwork for addressing the very real privacy management challenges arising when humans coexist with robots and other devices. As applied to the field of robotics, the purpose of privacy management is to govern the observation, movement, and recording activities of a robot in accordance with the expectations of the humans with which it interacts. A privacy management scheme for robots has several aspects that make it unique.

The first aspect is that the privacy management concerns stemming from the use of robots are qualitatively different from those encountered in conventional website and mobile device apps. The current paradigm of website and mobile app data privacy is incentivized by the lack of a viable economic model to monetize most web services and content publication. Thus, website and mobile app privacy tends to be defined by "notice and consent" techniques that are primarily concerned with obtaining broad permissions from consumers to sell their personal information or behavioral data to third parties for marketing purposes. Participants in this system have allowed this notion of information privacy and its associated notice and consent modality to define most aspects of the data privacy conversation, from its regulatory motifs to the design of the privacy setting user interfaces for giving or denying consent. Moreover, in privacy jurisprudence, privacy expectations are bounded by a notion of "reasonableness" that sets the threshold for Fourth Amendment protections and tort-based privacy protections [4]. This means that people are only protected against privacy violations when the intrusion is not reasonable and expected. The interplay of the notice and consent modality with the amorphousness of the "reasonable expectation" doctrine means that, over time, privacy becomes inexorably eroded as individuals hand out blanket

permission slips for broad swaths of personal information to web service providers in return for "free" use of their services and apps.

Robot privacy is a much harder and more nuanced problem than web privacy. To be sure, robot privacy includes some classic information privacy concerns like those in website data sharing, but it must also account for physical privacy. "Physical privacy," as understood here, includes concepts such as whether a robot can record, or even measure, a person's physical characteristics with sensors (e.g., audio recording or heart rate monitoring); a robot's physical proximity when interacting with a person in certain contexts; and whether, and in what manner, a robot can touch a person. These kinds of physical privacy are much more closely related to those protected by classic privacy torts such as "intrusion upon seclusion" and battery (*see, e.g.,* [4]). Traditional notice and consent mechanisms, considered by many commentators as largely ineffective even within their own purview (*see, e.g.,* [5] [6]), are likely to be completely insufficient when applied to robot privacy management, which needs to provide granular and scenario-specific restrictions on the range of actions a robot can take in a wide variety of environments, from assisting an elderly man in the shower to handing out brochures at a shopping mall.

The second issue is that robot privacy management is fundamentally dynamic and contextual. Unlike in web-based privacy models, people and robots are mobile—robots can move into different physical spaces inhabited by different people, and different people can enter or exit a robot's functional proximity at any time. Privacy expectations are also based on cultural norms, shared group values, and even on physical location. Sometimes, situational contexts such as an emergency will override all other concerns. Thus, any proposed solution must facilitate a common consistent standard that assists robots in acting in alignment with our contextually-informed values. Related to the concept of mobility is how to disseminate privacy management settings to the various robots a person may encounter in daily life that she neither recognizes nor controls. Maintaining consistency of privacy settings and ease of configuration across the totality of the robot environment are important problems to be solved. For example, how will a robot be informed of the privacy management preferences of an overnight guest?

Further complicating matters, privacy management becomes exponentially more difficult in real-world scenarios where robots must select appropriate governance actions to accommodate the potentially conflicting privacy needs of multiple people simultaneously occupying a home,

workplace, or public space. Robots will be required to dynamically navigate a matrix of complex privacy settings, customs, culture, and personal needs and, in some cases, the robot may need to ask people nearby for clarification or mediate compromise positions in order to take effective action. Notably, robots share many of these problems with other categories of devices that cohabitate with humans, such as drones and "Internet of Things" devices; therefore, solutions to robot-specific problems have wide applicability spanning many device types.

Despite rapid advancement in the perception, movement, and learning capabilities of robots, issues in robot privacy remain without an effective research program. We propose that technological solutions aimed at the nexus between privacy as a legal and sociological concept and robot control in multi-actor environments can advance the conversation beyond the normative posturing engendered by the information privacy milieu. In that light, this research explores the characteristics of a technical framework for sharing individualized privacy preference data, including a formal taxonomic schema that can be accessed by robotic device makers to inform sensor collection, data collection, storage parameters and constraints, and the permissible range of movements, motions, and activities of a device. Transport mechanisms for distribution are reviewed that allow individuals to "publish" privacy preference data in accordance with the centralized schema and robots to "subscribe to" that data when they encounter the individual. Privacy preference rule selection protocols and merger techniques are developed that allow for the resolution of rule conflicts that arise between individuals in multi-actor environments or ambiguous contexts. Accountability and audit mechanisms are discussed, as are security models for mitigating secondary privacy harms.

An advantage of robot privacy management as conceived in this paper is that it has the capability to move the reasonableness threshold in a positive direction, towards more privacy. In other words, one way of slowing privacy's seemingly inevitable erosion is to provide a method of contextually reactive granular control whereby a person can say "I expect to have privacy here, here, and here—I refuse to give it all away in one broad permission slip." Rather than making one big decision to permit everything, this gives us the chance to make numerous contextually-based decisions that keep some interactions private.

2. Research Scope and Objectives

This paper considers the basic robot privacy problem from an essentially cybernetic viewpoint: aligning command and control of robots with human expectations in an environmental context. We develop a reference technical architecture, or "framework," necessarily incomplete but arrayed as a multi-pronged research agenda, to define structural concerns and implementation options that can assist in meeting the privacy challenges entailed by this new robotic environment. While predominantly a technical framework, this work uses legal and sociological understandings to design a model that exposes systemic assumptions and neutrally adapts norms to account for cultural and contextual subtleties.

More specifically, the objective is to ensure that robot control functions—namely, sensor activation and recording, as well as movement and action—meet the contextually sensitive privacy expectations of individuals coinhabiting the robot's zone of influence. In light of the unique issues involved in robot privacy management, the research agenda is guided by the research questions and design necessities presented in Table 1.

Table 1: Research questions and design constraints for technical framework.

	Question/Constraint
Q1	How can individualized privacy expectations be indicated and communicated while accounting for cultural and contextual nuance?
1.1	How do humans indicate their privacy expectations in a way that robots understand?
1.2	How do humans communicate their privacy expectations to robots dynamically and in real-time, especially when they have had no prior interaction with the robots in a given environment?
Q2	How do robots use privacy expectations to take an appropriate action in a specific setting?
2.1	Consideration: Automated as much as possible to minimize the configuration and interaction burden on individuals.
2.2	How do we ensure that robots always have a path forward--i.e., are always able to make *some* control decision even in difficult ad hoc cases?
Q3	How can robots understand and respect the privacy expectations of multiple people?
3.1	How do we resolve conflicts in privacy expectations when multiple people are involved?

3.2	What design mechanisms allow for negotiation and consensus when the expectations of multiple people conflict?
Q4	Accountability
4.1	How do we ensure that robots (and their designers) are accountable for their adherence to any proposed model?
4.2	How do we gather consent, when needed, for legal purposes?
Q5	Technical Considerations
5.1	Efficient and applicable even to low-cost IoT systems.
5.2	Model is standardized but can be incrementally expanded.
Q6	How do we ensure that data stored in the framework cannot be obtained by actors who can use it to commit privacy violations?

In addition, we define an important simplifying concept known as the "robiota." A "biota" is "the animal and plant life of a particular region, habitat, or geological period" [7]. By extension, we use the term "robiota" to refer to the ecosystem of robot devices—including robots, and potentially even drones and IoT devices—in the immediate vicinity of an individual (see Figure 1). As typically used here, a person's robiota includes any devices that are effectively within sensor detection distance of the person, or that have the capability to move or act upon the person over a parameterized physical distance. With these questions and concepts in hand, we now turn to the technical architecture.

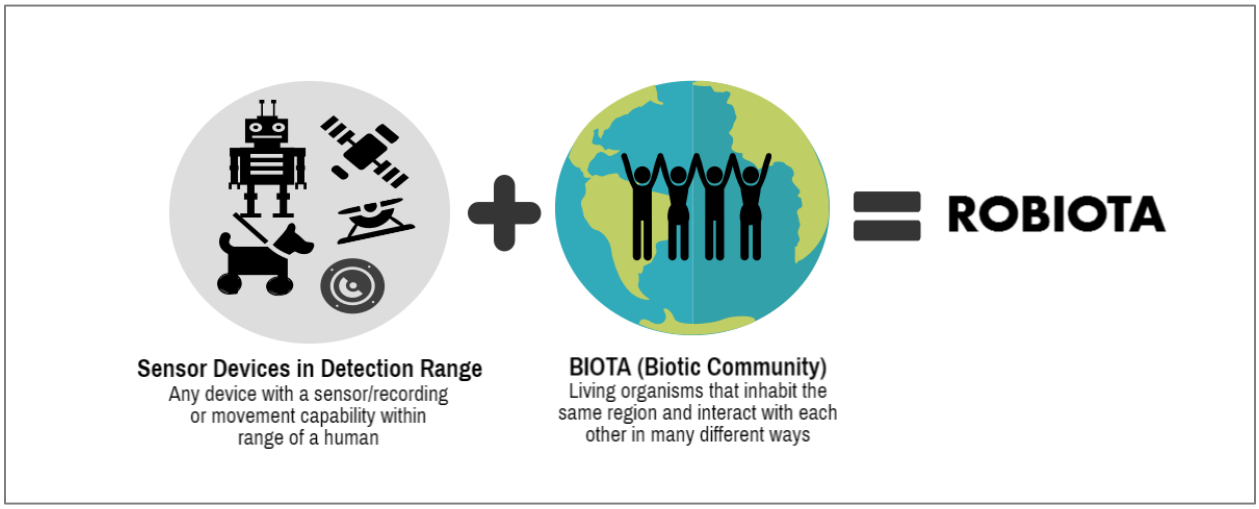

Figure 1: The "robiota" concept

3. Technical Architecture

3.1 Overview of Architectural Model

For robots to be reactive to the privacy expectations of individuals or groups, a system model for privacy preference data exchange is needed. Such a framework has several components, each of which may have multiple design options. Figure 2 shows a basic system architectural model that is used as a reference throughout Section 3.

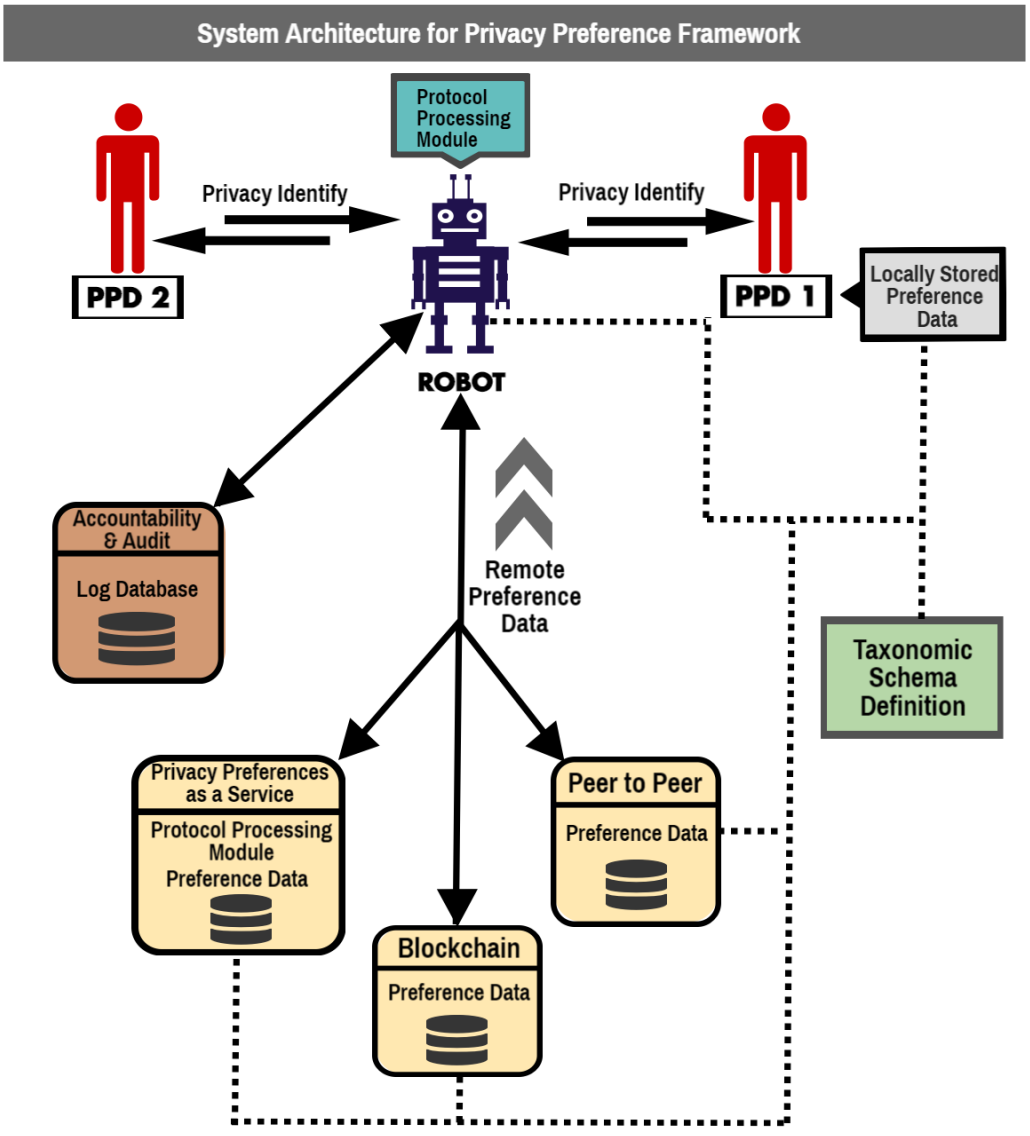

Figure 2: System architecture for the privacy preference framework.

As previously noted, robots need data about the privacy expectations of nearby individuals to inform and control their sensor activity, data collection, storage parameters and constraints, and the robots' permissible range of movements, motions, and activities. The most basic system component is a means for humans to indicate their presence to robots in their robiota ("PPD" in Figure 2). Another primary aspect of the system architecture is a standardized taxonomic schema to define the basic constructs for sharing localized individual privacy preference data with nearby robots. In the functioning framework, individuals "publish" their privacy preference data in accordance with the standardized schema and robots "subscribe to" that data when they encounter the individual. Additional system components needed for a functioning framework are models and transport protocols for the storage of privacy preference data and its interchange between system entities. Several storage and interchange options are possible, including a cloud service model (P^2aaS), peer-to-peer (P2P) model, a Blockchain model, and a local model.

Other system model constituents include processing modules ("Protocol Processing Module" from Figure 2) that enable the robot to retrieve privacy expectation information from preference data storage engines, as well as to sort through, combine, and make sense of preference data originating from multiple actors. These processing modules may execute their instructions on a centralized privacy service, locally on the robot, or a combination of both. When multiple actors' privacy expectations conflict, the robot needs to execute processing logic to resolve the conflicts using techniques ranging from completely autonomous to highly interactive. The system model also describes processing logic and remote storage that enables accountability and auditing of robot control choices ("Accountability and Audit" from Figure 2).

Each of these components and their design options are discussed in more detail in the sections that follow.

3.2 Enunciator

A central aspect of any design that allows robots to make ad hoc, real-time control adjustments to the privacy expectations of a dynamically-changing set of individuals is a mechanism for recognizing the presence of individuals as they move in and out of a robot's sphere of influence. One way individuals may make their presence and privacy preferences known to the robiota is by using a specialized device that serves as a personal enunciator or beacon, also called here a personal privacy device (PPD). Signals sent by the enunciator are detectable by robots within a

given range. This “detection zone” encompasses the robots’ zone of activity or sensor capability (see Figure 3).

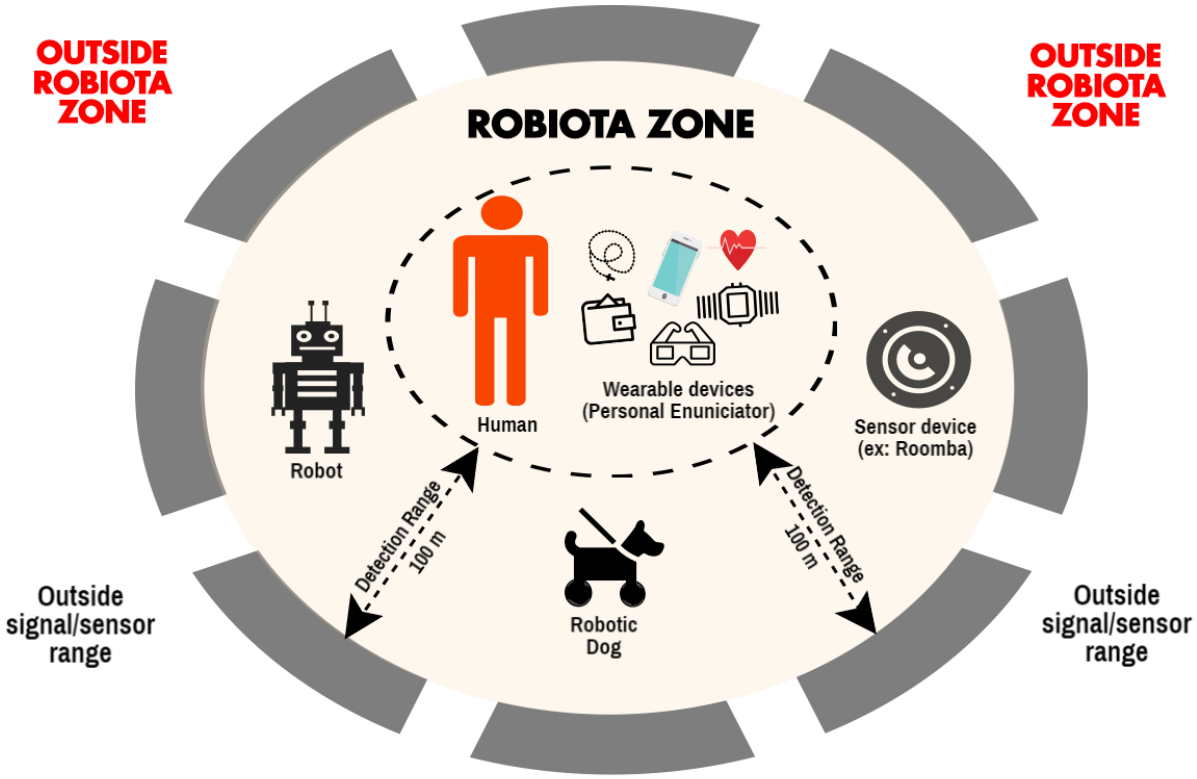

Figure 3: Individuals make their presence known to the robiota with a specialized device that serves as a personal enunciator or beacon, also called a personal privacy device (PPD). Signals sent by the PPD are detectable by robots within a given detection range aligning with the robot’s zone of activity or sensor capability.

A PPD may be embedded in a variety of hardware form factors, ranging from specialized wearable devices (e.g., a ring or other jewelry, clothing) to more generalized devices such as smartphones, smart watches, or fitness bands running enabling software. The appropriate form factor may depend, in part, on which storage models and interchange protocols are capable of providing the level of functionality needed.

One potential enunciator model uses Bluetooth Low Energy (BLE) [8]. BLE is a device-to-device networking technology supported on every major mobile and desktop operating system. BLE uses a variant of the classic Bluetooth wireless technology to broadcast a constant, low power

advertising or enunciation signal to any receiving devices in the nearby vicinity. Receiving devices in the robiota can detect the PPD's enunciation signal, which is branded with an identifier (UUID) unique to the sender of the signal. BLE has a typical maximum range of approximately 100m and consumes very little battery power or processing resources on the device sending the enunciation signal. The physical proximity of the sender can also be approximated by using the received signal strength indicator (RSSI) of the received radio signal. These characteristics (proximity sensing within the relevant distances for most robot activities, identity services, distance estimation, low resource use, and widespread OS support) make BLE a potentially useful technology for implementing a PPD in this framework.

In some cases "collective" enunciators may be used to indicate that the associated privacy preference data applies to every person present in a particular group, organization, or location. Collective enunciators may be valuable when the context governs the privacy environment, necessitating an override of individual preferences. For example, in a courthouse or substance abuse support group meeting, a collective enunciator might instruct robots to stop all long-term sensor recording, regardless of the privacy preference levels of the individuals nearby.

The technology for using beacons to announce the physical presence of devices to other nearby devices is well-established. Additionally, the potential of sensor-filled devices to actively announce and disseminate their own privacy policies over short-range wireless so that individuals nearby can opt-out of tracking has been noted by Langheinrich [9]. However, using beacon-like technologies, as proposed here, to communicate privacy preferences to robots, drones, and IoT devices extant in the real-time, physical environment is a new application that poses new challenges in performance, scalability, and security.

3.3 Storage Architecture

With individuals in the environment using a PPD to advertise their presence to the robiota, robots are aware of their existence and approximate location. Their privacy preference identity is also known, owing to the UUID capability of the PPD as implemented on BLE. The privacy preference identity (which may comprise their real or pseudonymous identity, but may be significantly less than that) can be used to obtain their unique privacy preferences.

The actual content of the privacy preferences remains unknown to the robiota until they are obtained via an organized exchange protocol from wherever they are being stored. There are

several possible architectural models and protocols for the storage and interchange of the privacy preference data, and the choice of which to use will depend on balancing the desired performance, privacy, and security factors. Any given storage and exchange architecture will also impact the range of device types which can serve as PPDs. This section explores several options and their relevant characteristics and tradeoffs, falling broadly into two categories: centrally-mediated and disintermediated.

In a centrally-mediated design, robots use the privacy preference identity (e.g., BLE UUID) to retrieve the person or group's privacy preference data from a central location. The most conceptually familiar model uses a centralized cloud-based architecture, in which a group of internet-accessible servers store the privacy preference data and transmit it to robots on request using common HTTP-based request-response APIs. This model is referred to here as "Privacy Preferences as a Service," or P^2aaS. The scale required for full implementation of this framework likely implies that a large cloud services provider with massive computing power and diverse geographic distribution (on the order of magnitude of Amazon Web Services, Microsoft Azure, Google, or Facebook) is needed.

The cloud service design has the advantage of allowing devices with low processing or storage power to leverage cloud-based resources, and makes available the widest range of PPD form factors because it requires little local processing. This design is also resource-efficient, familiar to users, and provides centralized management and backup. However, it has several disadvantages. For one, wireless networking connectivity to the Internet may be unavailable or poor in some locations or at some times. Second, the functioning of the framework is dependent on externalities such as latency, server load during peak processing times, service scalability, and resilience to cyberattacks. Perhaps most concerning is that the storage of privacy preferences—and, potentially, the privacy behavioral data used in autonomous training models (see Section 3.8)—on a central cloud server has merely pushed well-known information privacy pitfalls up a layer in the architecture. The rich data set is a ripe target both for cybercriminals and for business models that would strive to monetize the privacy preference and behavioral data in some way. This makes the appeal of the proposed framework to robot manufacturers and consumers dependent, at least in part, on trust and confidence in the cloud service provider.

Given these disadvantages, a disintermediated approach may be preferred in some cases to maintain data security or privacy isolation from centralized cloud service intermediaries. In one simple disintermediated approach, the PPD stores the related privacy preference data locally. When a robot needs preference data from a nearby entity, it requests the information and the PPD shares it using a local personal area networking technology, such as BLE, Zigbee, or DotDot. In a hybrid design, a local storage model could be combined with a cloud service model that provides secure preference data backup and redundancy.

Disintermediation can also be achieved using peer-to-peer (P2P) networks. In a P2P network, devices (nodes) join together into networks to share resources. Here, individual PPDs and robots would become nodes in a common network, each node providing an allocation of its storage resources to the network for the collective goal of storing the privacy preference data of all members. P2P networks are generally well-known, and examples in the storage/file sharing arena include BitTorrent and the original Napster music sharing service, now-defunct.

A special category of P2P network is the "blockchain," the idea for which originated with and forms the foundation for Bitcoin [10]. The blockchain is a public database or ledger of transactions consisting of "blocks" recorded by a network of nodes in chronological order. Using the blockchain, a network of nodes can reach consensus on a particular state of affairs and record the consensus in the "blocks" of the distributed network, without any need for a controlling authority. A full account of a blockchain-based privacy preference system is beyond the scope of this paper, but it would be possible to use the blockchain as a storage architecture for privacy preference data and as an auditing mechanism for recording the results of privacy-related decision-making.

3.4 Privacy Preference Taxonomic Schema

Deferred until now is the question of what content comprises real-life privacy preference data. Both content and meta-content are needed to support the features and mechanisms of a nuanced privacy preference–robot control interchange. "Content" in this scenario refers to the actual value of any particular privacy preference setting or rule that might inform an actual robot control instruction in a real scenario. For example, "Turn off audio recording when Alice is at a friend's house." This specific rule defines a specific sensor state for a specific person in a location-determinant context. In theory, it may be possible to enumerate every possible distinct privacy

preference for every person in every situation; however, in practice, a complete set of all possible content can never be found.

For the purposes of framework design, more interesting than the precise content of any particular privacy preference is the structural template or "meta-content." The objective of this section is to explore the key design elements of a functioning taxonomy for privacy preferences in a robotic environment, though it may be more precise to say that the objective is to explore the governing structural template or schema for the taxonomy. The determination of a completed entity classification of privacy preference data is beyond the scope of this work and is the subject of extensive future research.

3.4.1 Rule Form and Components

Discovering the structural template of privacy preference rules requires examining examples with the goal of eliciting their canonical structure. Returning again to the example preference setting for Alice, a recognizable structural form is apparent that may be generalized to the wider category of privacy preference settings relating to sensor control: SENSOR—SENSOR STATE—ROLE—CONTEXT. Although additional rule forms for robot movements/actions and other meta-content scaffolding will also be necessary, it will soon be apparent that this initial structural template for rule taxa provides a surprisingly useful generalization for robot sensor control with respect to privacy preferences. We will now examine the components of this rule structure in more detail.

Robots, drones, and IoT devices have sensors for a variety of reasons: to orient themselves in their environment, to identify important objects or people, to attenuate the force they apply when performing movements or other actions (e.g., grasping), to determine the operating state of a machine or device they are controlling, and to record the movements, actions, sounds, or other telemetry data of people or other entities for historical, accounting, behavioral analysis, and pure surveillance purposes. A robot of any sophistication probably has dozens of sensors representing almost every one of these categories. Depending on how and when a sensor is used and the duration for which it saves its sensor data, any sensor has the potential to violate a privacy preference.

With respect to the sensor entity itself, each model of sensor being used in robotic devices collectively is a sensor taxon in the framework taxonomy. Each sensor model, in turn, has a set of specifications that describe its capabilities, operative ranges, and potential operating states. For

example, the "Waveshare Obstacle Detection Laser Sensor [RB-Wav-23]," available from Robotshop.com, is used in some devices to detect obstacles in the ambient environment or count objects moving by it. This laser sensor has an effective detection range of 0.8m and a maximum range of 1.5m. A different category of sensor, the "BlackBird2 3D FPV Camera," sports two cameras for stereovision (3D) capability. Many models of sensor with varying capabilities exist across the robot sensor parts supplier market.

For representational efficiency and compactness, in some instances of the sensor rule structure SENSOR may indicate an overall sensor capability. A sensor capability is a higher-level descriptor indicating what a sensor can do, such as "Video", "Audio", "Detect Motion", etc. Each sensor model may be grouped into one or more sensor capability categories. For example, when a rule instance such as "Video recording off for Alice in bathroom" is encountered, it is known to apply to all sensor models having the "Video recording" capability. In this manner, any robot that encounters a rule in operating mode will be able to process the rule with reference to its own sensor capabilities irrespective of the exact sensor model the robot is using.

The SENSOR STATE rule structure component represents an additional category of taxa for potential operating states of a sensor. The most simple and basic SENSOR STATE might be ON or OFF. Depending on the sensor model/capability category, more granular states can be described. For example, a video sensor capability could have the sensor states: ORIENT-ONLY (camera is only used for the robot's own orientation purposes), RECORD PERMANENT (a permanent video record is stored on a robot's storage device), RECORD TRANSIENT [TIME] (the video record is kept for the designated time value, then disposed of, and RECORD OFF. As with SENSOR, SENSOR STATE can be represented by higher-level descriptors that encompass more than one category.

Supporting the operation of the putative framework in real-world environments will require the assistance of manufacturers and designers. Parts suppliers for sensors frequently used in robots will need to participate in the codification of sensor models and capabilities according to the taxonomic scheme. In addition to using the framework systems and structures for privacy preference processing, individual robot manufacturers will need to identify and enumerate the sensors and sensor capabilities used in their devices according to the model/capability/state taxonomy so that they can use the framework to obtain and process the relevant privacy rules from localized PPDs.

Another key property of the rule taxonomic structure is that rules should be representable in role-based forms, not merely identity-based forms. This requires a taxonomic structure that can enumerate a privacy identity's roles, and the time, location, or other context in which those roles apply. The reason for this requirement is that, while a PPD and its associated rules represent the collective privacy expectations of a *privacy identity* (e.g., a person or group), a rule itself pertains to the privacy expectations of a privacy identity's *role in an identified context*.

A short example illuminates the distinction: Cathy is a medical doctor who carriers her PPD everywhere with her. She works in a large hospital that uses numerous robots for a wide variety of purposes, including assisting in surgery and medical procedures, administrative functions, maintenance, and counseling. Patients also have privately-owned personal assistance robots that accompany them to the hospital. When Cathy enters the hospital as an employee, she expects that audio and video of her activities will be recorded by any and all robots in the hospital robiota. These recordings are required by hospital policy and maintained for medical malpractice accountability. A possible rule defining a recording rule for her "doctor" role might be: "Video Record Permanent for [Me] As Doctor in Hospital." On the other hand, if Cathy enters the hospital as a patient, she is subject to various legal and ethical constraints on patient privacy that entail different privacy preference rules. A possible rule for her "patient" role might be: "Video Orient-Only for [Me] As Patient in Hospital."

In addition to accurately reflecting real-world scenarios, role-based rule capability has the additional advantage of enabling very generalized preference categories. This allows some rules to efficiently and compactly represent privacy preferences applicable to large numbers of privacy identities. In fact, some rules may be generalized to such an extent that they encode valid "default" rules for certain contexts that largely apply to every privacy identity in that role-context. The doctor and patient rules above, for example, are fair representations of default video recording rules for physician-employees of hospitals, and for all patients of hospitals, respectively. This default rule capability will be valuable both during the initial populating of an operational taxonomy and also in easing the burden of individuals to configure their own preferences.

Despite the potential for default rules, it is to be expected that some assistance may be needed from the owners of PPDs. It will be incumbent upon the owners of PPDs to enumerate and configure their roles/contexts adequately both during initial PPD setup and as new contexts arise.

One possibility for minimizing this burden is for configuration devices present in certain locations or group environments to be used to assist in the initial configuration of individual PPDs with defaults. For example, the "hospital" configuration device could imprint one or more default rules on patients' PPDs when they first enter the hospital or assent to the hospital patient privacy policy. The hospital's configuration device might also imprint "doctor" rules on doctors' PPDs during their new employee orientation process.

The CONTEXT component of the privacy preference rule structure has been touched upon, but not fully explored, in the discussion of roles. The presence of a CONTEXT component recognizes that privacy choices are often context-dependent at a very granular level. For example, whether a person would want to be recorded by a robot may differ depending on whether she is having lunch at a busy restaurant or is conversing with her family in her living room. The notion that individuals' privacy choices are context-dependent has been discussed by Nissenbaum and others in their work on information privacy notice and consent [11] [12].

In this taxonomy, several conceptual layers of context are potentially important. A context taxonomy could be effective with at least the context layers shown in Table 2 with example taxa.

Table 2: Context layers in privacy preference rule taxonomies

Context Layer	Notes and Example Taxa
Cultural	Cultural background of a region, religious affiliation, ethnic group [13] [14]
Societal	Economic system, political structure
Group	Voluntary or involuntary affiliation with a societal segment or group, such as charity, church, advocacy group, support group, political affiliation, formerly incarcerated
Locational	Home, place of employment, private meeting, friend's house, medical facility, region, country, state, city, country
Individual	Pertains to the individual or collective owner of a privacy identity
Situational	Ad hoc situations, emergencies or times when security or safety of self/others is impacted
Legal	Constitutional, statutory, or regulatory constraint, e.g., compliance with a privacy law or judicial/law enforcement rule of conduct

Trust Relationship	Explicit or implicit interaction relationship with the robot entity, such as a robot one personally owns or that inhabits a place of employment; functional relationship, such as a personal care robot

These layers may be organized hierarchically such that some layers are more generalized than others. Lower layers in the hierarchy can override the higher layers when rules are processed by robots to determine control behaviors. This enables the creation of powerful generalized default behaviors that can be used to apply sensible privacy controls to large groups with conceptual and representational efficiency. Meanwhile, the descending layers of hierarchy maintain the capability of the taxonomy to adapt to more granular choices at any arbitrary level of the context hierarchy. For instance, if culture is organized highest in the hierarchy of context layers, locational context rules, if present, override cultural rules for a matching SENSOR—SENSOR STATE—ROLE tuple.

3.4.2 Action Rule Structure

So far, the discussion has centered around the general structure for a privacy preference rule for sensor state/control. However, other forms of robot activity control may be relevant or necessary for a comprehensive approach. In addition to sensor recording, for instance, the physical proximity of a robot may be another factor in humans' perception of creepiness; e.g., when a robot passes a human too closely, the person might anthropomorphize this behavior as rude or dismissive, concluding that it is a creepy violation of personal space. Naturally, context governs what is perceived as "too close"—city-dwellers and some cultures have a higher tolerance for close living than others.

As this example shows, some privacy expectations are met only when control can be exercised over a robots' proximity, motility, location, limb movements, functions, and conversation topics or interruptions that may be considered intrusive. Therefore, it is expected that one or more additional action rule structures may be necessary to enable a complete set of privacy preference descriptors. It is unknown at this time what form such rules would take or how action rules would relate to or be prioritized with respect to sensor rules—let alone the population of the action rule taxonomy. However, several aspects are clear from the discussion of sensor/sensor-state taxa. First, a taxonomy of robot *activities* at the functional level must be developed, likely with the participation of robot device manufacturers. Second, an understanding of the kind of activity

constraints relevant to privacy must be determined. Third, activity constraints must be associated with each functional descriptor.

Considering that context and role remain relevant qualifiers on the action rules, just as they were on sensor control rules, a generalized rule structure may potentially be found that encompasses predicates not only for sensor/sensor state actions, but also for all forms of activity control and location prohibition. This kind of general rule structure may be helpful for achieving notational and grammatical efficiency in the rule processing logic. Even if a general form can be found, taxa for specific activities and activity constraints must be enumerated in addition to their grouping hierarchy/relationships.

3.4.3 Extensibility

Extensibility is another design characteristic of a privacy preference rule taxonomy. As robots begin to be used more frequently, new types of robots will be developed for a wider-range of applications, purposes, and usage contexts. Manufacturers of robot components will develop new sensor devices and even new categories of sensor measurement. As people begin to understand the subtleties of aligning privacy expectations with increasingly complex robot interactions, the taxonomy will have to reflect and specify those nuances on a very granular level. In fact, there is unlikely ever to be a "complete" taxonomy that fully describes every sensor, movement, context, role, etc., for every possible interaction.

Therefore, a key characteristic of the privacy preference framework is that the schema, and the associated methods of processing the schema grammar, should not "break" (i.e., require upgrade) in order to codify new taxa or even to introduce new subdivisions of taxa. Existing robots and PPD devices should be able to download the most current taxonomic schema at any time without concern that changes will affect current operations. The schema should also be self-describing to the extent possible so that the structure of the schema itself can be processed according to automated methods and self-validate. Existing data representation technologies, such as XML and XML Schemas (XSD) [15], are capable of satisfying both the extensibility and self-describing requirements.

3.4.4 Rule Computability and Consistency

A putative taxonomic structure for individual privacy preference rules has been described, as well as the broad notion of hierarchies in rule attribute taxa. In any given situation, numerous

preference rules associated with a specific privacy identity may be applicable to the current context owing to the hierarchical nature of context descriptors. For instance, a "situational" context rule may apply to the current robot context at the same time as a "group" rule. Which rules have priority based on their context? A further concern arises in scenarios where multiple privacy identities may exist in the current environment. How should rules from multiple privacy identities that direct the same robot control functions be merged, evaluated for consistency, and, if necessary, reconciled?

One of the stated design conditions of this framework is that robots always have an available and reasonable resolution path for applying privacy preferences to their functions. In simple design terms, this means that robots can always take *some* action, and not freeze, crash, act erratically, or choose such a poor course of action that it disappoints the expectations of all concerned. More formally, an important property of the privacy preference schema is that the robotic control function outcome of any arbitrary collection of rules associated with the currently applicable context be *computable* and *decidable*.

"Decidability," as understood here, means that all control state processing paths in the privacy preference rule set applied to a single privacy identity lead to exactly one outcome, even if that outcome is a "default" rule at a very high level in the context hierarchy. A more complete description of the robot control state–taxonomy interchange protocol is in Section 3.5, but a short summary here is illustrative: The basic activity of a robot in applying the framework is to decide the most correct sensor state and activity constraint outcomes given the current "operating context" and the privacy preference rules of the localized privacy identities. In other words, a robot's job is to determine the current context, obtain the applicable privacy preference rules from any nearby privacy identities, and process the rules to find the exactly one sensor state/activity constraint that applies to each sensor/activity in this specific context. To more formally represent the sensor rule structure (which is easier to grasp than the activity constraint rule structure): Given a current context *c,* a privacy identity's current role *r*, and a privacy preference rule set P, for each sensor *s* in a robot's set of active sensors S, there is exactly one sensor state outcome *so.*

The condition of decidability further implies that all taxa hierarchies (e.g., for role, context, and sensor group) are internally consistent. This means that there is a clear hierarchy of priority between taxa hierarchy levels wherein lower levels (i.e., more granular levels) of taxa have higher priority than higher levels. This constraint also requires that all rule sets contain at least a single

master default rule for each sensor/sensor group so that if a given rule set does not have any lower level rules for that sensor, some sensor state is determinable. A further corollary to this is that only a single taxon at any hierarchy level applies to the current context. However, it should be noted that this is a formal property of the taxonomy rather than an overall condition of a real-life implementation. In fact, in practice it will likely be quite common that the current context is ambiguous or difficult for the robot to determine. In these situations, the robot must have valid resolution paths for the ambiguity, one of which involves applying the associated rule at the next higher level of the context taxa hierarchy. For example, if the robot cannot determine if it is in a public place or private parking lot, apply the "city" rule. On the other hand, practically speaking, an ambiguity will only require resolution if rules exist for each of the ambiguous options—otherwise the robot would already be applying the higher level rule. Another resolution path may include simply asking the person represented by the PPD, a unique capability of robots that will be discussed later in Section 3.6.

Until now the discussion has centered around the decidability of robot control states from preference rule sets obtained from a single privacy identity (i.e., the preferences of one PPD, person, etc.), and a property of single rule sets is that they be decidable and internally consistent. However, as previously noted, a challenge and a design constraint of the proposed framework is that it be operative in multi-actor environments. That means that when privacy preference rule sets pertaining to more than one PPD constrain the behavior of the instant robiota, preference rule sets emanating from multiple PPDs must be capable of being merged and, when conflicts exist, of being reconciled. Rule conflicts and a protocol for resolution are the topics of Section 3.6, but an introduction to the merger problem and the elements of the taxonomic structure that support merger are relevant here.

Computationally, merger of a plurality of rule sets may have two types of result: a set of control state instructions that *are* in conflict (i.e., the resolved rules direct opposing control states, such as that audio recording be "ON" to satisfy one PPD and "OFF" to satisfy another PPD) and a set of control settings that are *not* in conflict (i.e., the resolved rules indicate the same control state for all PPDs). It is foreseeable that conflict states will occur frequently. Therefore, across multiple preference rule sets, the property of formal decidability simply will not hold when the set of conflicting control states is not empty. However, it is possible to say that the system overall (when functioning across multiple actors) exhibits the property of "relaxed decidability." Relaxed

decidability holds because a robot must be able to take *some* action, even if the robot has to engage in a process of conflict reconciliation that includes dialogue, negotiation, and consensus-seeking with or between the relevant persons (this resolution protocol is a subject for Section 3.6).

A robot ought to be able to combine privacy preference rule strings and quickly determine logically whether the preferences are conflicting or an irresolvable conflict requires a resolution mechanism. Several features of the taxonomic structure can support merger operations, even if the robot has to appeal to higher-order functions (i.e., the resolution protocol) to determine a single definitive control outcome per sensor. These features are loosely characterized here as the "computability" of merged preference rule sets. First, the taxonomic structure should be described in a formal notation that allows for validity checking both within and across rule sets. In single actor rule sets, validity checking would expose internal inconsistencies, such as duplicate rules, conflicting rules, and invalid hierarchy chains. In composite rule sets from multiple actors, validity checking functions would quickly expose potential or actual conflicting control sets. Supporting functions allow the robot to efficiently compute whether a conflict exists in control directives given a context, role, sensor, and simple or compound rule set. Second, the formal notation would also provide the corollary capability of enabling basic logical operations on rules so that they can be combined into complex statements with single truth values. A third, but related, feature provides taxonomic support for rules to indicate allowable resolution options when conflict arises. For instance, properties attached to the person, role, context, or the rule itself could indicate which resolution options are available to the robot in particular circumstances. Advance knowledge of available resolution options may allow the robot to truncate or simplify the standard resolution protocol processing (or automate it entirely) to more quickly resolve conflicts or avoid asking clarifying questions (see Section 3.6.3).

3.5 Algorithms and Protocol

With a realistic taxonomic data structure in hand, we turn now to processing characteristics of a functioning system. This section describes a processing protocol for robot control functions mediated by the privacy preference framework. The overall objective of the processing protocol is to determine which privacy identities fall in the robot's sensor detection range or in the potential zone of impact of its actions, obtain the relevant privacy reference rules from each PPD relating to the context and role, determine if preference conflicts exist between multiple privacy identities

and, if so, engage in a resolution protocol to determine final control states. Then, the robot enacts both the non-conflicting and resolution control states and records the environmental properties and resultant actions for auditing and learning purposes.

Protocol 1 shows pseudocode for a high-level algorithm for this processing protocol. While Protocol 1 is essentially self-describing, some aspects are worthy of further discussion. First, the processing protocol can be iteratively executed by a robot on a timed basis, in response to a control decision that needs to be made by the robot to perform some task, or in response to a trigger or other notification when an enunciator has moved into or out of the robot's detection range.

Protocol 1: Basic processing protocol for privacy preference framework interactions.

Require: R, a robot; S_R, the robot sensor array

ENUNCIATORS $E \leftarrow$ *get_enuciators_within_detection_range* (R)
PRIVACY-IDENTITIES $P \leftarrow$ *get_privacy_identities_from_enunciators* (E)
OPERATING-CONTEXT $c_{op} \leftarrow$ *get_current_operating_context* (R)
for each PRIVACY-IDENTITY $p \in P$ **do**
    VECTOR $d_p \leftarrow$ *determine_vector* (E_p)
    ROLE $r_p \leftarrow$ *query_privacy_identity_role* (p, c_{op})
    for each SENSOR-CONTROL $s \in S_R$ **do**
        CONTROL-RANGE $sr_s \leftarrow$ *query_control_range* (s)
        if *within_control_range* (sr_s, d_p) = true **then**
            RULE $y_{p,s} \leftarrow$ *retrieve_rule_from_hierarchy* (p, s, c_{op}, r_p)
            RULE-SET $Y \leftarrow$ *add_composite_rule_set* ($y_{p,s}$)
        end if
    end for
end for
RULE-CONFLICTS $Y_c \leftarrow$ *validity_check* (Y, "conflict")
for each RULE-CONFLICT $y_c \in Y_c$ **do**
    CONTROL-STATE $t_y \leftarrow$ *resolve_conflict* (y_c) //see Merger Protocol
    apply_control_state (t_y)
end for
RULE-SET $Y_{ok} \leftarrow$ *validity_check* (Y, "no-conflict")
for each RULE $y_{ok} \in Y_{ok}$ **do**
    CONTROL-STATE $t_y \leftarrow$ *get_control_state_from_rule* (y_{ok})
    apply_control_state (t_y)
end for
Log_outcomes (Y_c, Y_{ok}, P, c_{op}, *time*)

Second, although a robot performs the detection of enunciators and drives the logical back-and-forth of the protocol, certain processing activities will cross device boundaries, and the device that performs a given function will depend on the chosen implementation architecture. For instance, the function call to determine privacy identities from nearby PPDs/enunciators (*get_enuciators_within_detection_range)* will be performed by the robot device, as will the function to apply the final control states to the robot's controls. The logic for retrieving a privacy identity's preference rules for sensors (*retrieve_rule_from_hierarchy*) may reside in a centralized cloud service (P^2aaS) available to robot manufacturers or PPD owners, or it could be stored locally on the PPD/enunciator. Validity checking across the compound rule set could be performed locally by the robot itself, or remotely by the P^2aaS. Finally, the rule conflict validity checking function (*validity_check*) is described in more detail in the protocol for conflict resolution (*resolve_conflict*), which will be unpacked in Section 3.6.

3.6 Merger and Resolution Protocol

Section 3.4.4 already touched upon both the impetus for privacy preference rule merger and resolution techniques, as well as upon some of the taxonomy-related design elements. To recap briefly, the overall goal of the merger and resolution protocol is to enable a robot to take appropriate action in spite of the inevitable situational ambiguities and conflicts between the privacy expectations of multiple actors that will arise in every day usage. The merger and resolution protocol's secondary design considerations include: (1) the protocol directs a robot's control behaviors toward contextually normative "best-case" outcomes whenever possible; (2) conflicting actors have the chance to discuss, clarify, and amend their privacy preference positions when necessary; (3) notice of sub-optimal control outcomes, as well as the opportunity to engage in mitigating actions, is provided to actors whose preferences were subjugated; and (4) the protocol acts autonomously, whenever possible, after considering the risk of harm.

Figure 4 shows a prototypical merger and resolution protocol that illustrates actions a robot can take to resolve the control paralysis induced by preference conflicts. Figure 4 groups the actions roughly according to the amount of guidance or intervention required of participating human actors. It is important to note that, while Figure 4 shows a broad default order, a robot's order of stepping through the actions of the resolution protocol may vary in each specific situation. To wit, the availability of some types of resolution actions, as well as their ordering, may depend

on such factors as the context, the actors' roles in that context, the exact nature of the preference conflict, and the resolution option directives in an actor's personal privacy preference data.

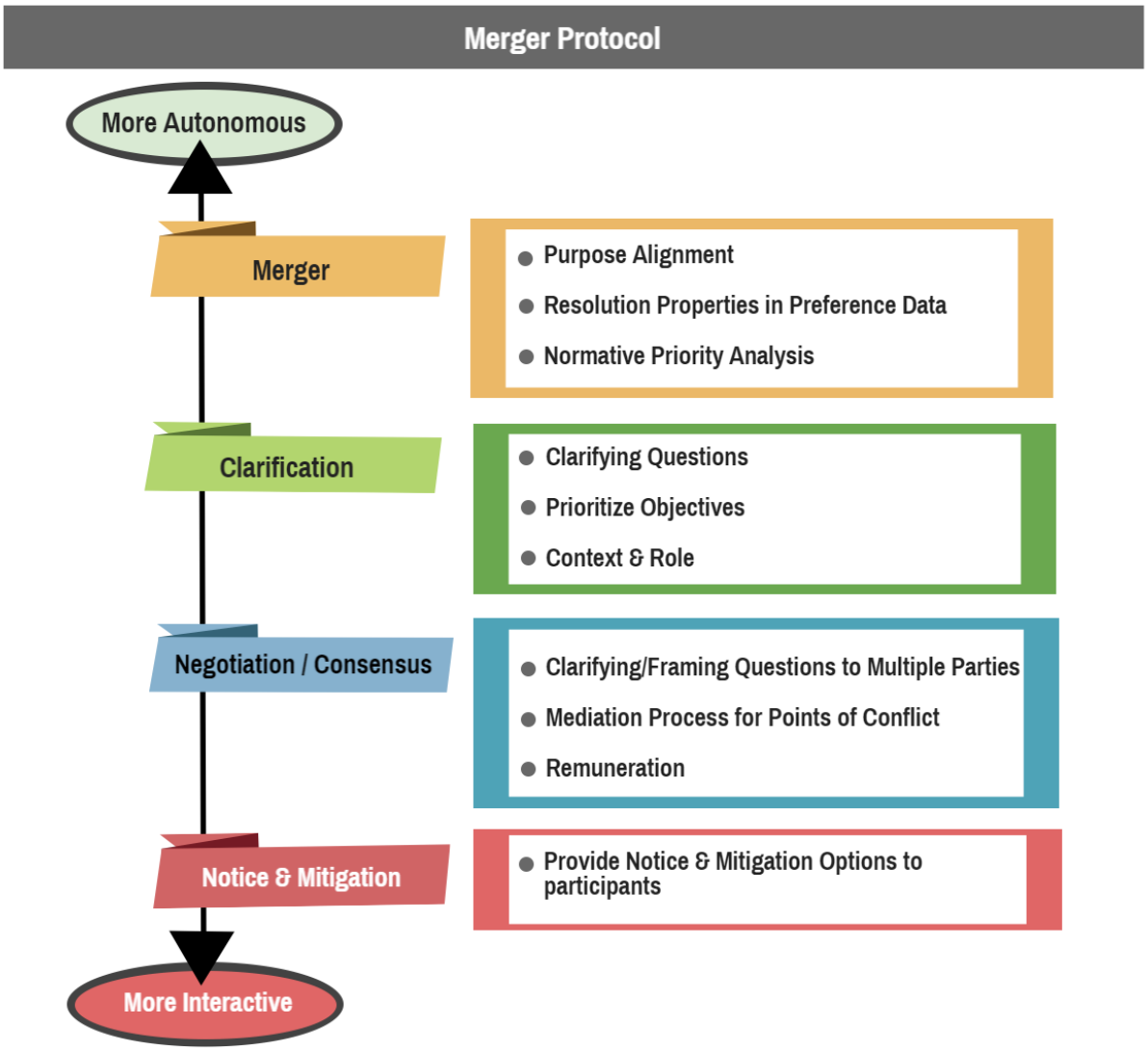

Figure 4: Merger protocol.

3.6.1 Validity Checking and Conflicts

At the outset, it is important to understand the basic meaning of a "conflict" in control directives for the purposes of this framework. Using sensor control rules to illustrate, a conflict arises in at least two ways given the prototype tuple form SENSOR—SENSOR STATE—ROLE—CONTEXT. In the first type of conflict, two or more rules with matching sensor, role, and context content direct a

robot to set conflicting or contradictory sensor states. In most cases, this type of conflict would be exposed quickly by performing validity checking functions of the kind described previously. Conflicts of this type would be exposed in single-actor rule sets at the time the rule was made, whereas in compound rule sets involving multiple actors these conflicts would be determined dynamically in real time. The second and more subtle type of conflict also occurs when the rules of multiple actors direct contradictory control states for the same sensor, but the role and/or context designation maps to a different level in the hierarchy. Table 3 shows examples of both kinds of conflicts.

Table 3: Examples of rule conflicts

Person	Sensor	Sensor State	Role	Context	Conflict Type
P1	Video	Orient Only	Friend	Other's House	Type 1 Direct
P2	Video	Record Persist	Me	My House	
P1	Video	Orient Only	Patient	Hospital	Type 2 Context Hierarchy
P2	Video	Record Persist	All	Public Place	

3.6.2 Purpose Alignment

We now consider each of the protocol's resolution action types from Figure 4. The initial layer of resolution examines the robot's purpose in the environment, as well as the purpose of the specific interaction. First, a robot is presumptively in any given space for a discernable reason—what is the reason? For example, is the robot there to perform a small range of tasks (such as bringing tea or vacuuming), to provide personal services like home health care, to provide companionship, or to greet the public? Second, does the control conflict relate to (or block) a specific robot activity that directly relates to the robot's reason to be there? How necessary is that specific activity to the robot's overall purpose?

The robot can then use these background properties to resolve conflicts by eliminating conflicting rules that so clearly subvert that purpose that they may be safely ignored. For instance, if a robot's purpose is to perform home health care services, then the robot will need to perform specific tasks like undressing the patient for a bath; control directives from the patient's spouse forbidding the robot to enter the master bathroom would, in this hypothetical, be contradictory to

the robot's purpose. The conflicting control directives from the patient's spouse could then be reasonably disregarded to resolve the conflict. In regard to a future research agenda for the proposed framework, it is worth noting that, for such a mechanism to work autonomously, methods of codifying and formally expressing robot teleology must be developed.

3.6.3 Resolution Properties

Appeals to the robot's purpose, while helpful and potentially definitive, will likely not suffice to resolve most control conflicts. Another conflict resolution stratagem (briefly mentioned in Section 3.4.4) concerns taxonomic structures that allow individual privacy identities to indicate their willingness to resolve certain kinds of control conflicts automatically. Resolution properties of the taxonomy can be used both to express the range/ordering of acceptable conflict resolution methodologies as well as to indicate quantitatively the "importance" of a privacy preference rule to an actor. For example, with resolution properties, actor P1 may indicate that a rule preventing a robot from recording video while at a friend's house is very important by assigning it a "5" on a scale of 1-5. Her friend, P2, indicates that she prefers video recording to be "on" at all times in her own home; she assigns it a moderate preference of "3." When the two friends get together, P2's robot can autonomously resolve the rule conflict by turning off video recording based on P1's higher importance rating. Resolution properties also allow actors to assign preference to certain kinds of resolution strategies over others. Continuing the example, in addition to or instead of assigning an importance level, P1 and/or P2 could also indicate a preferred resolution strategy such as "negotiate" (i.e., "help me negotiate a compromise with other actors when this rule is in conflict").

Resolution properties can be assigned to various taxonomic identifiers in the hierarchy—actor/identity, context, role, as well as to individual rules. The resolution properties possesses an expressed hierarchy of priority in that properties attached to the higher levels are more general, and also more overridable, than properties assigned to lower levels of the hierarchy. Individual rule resolution properties supersede role properties, which supersede context properties, which supersede those assigned to the overall privacy identity. In this way, resolution properties can be applied broadly to many control scenarios without losing the capability to target very specific situations when necessary.

3.6.4 Normative Priority Analysis

Some conflicts may be simplified by or completely automatable with the assistance of normative priority analysis. Normative statements generally take the form of a rule of conduct phrased as an imperative; for example, the "Golden Rule" ("Treat others as you would like to be treated.") is a normative statement. Normative statements claim how things "ought" to be and, by doing so, can serve as shortcuts when ambiguities arise in moral/ethical evaluation.

Similarly, normative statements can sometimes be applied to rule conflicts to simplify processing and cut through ambiguities so that the robot can take action. However, normative rules, as used here, remain contextually sensitive. Though a few normative statements may posit "universal" beliefs of humankind, most statements are likely to be based on cultural, national, or religious worldviews. Therefore, the framework's taxonomic schema must be able to support normative rules in automated resolution analysis that are applied from these perspectives in light of the robot's current operating context.

Examples of the kinds of normative statements that may be useful in this framework are shown in Table 4. Consider these rules in action: a robot is in a busy public space with twelve privacy identities currently in its detection range. On average, ten of the identities direct robot control state A, a lower privacy state, while the other two direct robot control state B, a higher-privacy state. However, the privacy identities are generally only in the space temporarily and seldom engage in twelve-way interactive negotiation and consensus. What should the robot do to make a decision on the control state? It can apply the normative rule "the majority state is controlling" and set the control state to A; or, it can apply the normative rule "default to the highest privacy for everyone" and set control state B. Which normative rule the robot decides to apply is likely a function of the cultural values embedded in the operating context.

Table 4: Examples of normative priority analysis rules

Example Rule	**Applicability**
Treat others like you would want to be treated	Universal ?
Prefer one's family over strangers	National
Majority wins	Group, National
Highest privacy setting for everyone	National
Set the highest privacy for the most people	Group, National

Least harm to the most people	Situational
Respect your father and mother	National, Cultural
Owner of a private space's PPD controls	Legal, Processing
Rule designated at the lowest level in the context hierarchy controls	Processing

Some normative rules may identify an implicit or explicit hierarchy of priority based on the privacy identity from which the conflicting rules originate. Encoding the priority rights as normative constructs allows them to be exposed and analyzed. For example, does the PPD representing a private place of business, like a coffee shop, have priority over its customers such that the coffee shop "wins" rule conflicts? Do a parent's rules trump those of her minor child? The principle of privacy identity priority can also be extended to the robot itself, which in some circumstances may represent a non-present privacy identity. For instance, does a law enforcement drone looking for a fugitive outrank conflicting privacy expectations of members of the public? Whether a coffee shop proprietor, parent, or law enforcement drone has these priority rights is likely based on cultural or legal context.

Other rules may resemble processing constructs more than cultural norms or precepts. The last rule in Table 4, "the rule designated at the lowest level in the context hierarchy controls," codifies the normative belief that individuals who designate their privacy preferences at a more detailed or granular level should somehow be prioritized over those who set their privacy preferences less specifically. In practice, this normative rule means that rule conflicts may be resolved by looking at the level of the context hierarchy from which the conflicting rules originate, i.e., a preference rule set at the individual level of the context hierarchy by Alice supersedes a conflicting rule originating from Bob's "national" level. Keep in mind that the rules in Table 4, including processing rules, are merely exemplary and contextually dependent—whether the framework would wish to enact any particular processing rule is another matter. However, a mechanism of this sort that even codifies processing rules has the advantage of exposing system assumptions and biases, which at least allows them to be scrutinized and used in a contextually sensitive way.

3.6.5 Interactivity and Clarifying Questions

In many cases, autonomous methods will fail to completely resolve conflicts between actors' privacy preferences, either because additional clarifying information is needed from an actor about the situation, or because the robot needs to communicate the conflict and give the actors the option

to negotiate a resolution. In Figure 4, these less-automated conflict resolution methods are referred to as "interactive" and are loosely grouped under the labels "clarifying questions" and "negotiation/consensus". Clarifying questions require interactive communication between the robot and at least one actor, whereas negotiation/consensus techniques generally need the robot and all of the conflicting actors to communicate.

An ability to ask clarifying questions allows the robot to request additional information about a human actor's goals or priority of objectives. A robot can also ask further questions when it is having trouble discerning the current context from ambient conditions, or when it is uncertain which taxonomic context maps to the current context. Sometimes, an actor's role in a given context may not be immediately apparent from preference settings provided on the actor's PPD. A robot can vocalize and receive assent to a putative control action when common knowledge assumptions seem to be violated in a given situation. Clarifying communications can even enable an actor to spell out the modalities of resolution he or she is willing to undertake to resolve conflicts.

All interactive methods, i.e., both clarifying questions and negotiation/consensus methods, are dependent on the availability of robust and dynamic methods of ad hoc information exchange between robots and humans. In most cases this means that the robot will need primary natural language interpretation and speech capability. Basic capabilities for natural language processing are becoming more frequent features of robotic devices and will likely become routine features within a few years. Veloso [16] has recognized the need for AI systems to provide plain-language explanation of their activities to improve trust and allow humans to correct their autonomy; Veloso describes a "verbalization space" wherein explanation of activities across multiple layers of detail is possible.

For the purposes of this framework, however, the more direct concern is not the general problem of natural language processing or of representing complex internal states in human language, but the development of linguistic techniques and lexica so that robots can pose clarifying questions, describe control rule conflicts in everyday language, present options to actors, engage in mediation between parties to build consensus, and, of course, understand the reciprocal human dialogue resulting from their speech and translate that speech into resolution actions. In one aspect, these techniques depend on formulating privacy preference rules into natural language grammatical

forms such as interrogatories, imperatives, and subject-predicate constructions that describe control states—a feature of the taxonomy already alluded to in Section 3.4.4.

The development of these linguistic techniques and the methodologies for their continued improvement over time is no small task. However, robots have two important advantages that place them in a unique position to empower the development of these techniques. First, robots have the computational power for robust natural language processing capability. Moreover, we *expect and desire* robots to have language capabilities in a way we do not in other AI contexts. Second, the omnipresence of our own robot devices as they live and work with us enables distinct advantages over other kinds of AI systems we use: robots we "know" can pose hypothetical scenarios to us as part of everyday conversation. This allows robots to subtly learn our preferences from us over time. As a robot learns from us, our privacy preference data becomes more specific, accurate, and nuanced by virtue of the fact that a robot's communication with the privacy preference framework is bidirectional. In addition, this conversational learning capability helps to solve a well-known and difficult general problem in the design of privacy preference user interfaces: how to create organized and accessible privacy preference UIs that people will actually use [17] [6] [5].

3.6.6 Negotiation and Consensus

The basic strategy of negotiation and consensus protocols is to facilitate a conversation between parties whose privacy expectations have conflicts. Negotiation/consensus protocols can be considered to have two sub-stages: multi-party clarifying/framing questions and mediation.

In the first sub-stage, a robot asks any clarifying/framing questions that require answers from multiple actors. These questions are clarifying in that they generally seek information that indicates or illuminates ambiguities in the operating context or conflicting knowledge assumptions. They also act to frame—that is, to make obvious to the human actors with the conflict their own underlying presuppositions about the context or other matters concerning the interaction. Some types of questions that could be asked include:

- clarifications about the priority of actors and rationale (e.g., "Do you think your preferences deserve priority here, and why?")
- the importance of the specific conflicting preferences to each party ("How important is it to you not to be recorded while having lunch here, and why?")
- the actors' willingness to subvert their own preferences for the sake of consensus

Questions may be in a compound form, having both a choice/rating query wherein the answer can be quantified and acted on by the robot, and a free-form explanation query wherein the answer is mostly intended for its framing effect on the other party, not for extensive parsing by the robot. Framing questions in themselves are sometimes sufficient to expose the disparity between positions so that conflicts can be resolved, but even if they are not sufficient they reveal any positional differences between the parties.

The second sub-stage is for the robot to engage the conflicting parties and facilitate a mediation process between them. To do so, the robot may engage in well-known and standard mediation techniques used, for example, in alternative dispute resolution proceedings for legal matters; or, the robot may engage in softer forms of mediation based on other fields of study. The origin of the mediation techniques is less important than that the robot is capable of discerning the conflict points from the rules themselves, the automated merger process, and the clarifying/framing questions. It can then use those conflict points to fill in a well-defined mediation template in a manner ripe for resolution. If the parties are ready to yield on a conflict point, then the robot can enact the new control state and document the outcome of the process (*see* Section 3.7 on accountability).

3.6.7 Remuneration

When attempts at clarification of interests and voluntary negotiation have failed to resolve a privacy preference control conflict, remuneration for consensus is a potential resolution option, as it is in many negotiation scenarios. As the robot has arrived almost at the end of its available options to resolve conflicts, offering an actor the opportunity to yield for value is arguably more fair than the alternative, which is to subjugate the privacy preferences of one or more actors in order to be able to take *some* kind of action. Remuneration provides robots a powerful mechanism to impart a tangible benefit on actors who cannot otherwise agree. Of course, there is a principled objection to whether we should be able to monetize our own privacy in this way in the free market. Hull, for example, argues that the common privacy self-management model pushes privacy into the free market and disrupts the functioning of more socially responsible options [18]. Rather than engage in an analysis of whether privacy should be commodified, we make two observations. First, individuals already do monetize their privacy this way when they use free apps and web services in return for giving the service provider broad latitude to sell their personal information and

behavioral data to advertisers. Second, our resolution rubric places monetization options at the end, not the beginning, of a number of options aimed at voluntary consensus. Indeed, the impetus for this framework is to move privacy preference selection back toward a more socially responsible model than the broad website-style notice and consent methods that will likely be used by robot manufacturers if no such framework is developed.

Many details need to be worked out before a remuneration model could be effective and not burdensome to use. One significant question is Who would pay? Robot manufacturers have an interest in their robots respecting the privacy of their customers and third parties, while at the same time not becoming paralyzed by control conflicts. As a result, manufacturers might reward consumers with money, discounts, robot repair credits, or other media of value to encourage consensus. Alternatively, the payor might be the actor whose privacy preferences "win" the conflict. An actor-payor system allows those actors who highly value a given preference outcome to encourage consensus in their favor. To minimize unfairness resulting from wealth disparities, an actor-payor system could be built on a platform of micropayments wherein the exchange medium is not real money but "privacy credits." Actors could use privacy credits gained during negotiations where they were more flexible about the outcome to fund their desired outcomes during other negotiations. A decentralized value exchange platform such as the bitcoin blockchain [10] is an intriguing architectural foundation on which to base a privacy credit micropayment system. Another detail to be worked out is the taxonomy support for properties that allow a PPD owner to specify acceptable remuneration amounts for "yielding" on particular rules or in particular contexts.

3.6.8 Notice and Mitigation

In some—hopefully rare—circumstances, a robot may have failed to resolve preference conflicts using autonomous or negotiated modalities. Still, the robot must act. The robot's remaining resolution strategy is to determine a final control state and provide notice and, if possible, mitigating actions for those actors whose preferences were unavoidably subjugated. To determine the "best" final control state, the robot may revisit normative rules that were considered during the earlier autonomous processing. For example, the robot may implement the rule that enacts the strictest privacy control state, or imparts the best privacy outcome to the most people. In cases where the best outcome according to any normative rule is indeterminate, the robot may

simply make a random choice between the conflicting control states. Once the final control state is known, the robot should communicate its choice to the affected persons along with any mitigation options. Mitigation options could include providing a lag time before enacting the control so that the affected person can leave the area or otherwise prepare herself. As an example, P1 owns a personal service robot he has permitted to enter a room where he is nude, but P2 (P1's spouse) has elected not to allow this. Lacking any other way to resolve a rule conflict, the robot might say: "Your husband has been in the shower for too long; I need to open the door and move through the bedroom to check on him. I will open the door in 10 seconds, so please cover yourself." Another kind of mitigation option could be for the robot to mask sensor detection of the subjugated party's face or voice by blurring the video or transforming the audio with phase shift/reverb effects.

3.7 Accountability: Logging and Audit

It has been a fundamental tenet of this proposal that one key to the widespread adoption of robots is overcoming the "creepiness" factor by providing effective and transparent technical methods for respecting individuals' privacy expectations. Commenters have also increasingly come to believe that autonomous systems need mechanisms whereby their decision-making processes are more transparent to outside agents (whether human or automated) (*e.g.*, [19] [20]). Extending that reasonable premise to the proposed privacy preference framework, it becomes clear that robust accountability mechanisms are essential given the number of interacting inputs and the complexity of the decision-making processes as a robot moves through rule determination, merger protocols, and action determination. Moreover, any specific instance of rule conflict processing may involve negotiation and consensus between actors who make ad hoc decisions or compromises, give consent to override previously indicated preferences, or provide additional information the robot must accept as true. This ad hoc external information must be documented not only for liability purposes, but also because the legal constraints governing privacy demand actor notice and consent for some kinds of privacy intrusions.

Strong instrumentation of the input variables, assumptions, and decision-making processes are essential to transparency, to any improvement methods based on machine learning, and ultimately to machine and manufacturer accountability. Based on the system architecture thus far, it is proposed that framework processing protocols be built from the ground up for instrumentation of

at least the data elements, decision points, and external influences shown in the representation in Figure 5 for every iteration of a robot's control state evaluation protocol.

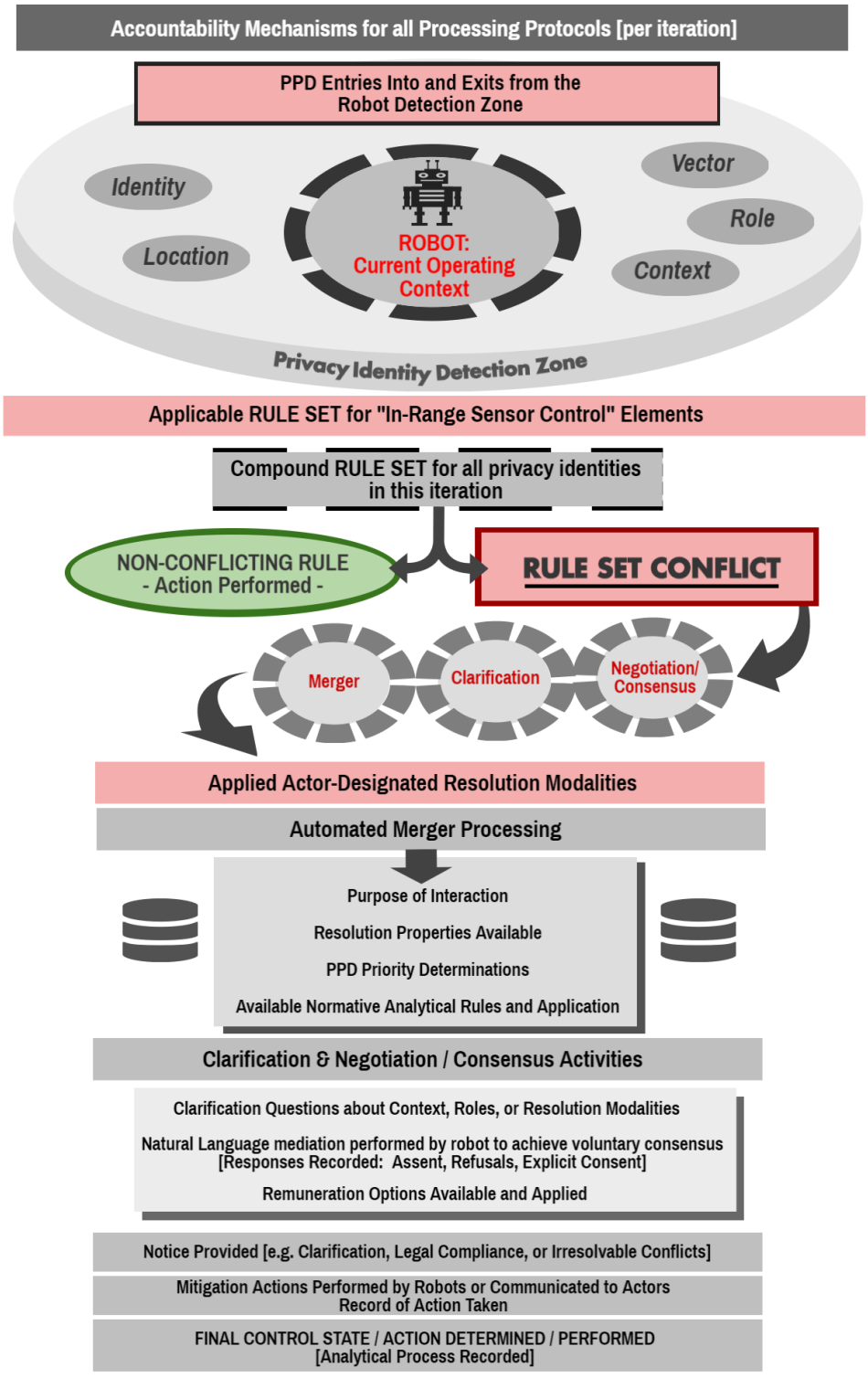

Figure 5: Accountability mechanisms and associated properties.

In addition, the provider of privacy preference data—whether the provider is a centralized privacy preference service, a PPD, or other decentralized model—logs all data requests by robots and the privacy preference data that was provided in return. PPDs also record the identifiers of robots entering and exiting their own robiota.

Instrumentation such as this allows accountability and compliance auditing to flow simply and naturally from system design. Manufacturers that participate in the privacy preference framework agree to comply with responsible use policies and to design their robots to act in compliance with its outcomes. In most cases, compliance auditing is confined largely to analyzing whether the control actions the robot took fit the parameters. The codified nature of the processing, as well as of the instrumentation, means that autonomous systems can perform most audits automatically. In automatic auditing, an autonomous system randomly selects a control state evaluation protocol instrumentation log from the pool of instrumentation logs periodically uploaded to a central repository by robot devices. The autonomous agent virtually performs the scenario as the robot performed it and determines if the activities and outcomes plausibly follow from the conditions. Manufacturers who fail audits are subject to incrementally escalating disincentives.

Draconian as compliance auditing may seem, it is worth noting that some robot device makers will sign on to use P^2aaS services, in which case protocol processing will be performed at a remote layer of the architecture, and thus not be subject to flawed implementation or other manipulation that would generate auditing violations under most circumstances. Moreover, the dynamic and ad hoc nature of processing in real-life environments means that many times privacy outcomes will be conditional on actors' responses and other externalities. Therefore, heavy instrumentation of these externalities plays an important role in a robot device manufacturer's ability to show compliance for liability purposes.

3.8 Security and Privacy of Preference Data: Access Control, Trust, and Privilege

One potential area of concern relates to the privacy and security of privacy preference data itself. The act of making available privacy preferences and privacy-related behavioral data to the range of subscribing devices in the robiota may itself have security and information privacy pitfalls. Privacy preference rules, roles, contexts, and associated options encode a variety of sensitive data about the underlying actor manifested in the privacy identity. If uncontrolled, this information could be used by malevolent devices to obtain important personal information about

the actor, his usual activities, and his relationships with others. Preference data contains enough data points to allow advertisers to build a shadow profile containing the desires, proclivities, and associations of an actor. The data could also be used to foil individual choices for anonymity or pseudonymity, in much the same way that website behavioral data is cross-matched to other data to deanonymize users on the web today.

Various technical methods may be used to help address these concerns. At the outset, we recognize that the visibility and level of detail of the preference data provided in a given situation is a function of the level of trust a privacy actor has in the requesting robot or device. A PPD's private taxonomy might encode devices of high, medium, and low trust. A person's own devices might have high trust, for example, while workplace or friends' devices are medium trust and the robots one encounters in public places are low trust. The various trust levels can then be used to segment the data returned about a privacy identity into categories of more or less quantity, detail, elasticity of usage, and permanence.

In terms of quantity and detail, most privacy preference data requests should be need-based. That is, only in rare circumstances of extremely high trust and explicit purpose should a device be able to request all or a significant share of a PPD's associated privacy preference data. In most scenarios and with devices of all but the very highest trust levels, processing logic of the framework should only yield preference data relating to the current operating context. PPDs interacting with low trust devices may even perform context validation to ensure that the low trust device is not "spoofing" an operating context in order to obtain additional or more sensitive preference data. A PPD or central privacy preference service can, for example, validate the low-trust device's issued context against its own determined operating context using environmental conditions, or against the context determined by other devices in the same robiota. Furthermore, in most cases, the rule sets returned should also be validated against the canonical sensor/control taxa enumerated by the manufacturer of that robot type; this ensures that devices cannot pretend to have sensor/control capabilities in order to have access to more preference data.

Trust levels may also limit the elasticity of usage of privacy preference data. A robot with low-trust privilege may only be able to obtain control state outcomes relating to the operating context for its canonized sensor/control taxa. In contrast, devices with higher privilege might be able to process custom control actions based on access to raw preference rules, perhaps even for non-

canonized control types. Access to richer data can enable a trusted robot to perform stages in the merger protocol with greater accuracy. Potentially, the taxonomic schema might support customization of the available resolution modalities, remuneration values, or resolution properties based on trust level. For example, an actor might choose not to allow low-trust public robots to use remuneration as a resolution modality. Relatedly, access to read privacy preference behavioral data should be limited to devices that have a valid purpose for doing so, such as highly trusted personal devices. The ability to write behavioral data back to the privacy preference data set for direct use or training algorithms may be similarly constrained by trust levels.

Techniques may also use trust levels to determine the permanence or persistence properties of data. For instance, lower trust devices "forget" the preference data they obtained about any PPD when they are out of range of the PPD, mitigating the security problems associated with long term data storage. Trust levels may also determine whether accountability logs are written directly to a central audit log repository instead of being stored on the robot for a time. Also, to maintain privacy and security of accountability and audit data, auditing mechanisms can be built such that they anonymize or otherwise obscure the underlying privacy identities. For legal compliance, split-data logging techniques can be used so that control state evaluation protocol instrumentation logs include transaction identifiers written only to PPD logs; this way an actor, through her PPD, can find associated notice and consent data, but a robot or auditor cannot determine PPD identity directly from the log data.

4 Cataloging Contextually-Based Privacy Expectations

Most of the analysis thus far has centered on the structural design and functional constraints important to a comprehensive technology-oriented approach to the robot privacy problem. We have developed a system architectural "shell" that is ready to codify and exchange privacy preference data to use for robot control functions, but as yet have not identified the specific preference data we would put into the system data structures, or even a means for determining the preference data and continually improving its quality. To put it another way: What methods can be used to catalog the contextually-based privacy expectations of human beings, and how can we encourage humans to designate their individual preferences? While these questions cannot be definitively answered here, a brief review of research in this area is helpful to understand some aspects of the possible answers.

The first strand of research relates to the general question of privacy-related taxonomies and the usability of preference systems. In the area of information privacy as it relates to privacy policy notice and consent, Cranor [6] recounts early efforts to use technical methods to help simplify and automate user choices. The Platform for Privacy Preferences (P3P) was initiated in the late 1990s to produce a vocabulary of privacy policy terms for codifying "compact policies" so that web browsers could read the privacy policies and take action of users' behalf without interfering with browsing [6]. When Internet Explorer cookie blocking became tied to P3P, many companies (e.g., Amazon and Facebook) began active circumvention of compact policies due to its perceived inability to represent the full range of notice and consent semantics [6, p. 296]. Ultimately, Cranor concludes that P3P failed not because of difficulties with vocabulary taxonomy or technical issues, but lack of incentives for consumers and industry to adopt it [6, p. 295]. Efforts such as eTRUST (later, TRUSTe) and IAB CLEAR have pursued the goal of XML-based privacy policies with varying degrees of success. The Usable Privacy Project [21] has the goals of performing privacy preference modeling based on the most salient privacy policy constructs and using natural language processing to automate the comprehension of privacy policies. More recent research by Schaub et al. [5] describes a taxonomy of privacy notice approaches that can be used by designers to build more effective and usable user interfaces.

Nissenbaum has outlined a theory of contextual integrity for information privacy that describes information flows as dependent on the context in which an information request is made, the role played by the agent doing the requesting, and the type of resource being accessed [11]. Although, as previously noted, the information privacy sphere differs in several ways from the robot privacy problem as understood here, subsequent researchers have followed the contextual integrity inquiry to develop two potentially fruitful strains of research for our privacy preference framework. Wijesekera et al. performed user studies that show, first, in contrast to current mobile device permission models, users do in fact make their privacy decisions in a contextually-based manner and, second, that machine-learning models could be built that can accurately predict users' privacy decisions 95.7% of the time [12]. Shvartzshnaider et al. investigated using crowdsourcing methods for automating the discovery of contextual norms, showing that survey questions could successfully be generated to target privacy norms in any context [22].

Soares and Fallenstein have approached the problem of human value alignment from the perspective of trepidation about "superintelligent" AI systems [23]. Their research agenda touches

upon two areas that have particular relevance here. The first area concerns the problems of inductive value learning and operator modeling—or, how to build intelligent agents to learn values from training data and how to model an operator so that the operator's preferences can be extracted [23, p. 10]. The second area involves the question of building a reasoning model to ensure that AIs do not develop their own orthogonal incentives to manipulate or deceive humans. More pointedly, will it be possible to develop AI learning systems that are utility indifferent, i.e., capable of switching their preferences on demand without polluting the outcomes with their own goals [23, p. 9]? For our framework, Soares and Fallenstein's first research area relates to the automated determination of values and norms, and the second to aspects of the merger protocol involving goal clarification and normative analysis.

5 Conclusion

In this paper, we began with the thesis that the omnipresence of robotic devices in our environment gives rise to unique privacy problems unlike those in other domains. To solve those problems, we delineated a technical framework that includes: the overall components of a system architecture, taxonomic data structures for mapping privacy preferences to robot control states, a processing protocol for enacting robot control states and identifying preference conflicts between multiple actors, automated and interactive methods of resolving those preference conflicts, accountability and audit mechanisms, and methods for ensuring the privacy of preference data itself.

Looking ahead, additional research and coordinated action is needed. Although our taxonomic structure and processing architecture are built hierarchically so that a functioning control rule set can be developed from very few rules, an approach to the problem of populating the data structures of the framework—particularly the taxa for context hierarchies, roles, and the default control state rules that associate with them—must be selected from the available strategies or developed anew. Usable methods of enabling individuals to understand and self-select their own rules are also needed, though many robotic devices will provide the distinct advantage of natural language capabilities to facilitate users' vocalization of their preferences. For instance, a robot could assist its owner to configure his or her PPD preferences by engaging in a templated, but dynamic, dialogue with them over a period of time. In addition, research on the appropriate techniques to use for ongoing autonomous improvement across several categories is needed: learning/refinement

of individual preference rules and broad default contextually-based preference rules, system training to improve the recognition of operating contexts and the mapping of goals to outcomes, and learning techniques that help robots better ask and understand clarifying questions and facilitate negotiation.

The overall success of the framework's approach will depend on widespread input, support, and adoption by robotics industry stakeholders, standards bodies, researchers, and the general public. Fortunately, many benefits will accrue from approaching the robot privacy problem from the coordinated perspective described here. Many of these advantages have been touched upon before, but a few are worth reiterating here. First, for individual users, the proposed model presents a comprehensive technological structure that empowers people to make choices reflecting their cultural and personal values. The PPD-enunciator concept helps individuals maintain and configure consistent privacy preference settings across the totality of the robiota. Instead of defaulting to lowest common denominator approaches, robots can use these personalized settings to achieve optimum privacy expectation alignment even in ad hoc and unforeseen privacy scenarios, and even when multiple actors are involved. A rich and interactive resolution protocol ensures that users are able to engage in participatory and adaptive notice, negotiation, and consent activities when conflicts do arise.

Second, manufacturers have much to gain from a common, standardized approach to meeting individuals' privacy expectations. Public comfort that robots are attempting to be respectful of their privacy choices in most situations will help stave off the perception that robots are surreptitiously watching and recording them, and hence do much to drive marketplace adoption of these technologies. Furthermore, agencies such as the FTC are likely to become increasingly active in regulating consumer privacy in robot devices, and consumers are increasingly likely to litigate against manufacturers for privacy harms. By *actually* approaching privacy design in a detailed and thorough manner—a design that respects people's choices, documents their consent, and in which the robot chronicles and is accountable for its decision-making—robot device-makers erect a powerful natural defense against regulators and other legal challenges. Finally, by widening our vision from robot privacy momentarily, a functioning framework of the type described here could be extended to create a generalized human expectation–robot control language and supporting architecture for all sorts of robot-human interactions. Robot makers and researchers would benefit from a generalized solution to the problem of how robots can describe their activities to humans

and meet their expectations, as well as how humans can interact/converse with the robot to make their expectations known.

Society and policy stand to gain from an approach enabling a renewed conception of privacy that re-centers itself on the individual. Instead of the privacy-eroding blanket "consent" paradigm that has grown up around current web monetization models, a new paradigm matures—one where individuals are able to make fine-grained choices about privacy expectations that are specifically and painstakingly enacted by robot devices. An individual's option to make a choice and have it respected, in itself, re-energizes the legal conversation about the "reasonable expectation" of privacy in almost any domain. Our approach creates another type of empowerment of the individual arising from the remuneration model for resolution of conflicts: the rewards for subverting our privacy goals at least accumulate to individuals, rather than to ad networks and big data brokers. Furthermore, robots and their makers are responsible for and accountable to individuals because it will be possible for an individual to review the accountability instrumentation to see why a robot made the control decision it did. Also, since our design attempts to be contextually-neutral about normative analysis, value judgments, and rule hierarchies, it encodes those evaluative and processing constructs in a way that exposes hidden assumptions and biases. These accountability attributes represent a significant improvement to the way that automated systems conventionally operate—as "black boxes" that provide no insight into their contrivances.

REFERENCES

[1] "Privacy 'impossible' with Google Glass warn campaigners," BBC News, 26 March 2013. [Online]. Available: http://www.bbc.com/news/technology-21937145. [Accessed 16 March 2017].

[2] H. Kelly, "Google Glass users fight privacy fears," CNN, 12 December 2013. [Online]. Available: http://www.cnn.com/2013/12/10/tech/mobile/negative-google-glass-reactions/. [Accessed 16 March 2017].

[3] R. Carroll, "Goodbye privacy, hello 'Alexa': Amazon Echo, the home robot who hears it all," The Guardian, 21 November 2015. [Online]. Available: https://www.theguardian.com/technology/2015/nov/21/amazon-echo-alexa-home-robot-privacy-cloud. [Accessed 16 March 2017].

[4] J. L. Mills, Privacy: The Lost Right, New York: Oxford U. P., 2008.

[5] F. Schaub, R. Balebako, A. L. Durity and L. F. Cranor, "A Design Space for Effective Privacy Notices," in *USENIX Association Symposium on Usable Privacy and Security*, 2015.

[6] L. F. Cranor, "Necessary But Not Sufficient: Standardized Mechanisms for Privacy Notice and Choice," *J. on Telecomm. & High Tech. L.,* vol. 10, pp. 273-308, 2012.

[7] Oxford Dictionaries, "Biota," [Online]. Available: https://en.oxforddictionaries.com/definition/biota. [Accessed 14 March 2017].

[8] Bluetooth SIG, "Bluetooth Low Energy," [Online]. Available: https://www.bluetooth.com/what-is-bluetooth-technology/how-it-works/low-energy. [Accessed 1 March 2017].

[9] M. Langheinrich, "A Privacy Awareness System for Ubiquitous Computing Environments," in *Proc UbiComp '02*, Springer, 2002.

[10] S. Nakamoto, "Bitcoin: A Peer-to-Peer Electronic Cash System," 1 November 2008. [Online]. Available: https://bitcoin.org/bitcoin.pdf. [Accessed 11 March 2017].

[11] H. Nissenbaum, Privacy in Context: Technology, Policy, and the Integrity of Social Life, Stanford Law Books, 2010.

[12] P. Wijesekera, A. Baokar, L. Tsai, J. Reardon, S. Egelman, D. Wagner and K. Beznosov, "The Feasibility of Dynamically Granted Permissions: Aligning Mobile Privacy with User Preferences".

[13] E. Dincelli and S. Goel, "Can Privacy and Security Be Friends? A Cultural Framework to Differentiate Security and Privacy Behaviors on Online Social Networks," in *Proceedings of the 50th Hawaii International Conference on System Sciences*, 2017.

[14] E. Dincelli and S. Goel, "Research Design for Study of Cultural and Societal Influence on Online Privacy Behavior," in *Proceedings of 2015 IFIP 8.11/11.13 Dewald Roode Information Security Research Workshop*, 2015.

[15] "Extensible Markup Language (XML)," W3C, [Online]. Available: https://www.w3.org/XML/. [Accessed 16 March 2017].

[16] S. Marquart, "Complex AI Systems Explain Their Actions," Future of Life Institute, 28 November 2016. [Online]. Available: https://futureoflife.org/2016/11/28/cobots-manuela-veloso/. [Accessed 11 March 2017].

[17] D. J. Solove, "Introduction: Privacy Self-Management and the Consent Dilemma," *Harvard Law Review,* vol. 126, pp. 1880-1903, 2013.

[18] G. Hull, "Successful Failure: What Foucault Can Teach Us about Privacy Self-Management in a World of Facebook and Big Data," *Ethics and Information Tech.,* vol. 17, no. 2, pp. 89-101, 2015.

[19] N. Diakopoulos, "Accountability in Algorithmic Decision Making," *Communications of the ACM,* vol. 59, no. 2, pp. 56-62, 2016.

[20] F. Pasquale, The Black Box Society, Cambridge, MA: Harvard Univ. Press, 2015.

[21] N. Sadeh, A. Acquisti and T. D. Breaux, "The Usable Privacy Policy Project. Technical report, CMU-ISR-13-119," Carnegie Mellon University, 2013.

[22] Y. Shvartzshnaider, S. Tong, T. Wies, P. Kift, H. Nissenbaum, L. Subramanian and P. Mittal, "Learning Privacy Expectations by Crowdsourcing Contextual Informational Norms," 2016.

[23] N. Soares and B. Fallenstein, "Agent Foundations for Aligning Machine Intelligence with Human Interests: A Technical Research Agenda," Machine Intelligence Research Institute, 2014.