Toward an Ontology of Regulation: Socially-Based Support for Coordination in Human and Machine Joint Activity

Paul J. Feltovich, Jeffrey M. Bradshaw, William J. Clancey, Matthew Johnson
Florida Institute for Human & Machine Cognition (IHMC)
40 South Alcaniz Street, Pensacola, FL 32502, USA
Email: {pfeltovich, jbradshaw, wclancey, mjohnson}@ihmc.us

Abstract. In this chapter we explore the role of regulation in joint activity that is conducted among people and how understanding this better can enhance the efforts of researchers seeking to develop effective means to coordinate the performance of consequential work within mixed teams of humans, agents, and robots. Our analysis reveals challenges to the quality of human-machine mutual understanding; these in turn set upper bounds on the degree of sophistication of human-automation joint activity that can be supported today and point to key areas for further research. These include development of an ontology of regulatory systems that can be utilized within human-agent-robotic teamwork to help with mutual understanding and complex coordination.

Keywords: Coordination, culture, human-agent-robotic systems, joint activity, ontology, policy, predictability, regulation, teamwork

1 Introduction

One of the most important prerequisites for joint activity among people—and indeed for the functioning of human cultures—is the presence of regulatory systems by which such activity can be coordinated [13]. The kinds of joint activity we have in mind run the gamut of life, including processes as diverse as a conversation, a couple dancing, driving on a busy highway, and military operations.

The essence of joint activity is interdependence—that what party "A" does depends on what party "B" does, and vice versa (e.g., "One if by land, two if by sea" in Longfellow's account of Paul Revere's famous ride). As soon as there is interdependence, there is a need for coordination in time (e.g., timing a live, multi-party phone call) and/or space (e.g., designating a drop-off point), which in turn requires some amount of predictability and order. Such order has been described by Rousseau as the self-crafted bedrock of successful societies:

…the social order is a sacred right which serves as the foundation for all other rights. This right, however, since it comes not by nature, must have been built upon conventions." [31, p. 170].

In this chapter, we will explore joint activity and the kinds of "conventions," to use Rousseau's term, that serve to bring the order and predictability necessary for coordination among people, although we will call these "regulations" (for reasons that will be explained). By *"regulation,"* we mean any device that serves to constrain or promote behavior in some direction.

We conduct our investigation by identifying and unpacking the many and diverse systems of regulation that humans create and employ for achieving order. In what follows, we first discuss the nature of joint activity (Section 2), and then characterize some of the many culturally-based regulatory systems and their role in joint activity and coordination (Section 3). Finally, sections 4 and 5, respectively, outline some of the research challenges and applications of this work in the context of human-agent-robot teams conducting complex operations, such as space exploration, disaster response, and military operations [3; 5]. In the Appendix, we provide a brief illustration of major categories emerging in our ongoing effort to develop an Ontology of Regulation.

2 Joint Activity

Joint activity, interdependent activity, is important even to animals, which in some ways are better analogues than humans for the limited intellectual capacities of software agents [16]. One important way in which animals accomplish joint activity is through coordination devices, including signaling and display behavior. Animal biologist John Smith has identified ten signal types that he claims are nearly universal to (at least vertebrate) animals, although the specific manner of expression may vary across species [6; 16; 32; 33]. These are simple signals, basic to coordination: e.g., "I am available to take part in joint activity," (or not), "I am going to move," "I am monitoring something important," or "I am under attack." Such signals decouple action from intention and provide opportunity for other parties to join in or stay away.

With the increased complexity of our joint activities, predictability in support of coordination is even more important to humans. Cultural anthropologist Geertz [19] has argued that because of humans' under-determination biologically relative to lower animals, and because of our larger repertoire of behavior, we are in even greater need of means of coordination. Notably, for the most part, we are left to fashion these means ourselves. That is, we need to learn and be taught how to live, interact, and control ourselves—hence our relatively long apprenticeships under parents and other educational and training influences. In this view, human *culture* itself is a vast fabrication of *regulatory systems* for guiding and constraining behavior, especially interdependent behavior (see also [20] regarding "guided doings").

Among humans, we distinguish roughly three types of joint activity. These are

based on differences in the nature of their points of interdependence, in particular, interdependence among necessary resources only, among actions, and among motivations and goals [10]:

Sharing: This is characterized by interdependence among necessary resources only. Parties have independent goals, and there is no functional coupling of methods. An example is two groups trying to schedule a conference room they both need to use on a certain day. In sharing, constraints on resource allocation require negotiation.

Cooperation: In cooperation, there is interdependence of activities but not of motivations and goals. Often there is also interdependence of resources. Following the last example, two groups trying to conduct their own meetings within the same room at the same time would be a cooperation. So also, interestingly, are competitive games, such as football, where the two teams' actions are clearly interdependent while their aims are not the same and even contrasting.

Collaboration: Shared project objective is the hallmark of collaboration [11]. All parties are trying to achieve the same end (mutually defined), and there is also usually interdependence of actions (often involving different roles) and resources. Team members within *one* team in a football game (or a relay team in track and field) fit this description, as does a group of scholars working together to produce a genuinely multi-authored article on a topic of mutual interest.

2.1. Key Aspects of Joint Activity

We have asserted that one of the major purposes of regulation is the predictability and order it provides to support *coordination* of interdependent activities within joint activity. To make this coordination possible, participating parties need to 1) know some things in common with regard to their activity and, 2) use what they know in common to coordinate their interdependent interactions and moves. Following linguist Herbert Clark, we call the pertinent shared understanding "Common Ground" [14]. Common Ground consists of all the knowledge, beliefs, assumptions, and presuppositions that parties have in common with respect to their joint activity.[1] These knowledge components include the pertinent regulatory systems that apply to their joint activity, as well as related coordination devices that can be used to navigate coordination. Before addressing these directly in Sections 2.2 and 3, we first present other important components of successful joint activity.

---

[1] In itself, Common Ground makes no claim about quality; common ground may be well- or poorly-tuned to the joint activity. Parties need to maintain common ground that is good enough to at least keep the joint activity moving forward.

We have argued previously that, in addition to adequate Common Ground, joint activity requires a "Basic Compact" that constitutes a level of commitment for all parties to support the process of coordination to achieve group goals [25]. We may say that to coordinate effectively, parties must have the basic resources, including sufficient common knowledge, i.e., have the *ability* to coordinate (Common Ground), and also the willingness to coordinate, the *will* (the Basic Compact). The Basic Compact is an agreement (often tacit) to participate in the joint activity and to carry out the required coordination responsibilities to facilitate group success [25; 26]. We represent as the *ideal* of the Basic Compact to be that parties want to be involved in the joint activity, and they want it to be successful (e.g., relay racers on a track team). Of course, many influences can degrade these ideal conditions (e.g., someone being coerced to participate). Such degradations can affect members' loyalty to the Basic Compact. So also can a member's operational "stance," in terms of having adequate resources, being overloaded or fatigued, being distracted, and so forth.

One aspect of the Basic Compact is the commitment to some degree of goal alignment—typically this entails one or more participants relaxing some shorter-term local goals in order to permit group oriented long-term goals to be addressed. These longer-term goals might be shared goals (e.g., a relay team) or individual goals (e.g., drivers wanting to ensure safe journeys). A second aspect of the Basic Compact is a commitment to try to detect and correct any loss of Common Ground that might be disruptive.

We do not view the Basic Compact as a once-and-for-all prerequisite to be satisfied, but rather as a continuously reinforced or renewed agreement. Part of achieving coordination is investing in those things that promote the compact, as well as being sensitive to and counteracting those factors that could degrade loyalty to it.

All parties need to be reasonably confident that they and the others will carry out their responsibilities in the Basic Compact. In addition to repairing Common Ground, these responsibilities include such elements as acknowledging the receipt of signals, transmitting some construal of the meaning of the signal back to the sender, and indicating preparation for consequent acts. The Basic Compact is also a commitment to ensure a reasonable level of interpredictability; that is, agents acting, to the extent they can, so as to be *mutually predictable/understandable and mutually directable* by others.

What is the primary role of a "Basic Compact" in joint activity? We submit that a critical role has to do with *trust and predictability* in the operation of the whole interactive system. For example, when the Basic Compact is strongly in force, we can trust that other people are working on their assignments, are telling the truth about important matters, are going to send an item to another party if they say they will, and so forth. In another example, the Basic Compact requires that if one party intends to drop out of the joint activity, he or she must inform the other parties.   Hence, the Basic Compact "washes over" the entire enterprise of the joint activity, largely conferring a trust level in the operations of all the components. When it is functioning at its best, the Basic Compact contributes to the predictability of events within the joint activity, what we have argued is a primary role for regulation in the first place.

A certain way of interacting serves to maintain and even improve Common Ground. This way involves what is called the "joint action ladder" (JAL) [13]. When one party sends a message/signal to another, the second party, in reply, should 1) acknowledge that he has seen the signal arrive [attention], 2) "read" the signal [perception], 3) provide his interpretation of what it means [understanding] and, 4) indicate what he is likely to do as a result [action]. The latter two, in making understandings and intentions public, provide the opportunity for repair of Common Ground, the common understanding among the parties (e.g., through discussion of differences).

2.2 Coordination Devices

People coordinate through signals and more complex messages of many sorts (e.g., face-to-face language, expressions, posture). Human signals are also mediated in many ways—for example, through third parties or through machines such as telephones or computers. Hence, direct and indirect party-to-party communication is one form of a "coordination device," in this instance coordination by *agreement*. For example, a group of scientists working together on a grant proposal, may simply agree, through e-mail exchanges, to set up a subsequent conference call at a specific date and time. There are three other major types of coordination devices that people commonly employ: convention, precedent, and situational salience [13; 25].

Convention: Often prescriptions of various types and degrees of authority apply to how parties should interact. These can range from rules and mandated procedures, to less formal codes of appropriate conduct. These less formal codes include norms of practice in a particular professional community, as well as established practices in a workplace. Convention also often applies to activity devolving from more situationally emergent interactions in which we may engage, e.g., contracts we enter into, and debts and other kinds of obligations we take on. Coordination by convention depends on structures outside of a particular episode of joint activity.

Precedent: Coordination by precedent is like coordination by convention, except that it applies to norms and expectations developed within the ongoing experience of the joint activity. As a process unfolds, decisions are made about the mutually accepted naming and interpretation of things, standards of acceptable behavior and quality, who on the team tends to take the lead, who will enact particular roles, and so forth. As these arise and develop during the course of the activity, they tend to be adopted as devices (or norms) of coordination for the remainder of the activity.

(Situational) Salience: Salience has to do with how the ongoing work arranges the workspace so that next move becomes apparent within the many moves that could conceivably be chosen. Coordination by salience is produced by the very conduct of the joint activity itself. It requires little overt communication and is likely to be the predominant mode of coordination among long-standing, highly practiced teams.

Coordination devices often derive from regulatory systems, as will be discussed in Section 3.3, after further discussion of regulation itself, next.

3 Characterizing Regulation

Culturally-based regulatory systems are many and diverse, and go beyond what we normally construe as "law" or even "rules." In addition to law-like devices, they also include customs, traditions, work-place practices, standards, and even codes for acceptable everyday behavior. In this section we discuss some of what we have discovered in our attempts to characterize regulation.

3.1 Toward an Ontology of Regulation

Paul Wohlmuth, a philosopher of law, once wrote an introductory chapter for a special issue of the *Journal of Contemporary Legal Issues.* This issue focused on the "constitution of authority" [34]. By constitution of authority he meant, roughly, how different kinds of things come to have regulatory power over human activity. He used the example of an automobile traveling a bend in the road to illustrate the ubiquity and diversity of the authoritative forms that can come to bear on human affairs (see also 19; 29). Wohlmuth's full analysis is discussed in [16].

Starting from the numerous examples provided by Wohlmuth, we engaged in an exploration to identify some of the different kinds of devices that serve to govern human conduct. This has been a somewhat informal investigation. It started as a brainstorming effort when we began to build a list of all the different kinds of regulatory concepts we could think of. Then we used dictionaries, thesauruses, synonym finders and the like. The main criterion for inclusion in our list was that a device *constrains or promotes behavior in some direction.*

We needed to choose a word to represent the scope of the entire enterprise, and after considering several plausible candidates we settled on the generic term of "regulation" as one that seemed the most inclusive and accommodating of our purposes. As we continued our effort, rough categories for the numerous basic terms were created and refined. Once we had accumulated a reasonable seed set of regulatory mechanisms, we started circulating them among colleagues for reactions and contributions. As word has spread about the endeavor, friends and colleagues have kindly called in or e-mailed very welcome unsolicited contributions.

We should add that—at least for the present—our central interest has been in the gradual emergence of stable categories—e.g., the difference between a "law-like" device" and an "obligation-like" device—and we have not worried much about the precision of each definition. In fact, up to now nearly all of our rough definitions have come from ordinary dictionaries rather than from careful handcrafting. For this reason, similarities

and differences across broad categories seem more stable than distinctions among items within them. As an example, let us examine more closely laws and obligations.

Law-like devices are variants on coded, largely, but not always, written down rules. They characteristically have the power of the State or similar authority behind them. One can be jailed, fined, or otherwise sanctioned for violating them, all "legally." Some examples (including variants within the major category):

Law-like devices (have the power of the State/authority behind them)
- *Law*: All the rules of conduct established and enforced by the authority
- *Statute*: A law passed by a legislative body and set forth in a formal document
- *Bill*: A draft of a law proposed to a lawmaking body
- *Ordinance*: A custom or practice established by usage or authority
- *Mandate*: An authoritative order or command, especially a written one
- *Edict*: An official public proclamation or order issued by an authority; Decree
- *Decree*: An official order, edict, or decision, as of a church or government

While "law" generally develops over time, gets codified, and exists and is enforced over time in a relatively extended process, there is also "fast law," as in a "decree" or "edict." Hence we decided to treat the two kinds as separate categories; that is, within "law," there is both "fast" and "slow" law.

Let us now contrast "law" with "obligation." Obligation generally devolves from position, status, or special group, as in one becoming a parent, priest, Muslim, police officer, or the president of a social club. Some examples (including variants within the major category):

Obligation-like devices (accrue from one's position, role, state)
- *Obligation*: Duty imposed legally or socially. Activity that one is bound to as a result of a contract, promise, moral responsibility, position
- *Duty*: Any action necessary in or appropriate to one's occupation, role, or position. Includes duties to groups of people, e.g., one's elders, one's children
- *Responsibility*: Condition, quality, or fact of being responsible; obligation
- *Requirement*: Something obligatory or demanded as a condition

As in the law example, this category overlaps with some other items such as contracts and agreements—e.g., an agreement, or even a promise, can invoke an obligation—but contracts and agreements seem to have enough unique features that we treat them as a separate category, "Agreement-like things" (see Appendix).

One might wonder about the forms in which regulation exists. The current version of the Ontology contains over 220 regulatory concepts in about forty categories. Given this scope and diversity, one might argue that regulation is just about everywhere, in a myriad of forms. This condition of pervasive regulation has been noted by sociologists and social anthropologists for some time. In addition to Geertz (noted earlier), we cite Erving Goffman, who has claimed that:

> …one of the consequences of this learning program [socialization, learning the extant systems of regulation] is the transformation of the world into a place that is appreciably governed by, and understandable in terms of, social frameworks. Indeed, adults… may move about through months of their days without once finding themselves out of control of their bodies or unprepared for the impingement of the environment—*the whole of the natural world having been subjugated by public and private means of control* [20, p. 33, emphasis added, our annotations in brackets].

And here is the major link between regulation and coordination: When regulatory systems break down, predictability and order degrade, and coordination becomes impossible. As stated by Goffman:

> If the meaningfulness of everyday activity is similarly dependent on a closed, finite set of [interdependent] rules [and practices]… then one can see that the …the significance of certain deviant acts is that they undermine the intelligibility of everything else we had thought was going on around us, *including all next acts* [predictability], thus generating diffuse *disorder* [20, p. 5, emphasis added, our annotations in brackets].

Given the importance of regulation in successful joint activity, we have, as noted, been developing an ontology of regulatory systems. The current version of the Ontology of Regulation is organized under four main categories:

o *Regulatory Devices.* These are the many forms of regulation themselves that have the power to promote or constrain activity and were exemplified by "law" and "obligation" in our earlier discussions. This is the most highly developed category of the entries within the current Ontology -- with around 200 entries. (Some of the major categories and some examples are provided in the Appendix.)

o *Developmental processes for the Constitution of Authority.* As noted earlier, the constitution of authority refers to the ways in which things come to have regulatory power over human activity. In this category we include three types: The first we refer to as the "Origins and Derivations of Regulatory Devices" themselves. These are ways that regulatory systems come into existence. They include such processes as: force, the process of coming to an agreement, social emergence, legislation, court order, and divine intervention. The second we refer to as "Origins and Derivations of Officiation." These are processes and devices by which people or institutions are initiated into positions of authority to enforce regulation (e.g., a police officer), for example, by election, ritual, testing, or accreditation. The third are "Origins and derivations of Interpretive

Prerogative." These are processes and devices by which entities gain the authority to interpret regulation (e.g., judges). Examples are appointment, credentialing, or election. The first two categories listed in this subsection, together, are the second most highly developed categories in the current Ontology, with more than thirty entries.

o *Objects of Regulation.* These are the entities to which regulatory devices apply. They apply to the activities of people--including people in roles, such as a medical doctor, a priest, a certified public accountant, or a citizen of a county--and institutions (e.g., a publicly traded company), as they participate in such processes as a marriage ceremony, audit, or corporate merger.

o *Guardians of Regulation.* These are individuals or institutions empowered to enforce regulatory systems (as they apply to certain groups of people and their activities). Examples are a police officer with regard to the public, a parent with regard to his or her children, or the Securities and Exchange Commission with regard to publicly traded companies. These enforcing entities often gain their regulatory authority by virtue of some process involving the "Origins and Derivations of Officiation" as described above (e.g., by licensure, appointment, or election by a people).

o *Interpreters of Regulation.* Because the relationship between regulations and concrete applications in the world is often not straightforward (e.g., see Section 3.5 on the "Bureaucrat's Dilemma"), there must be people and institutions with authority to adjudicate alternative interpretations (e.g., judges or appeals courts). Such authority is usually granted through processes involving the "Origins and Derivations of Interpretive Prerogative." Examples are appointment of a judge and licensure of an attorney.

      The last three categories, "Objects of Regulation," "Guardians of Regulation," and "Interpreters of Regulation, except as illustrated, are largely undeveloped. It should also be noted that, depending on context, the same entity can a can appear in any of these three categories. For example, a parent may have a variety of regulatory authority over his or her children but, at the same time, be subject to the laws of the land.

      A sample of major categories of regulatory devices from the Ontology is given in the Appendix. We should note that many, if not most, of these regulatory devices—e.g., agreements or even contracts— are, themselves, best characterized as processes and joint activities (see [18]).

3.2. Fast and Slow Regulation

We take the "speed" of a regulatory device to have three main dimensions: how quickly it can be enacted and acquire authority, how quickly it can be changed, and how quickly it can be enforced. For good reason, human activity happens at many different speeds, and regulatory devices need to be appropriate to these different paces. We would not want to have to engage the mechanisms of changing constitutional law for guidance and reprieve

when someone has a gun to our face, for instance. The distinctions we address below have similarities to distinctions David Woods has made between activity at the "sharp end" (i.e., at the point of connection and impact with the world) and the "blunt end" (i.e., behind-the scenes culture and practice that bear on sharp-end activity) [15]. As a first attempt at addressing the speed of regulation, we have divided the major classes of regulatory devices into three types:

Systemic Schemas, including things such as folkways (e.g., customs, traditions, mores) and elements of "natural law." These are deeply engrained, pervasive guidance systems of a people, perhaps largely implicit, slow to develop, and slow to change. Interestingly, these types of regulation may be quick to enforce. It seems that in these kinds of matters people are inclined and feel authorized to take matters into their own hands when things go awry; that is, they may deal with enforcement personally and on-the-spot. For example, many citizens might be inclined to intervene upon seeing deliberate burning of their nation's flag or the desecration of their prophet's image or holy book.

Organizational Constraints and Allowances, including such things as authorizations, policies, practices, obligations, and codified rules (e.g., laws). These are more special-purpose devices that are explicitly and deliberately enacted and enforced by different configurations of people, within different socially constructed bodies (e.g., a club, agency, nation). While they can generally be enacted and modified more quickly than the Systemic Schemas, they can, in some instances, be slow to change through what may involve complicated processes (e.g., changing a country's constitution).

People tend to put enforcement in this category into the hands of some socially-sanctioned authority. Speed of enforcement varies by the degree of closeness/access of such an authority to the pertinent (regulated) activity. On the highway, for instance, for the *law* to be engaged as regulation requires spotting of the incident by a law enforcement officer and may subsequently involve courts and procedures and judgments by designated individuals. (As an aside, technology is now affecting speed of enforcement in many ways. For instance, cameras and sensors are now enabling on-the-spot detection of traffic violations. This technology can document the violation, gain the driver's identity and address from the license plate, and stuff and mail an envelope for delivery of the ticket and fine.) In contrast, good and bad manners (which are more like systemic schemas) on the highway (e.g., following too closely or cutting in too quickly after a pass) can and often are enforced by the individual participants—by honking horns or giving angry hand gestures—who feel they have authority in this context.

Action Guides. One kind of action guide is what we call "design of affordances." This involves what we make *hard or easy to do* by design, such as in the placing of doorways, sidewalks, streets, and bridges in particular places to channel traffic [28]. Not all design of affordances is physical. For instance, we can make some activities easy or hard to do through the allocation of resources (e.g., restricting gasoline). Another kind of action guide, what we call "transactional utilitarian devices," (see Appendix) are fast acting,

10

often fluid, regulatory devices that we set for ourselves in the process of conducting everyday affairs with others—e.g., making promises, agreements, appointments, gestures, and so forth. Creating and dismantling affordances can be slow or fast (putting up a roadblock vs. building a road). The key is that once they are in place, they have nearly immediate regulatory efficacy. On the other hand, almost by definition, transactional utilitarian devices can be quickly created, dismantled, or enforced at the point of activity.

3.3 Relationships between Regulatory Systems and the Coordination Devices

We have argued that regulatory systems are created by any social group to increase inter-predictability and order necessary for coordination in joint activity. Hence, it is not surprising that the main types of coordination devices (see previous Section 2.2) people use in the actual conduct of joint activity bear a strong relationship to various categories involved in the Ontology of Regulation (Appendix). Some of these relationships are clear, e.g., in the cases of agreement-like coordination devices, precedent-like devices, convention-like devices, and devices that are like "transient utilitarian devices/salience." Hence, the coordination devices can be thought of as the operational mechanisms of the more abstracted regulatory systems. It is an interesting challenge for further research to investigate more deeply these kinds of correspondences (e.g., identifying additional types of coordination devices by examining categories of the Ontology that do not match well to the four discussed previously, for example a new category of "schedules").

3.4. The Special Status of Norms Among Regulatory Devices

It is common in socio-cultural research to characterize social behavior as being subject to "norms" [12]. In this sense, norm means what Bourdieu called "habitus"—dispositions or schemas relating perception, thought, and action that socially structure experience and behavior, while dialectically structuring the social world [2]. We suggest that this characterization can be further specified. In our present treatment, we take behavior to be subject to regulation of *many and diverse* forms, as exemplified by the entries in our Ontology. Norms are but one these, indicating something about *personal and societal acceptability*. In this regard, norms still do have a special status among regulation types. In particular, there are at least three variations of norms, all of which possess some overlap, but at the same time suggest significant difference. All are socially constructed and enforced:

That which exists now: This norm, call it "Norm1," pertains to *what is* actually in place now (regardless of what the pertinent official rules, laws, etc., may be—see below). Regarding driving a highway, Norm1 refers to the largely self-organized, in-place, ambient traffic speed: maybe 10+ the posted speed limit in clear, dry weather—lower in ice or snow. What is extant is special for a number of reasons. First it has a certain

11

momentum and inertia by way of its development, implementation, and execution, including enforcement and interpretation [34]. "Possession is nine-tenths of the law," so to speak. Think of the intense societal processes associated with the regulation of abortion in the United States. Many people have devoted a good share of their lives (and spent a lot of money) to persuade the public, legislators, and judges to accept their point of view. Moreover, "what is" attracts a following, in particular all those who benefit from the current state of regulation, either in their personal behavior, by favor of their constituent voters (or not), in medical facilities that can or cannot operate, or in related official posts that were created, and people were subsequently able to assume, because of the state of affairs.

That which is socially tolerable (relative, of course, to some reference group): This Norm2 (the "stretched norm") refers to what people of a certain community will *actually still tolerate*, beyond in-place norms. For instance, Norm2 reflects what a driver can get away with without incurring the wrath of other drivers—e.g., their taking down a tag number or calling in the police, honking their horns, or trying to run him of her off the road. While drivers, following the in-place norm, routinely drive faster than the posted speed limit, if somebody drives *too* fast, or even at the *normal* rate in icy and rainy conditions, others may try to take some action. It is clear that there are ranges of public tolerance for deviation from social norms established as "what is." Audience members at the symphony may dislike but tolerate some degree of coughing or whispering, but may chastise louder talking or repeated cell-phone use.

A more formal benchmark or standard by which some thing or process is judged acceptable: These are the more **formal** standards for regulation of behavior or products in a community. They often lie behind norms of the first two kinds, employed only when there are breakdowns in more informal employment and enforcement. In driving the highway, Norm3 refers to such things as the posted speed limit and all the legal machinations behind and entwined within it (e.g., the motor vehicle code, the courts). Norms of type three include laws, ordinances, posted job rules, rules-of-the-game, and so forth. As an example, let us look at the three kinds of norms as they might apply to a junior lawyer in the (hypothetical) law firm of Dewey, Cheatem, & Howe:

o   Norm1 refers to what the young lawyer does, pretty much like the other junior lawyers do, on a day-to-day basis, when everything is going uneventfully, within routine.
o   Norm2 refers to what deviations people around him will tolerate without reprisal (e.g., taking off at noon each Friday to get a jump on the weekend, being persistently behind in logging his billable hours, doing a great deal of travel away from the office).
o   Norm3 refers to the formalized rules and regulations pertaining to "being a junior partner at the Dewey firm."

So, what is special about norms as a regulatory device, compared to all the others? We propose that as benchmarks and criteria for acceptability, essentially for necessary quality, norms apply in the execution *of all the other* regulatory devices. These other regulatory devices can be carried out, applied in ways that are themselves subject to extant norms of all three kinds: the ways that are currently honored in a social group, the deviations that are tolerated, and in the formal regulations that can be brought to bear under challenge. One can easily see how these might play out in the execution of an agreement, even more clearly with regard to a contract, but even with regard to such activities as engaging in a promise, giving and responding to a command, executing one's duties and obligations, or just "obeying the law."

Norms also apply to the internal workings of any joint activity itself, in the cues that signal the need for and initiation of the joint activity, in the coordination devices and reciprocal actions utilized by the parties within the execution of the joint activity, and in the signals that indicate the joint activity is coming to a close. In a simple example, in the activity of a customer checking out at the counter of a convenience store, the customer generally initiates the interchange by walking up to the sales clerk's counter and placing an item on it. It is not the norm, for this purpose, for the customer to continue clutching the item—which may in fact be a cue to the clerk that the customer is approaching to ask a question ("Where's the restroom?"). One can envision how norms might be involved in the other components of this seemingly simple, everyday transaction, including the coordination devices that are employed.

3.5. Qualifying Regulation: The Bureaucrat's Dilemma

Regulations may be implicitly or explicitly qualified in order to take account of context. I promise to drive my friend to the airport on Friday—*if* my car is out of the repair shop by then. It is okay to break the speed limit—*if* one is driving an accident victim to the hospital. When "authorities" do not tie the hands of "enforcers" too rigidly, the latter can recognize and appropriately adapt to such circumstances. However, because of their limited perceptual and reasoning abilities and their difficulty in grounding and adapting to context, this is a daunting challenge for software agents.

Consider this seemingly simple rule: *Congress shall make no law abridging freedom of speech.* Then we start thinking about the circumstances in which one would not want this rule to be honored—for example, the famous crime of (falsely) crying "Fire!" in a crowded theater. So, we further qualify: *Congress shall make no law abridging freedom of benign speech.* But what is benign? (or "speech" for that matter?).

For every qualification made, one is left with a choice. Either a judgment needs to be made (e.g., about what is "benign") or there needs to be yet another qualification (e.g., defining what constitutes "benign"). In human affairs, this can lead to absurdity if there is an attempt to expunge all human judgment from such interpretations—to specify so extremely that criteria are mechanically analyzable. An example has been presented in the literature about national workplace rules that require that all handrails be 28 inches

high, whether the users are professional basketball players or midgets [24]. Revisiting speech, and our example, such specifications might take the form: *Congress shall make no law abridging freedom of speech that lasts less than one-half hour,* a degree of specification that enables mechanical and reliable interpretation by artificial agents, even by relatively stupid ones, but which can lead to absurd rigidity and reversals of intended effects.

4. How Can We Make Software Agents Better Team Players?

A key challenge in any kind of joint activity is to successfully navigate points of interdependency among the participants. In this section, we will focus on two challenges for software agents to become better team players: 1) participation in the Basic Compact, and 2) the crucial "handoff" operation embodied in the "joint action ladder" (JAL) as it contributes to mutual understanding.

Being part of the Basic Compact means that participants want to participate in the joint activity and want the group's goals to succeed. This is an ideal, and variants often occur (e.g., forced participation). This affects trust and confidence throughout the enterprise, for example, the belief that participants will function and report appropriately.

Of course, except for rare software agents that can reason about the relative utility of their participation in competing tasks, the issue of an agent's dedication to the group work is moot. More germane to the subject for most agents is their responsibility to uphold certain kinds of standards in reporting and functioning. For instance, if an agent reaches a failure point and will not be able to complete its task, this needs to be reported to others, and agents can often do this. On the other hand, more nuanced kinds of "appraisals of progress" can pose daunting challenges for software agents. In contrast, highly expert human teams have ways of understanding and reporting about "how things are going" that can anticipatorily *foreshadow* success or failure.

With respect to the first rung of the JAL, *attending*, agents need to share a common repertoire of watched-for signals that call for entry (and exit) into interdependency cycles—for example, signals such as a customer placing a store item on the check-out counter at a convenience store. The signals (and more complex messages) that the agents send and receive ideally would be expressed at many levels of overtness and abstraction, depending on the situation. These could range from simple beeps and buzzes (e.g., the warning of a big truck backing up) to complex appraisals that would tax the best human experts (e.g., "The battle here is going badly").

In critical situations, agents cannot afford to miss important signals. This is a matter of attention management, both in one's self and in others. Sometimes attention is directed appropriately by the nature of the ongoing activity itself, through the coordination device of salience. However, in many cases where practice is less of a routine, the need for attention focus calls for the ability of agents to direct each other to help ensure that messages are not missed. Operative regulatory frameworks can help, as in certain circumstances cues to the start of joint activity may be mandated, conventional, part of accepted practice, and so forth (e.g., the reaching out of a hand for hand shaking).

Once arrival of signals/messages has been detected, the JAL prescribes that they must be *read* and *understood.* Analysis must reveal their "meaning in context." This is largely a matter of adequate Common Ground, the pertinent (to the activity) knowledge that agents either bring with them or gain in the course of ongoing participation, including knowledge of the regulatory mechanisms and coordination devices that apply. Knowledge of pertinent regulatory devices can help by constraining the space of possible meanings for both sender and receiver (e.g., conversation policies [21; 23]). A major challenge, however, is for the parties to construct mutually a model of *which* regulatory systems *apply*. This is particularly challenging because the question is highly context sensitive; that is, it depends on how the parties conceptualize "what they are doing now." Consider the pertinent forms of regulation that would be taken to apply if a couple (or an observer watching them) conceived that they were "inspecting a house" for possible purchase, versus "casing a house" for a robbery, or a group thinks it is "conducting a holy war" versus "committing criminal murder" [1; 9]. Fluidly adopting different perspectives is difficult even for humans [17]. A major challenge for software agent participation in joint activity is the sophistication with which agents can adopt such "points of view" regarding their activities or changing roles.

When an agent has created a candidate understanding, it must broadcast this understanding for public scrutiny, in particular, for appraisal by those others with whom the agent's actions are most interdependent. This provides opportunity for the detection of failures and slips of Common Ground among the parties. Additionally, the final *action* rung of the JAL, involving broadcasting the next steps one is likely to take on the basis of this current, tentative understanding, not only provides additional information about understanding, but it also decouples intention to act from action itself, allowing for corrective intervention before actual moves are made. Both of these mechanisms contribute to increasing the predictability and transparency (e.g., of intent, stance, and state) of interdependent agents to each other and, hence, contribute to productive coordination.

Finally, the parties need to participate in repair of their understandings, as needed, to enable successful further progress in the joint activity. Making mutual understandings visible aids this repair process, and, hence, improves the quality of Common Ground. Aspects of discordance among the parties' understandings can vary in transparency and abstraction, from those involving simple facts and observable states, to those involving more complex levels of interpretation and reasoning that may tax agents, e.g., relations among things, inferences, interpolations, and appraisals of progress [7]. In any event, discrepancies in the agents' understandings require that they be capable of some forms of *negotiation*. As with determination of meanings in general, knowledge of the regulatory systems under which the agents are operating may help to delimit the range of discrepancies that occur.

The sketch just presented of the nature of understanding and interaction needed for successful joint activity is a caricature, intended to overly pull apart and concretize operations that in practice are often more blended and abbreviated, especially as groups work together over time and learn about their activities and each other. That is, in reality,

things are often different. One example is "downward evidence" in the JAL [13], by which higher levels of response are taken as indications of lower ones (e.g., a response that conveys sensible understanding of a message also conveys that it was noticed and read). Other examples of real complexities that belie the simplifications presented here involve qualitative changes of state in items such as coordination devices and even regulations. For example, agreements over time can become precedents, at which point they no longer require operational deliberation. Procedures conducted with explicit reference to plans, over time can come to be directed mostly by situational salience. Practice over time can be transformed into law.

The treatment above, though compromised, is intended to accentuate the kinds of understandings and operations that are required to conduct successful joint activity in its rawest form, and hence to point to the most important places where software agent capability must be considered (and possibly improved over time) in creating successful, mixed human-agent working groups.

5 Applications to Human-Agent-Robotic Teamwork

We are applying the ideas presented in this chapter to facilitate joint activity in mixed human-agent-robot teams. We are doing this in two primary ways. First we are implementing regulatory systems that help coordinate joint activity through constraints of authorizations and obligations that we call "policies" [4]. In this we utilize KAoS Policy and Domain Services, a framework that can reliably and flexibly specify, analyze, enforce, and adapt policy. Applications include modeling a point-of-view, such as in "adversarial modeling," in which we attempt to model actions and reactions to events by groups of people sharing a culture greatly different from our own. Another is cross-domain information exchange (CDIX), in which we attempt to facilitate the sharing of information across different governmental and relief agencies, perhaps highly different in their rules, procedures, and organizational "cultures." We are also applying this work in the context of human-agent-robot teams within complex operations, such as space exploration, disaster response, and military operations [3; 5].

A second direction is to enhance our understanding of the regulatory systems operative in human social behavior and to develop our "Ontology of Regulation" that can be implemented to support our human-agent modeling systems. The ontology is a work in progress. We are currently formalizing its concepts and relations within an OWL ontology using the Cmap Ontology Editor (COE; [22]). (For a current version, contact Feltovich).

We believe that the unusual approach of modeling social behavior with more attention to operative regulatory structures, relative to individual behavior and cognition, will complement other modeling approaches in important ways. For example, much of a people's culture may well be more stable than individual behavior and cognition, and may well be more amenable to being represented in advance, when, for example, there is need to quickly ramp-up a model of a brand-new hostile social group.

16

Acknowledgements

References

[1] Anderson, R. C., & Pichert, J. W. (1978). Recall of previously unrecallable information following a shift of perspective. *Journal of Verbal Learning and Verbal Behavior*, 17, 1-12.

[2] Bourdieu, P. (1977). *Outline of a theory of prac*tice (Trans. by R. Nice). NY: Cambridge University Press.

[3] Bradshaw, J. M., Acquisti, A., Allen, J., Breedy, M. R., Bunch, L., Chambers, N., Feltovich, P., Galescu, L., Goodrich, M. A., Jeffers, R., Johnson, M., Jung, H., Lott, J., Olsen Jr., D. R., Sierhuis, M., Suri, N., Taysom, W., Tonti, G., & Uszok, A. (2004). Teamwork-centered autonomy for extended human-agent interaction in space applications. *AAAI 2004 Spring Symposium*. Stanford University, CA, AAAI Press,

[4] Bradshaw, J. M., Feltovich, P. J., Jung, H., Kulkarni, S., Allen, J., Bunch, L., Chambers, N., Galescu, L., Jeffers, R., Johnson, M., Sierhuis, M., Taysom, W., Uszok, A., & Van Hoof, R. (2004). Policy-based coordination in joint human-agent activity. *Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics.* The Hague, Netherlands,

[5] Bradshaw, J. M., Jung, H., Kulkarni, S., Johnson, M., Feltovich, P., Allen, J., Bunch, L., Chambers, N., Galescu, L., Jeffers, R., Suri, N., Taysom, W., & Uszok, A. (2005). Toward trustworthy adjustable autonomy in KAoS. In R. Falcone (Ed.), *Trusting Agents for Trustworthy Electronic Societies.* (pp. in press). Berlin: Springer.

[6] Cheney, D. L., & Seyfarth, R. M. (1992). Precis of how monkeys see the world. *Behavioral and Brain Sciences*, *15*(1), 135-182.

[7] Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science*, 5(2), 121-152.

[8] Clancey, W. J. (1997). *Situated Cognition: On Human Knowledge and Computer Representations.* Cambridge: Cambridge University Press.

[9]Clancey, W. J. (1999). *Conceptual coordination: How the mind orders experience in time.* Hillsdale, NJ: Lawrence Erlbaum,

17

[10] Clancey, W.J (2004). Automating Capcom: Pragmatic Operations and Technology Research for Human Exploration of Mars. In C. Cockell (ed.) *Martian Expedition Planning, Vol. 107, AAS Science and Technology Series,* (p.p. 411-430),

[11] Clancey, W. J. (2004). Roles for agent assistants in field science: Personal projects and collaboration. *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, May *34*(2), 125-137.

[12] Clancey, W. J., Sierhuis, M., Damer, B., & Brodsky, B. (2005). Cognitive modeling of social behavior. In R. Sun (Ed.), *Cognition and Multi-Agent Interaction: From Cognitive Modeling to Social Simulation*. New York City, NY: Cambridge University Press.

[13] Clark, H. H. (1996). *Using Language*. Cambridge, UK: Cambridge University Press.

[14] Clark, H. H., & Brennan, S. E. (1991). Grounding in communication. In L. B. Resnick, J. M. Levine, & S. D. Teasley (Ed.), *Perspectives on Socially Shared Cognition*. Washington, D.C.: American Psychological Association.

[15] Cook, R. I., & Woods, D. D. (1994). Operating at the sharp end: The complexity of human error. In S. U. Bogner (Ed.), *Human Error in Medicine*. Hillsdale, NJ: Lawrence Erlbaum.

[16] Feltovich, P., Bradshaw, J. M., Jeffers, R., Suri, N., & Uszok, A. (2004). Social order and adaptability in animal and human cultures as an analogue for agent communities: Toward a policy-based approach. In *Engineering Societies in the Agents World IV. LNAI 3071*. (pp. 21-48). Berlin, Germany: Springer-Verlag.

[17] Feltovich, P.J., Hoffman, R.R., Woods, D., & Roesler, A. (May-June, 2004). Keeping it too simple: How the reductive tendency affects cognitive engineering. *IEEE Intelligent Systems,* 90-94.

[18] Feltovich, P. J., Spiro, R. J., Coulson, R. L., & Myers-Kelson, A. (1995). The reductive bias and the crisis of text (in the law). *Journal of Contemporary Legal Issues*, 6(1), 187-212.

[19] Geertz, C. (1973). *The Interpretation of Cultures*. New York, NY: Basic Books.

[20] Goffman, E. (1974). *Frame Analaysis*. New York City, NY: Harper and Row.

[21] Greaves, M., Holmback, H., & Bradshaw, J. M. (1999). What is a conversation policy? M. Greaves & J. M. Bradshaw (Ed.), *Proceedings of the Autonomous Agents '99 Workshop on Specifying and Implementing Conversation Policies,* (pp. 1-9). Seattle, WA,

[22] Hayes, P. J. (2005). Speaking informally... P. Clark & G. Schreiber (Ed.), *Proceedings of the Third International Conference on Knowledge Capture (K-CAP 2005)*. Banff, Alberta, Canada, ACM,

[23] Holmback, H., Greaves, M., & Bradshaw, J. M. (1999). A pragmatic principle for agent communication. J. M. Bradshaw, O. Etzioni, & J. Mueller (Ed.), *Proceedings of Autonomous Agents '99,* (pp. 368-369). Seattle, WA, New York: ACM Press,

[24] Howard, P. (1994). *The Death of Common Sense: How Law is Suffocating America*. New York City, NY: Random House.

[25] Klein, G., Feltovich, P. J., Bradshaw, J. M., & Woods, D. D. (2004). Common ground and coordination in joint activity. In W. B. Rouse & K. R. Boff (Ed.), *Organizational Simulation*. (pp. 139-184). New York City, NY: John Wiley.

[26] Klein, G., Woods, D. D., Bradshaw, J. M., Hoffman, R., & Feltovich, P. (2004). Ten challenges for making automation a "team player" in joint human-agent activity. *IEEE Intelligent Systems*, 19(6), 91-95.

[27] Locke, J. (1689/1962). An essay concerning the true original, extent, and end of civil government. In E. Barker (Ed.), *Social Contract*. (pp. 1-143). New York City, NY: Oxford University Press.

[28] Norman, D. A. (1988). *The Psychology of Everyday Things*. New York: Basic Books.

[29] Norman, D. A. (1992). Cognitive artifacts. In J. M. Carroll (Ed.), *Designing Interaction: Psychology at the Human-Computer Interface.* (pp. 17-38). Cambridge: Cambridge University Press.

[30] Norman, D. A. (1992). Turn signals are the facial expressions of automobiles. In *Turn Signals Are the Facial Expressions of Automobiles.* (pp. 117-134). Reading, MA: Addison-Wesley.

[31] Rousseau, J. J. (1762/1962). The social contract. In E. Barker (Ed.), *Social Contract.* (pp. 167-307). New York City, NY: Oxford University Press.

[32] Smith, W. J. (1977). *The Behavior of Communicating.* Cambridge, MA: Harvard University Press.

[33] Smith, W. J. (1995). The biological bases of social attunement. *Journal of Contemporary Legal Issues*, 6, 361-371.

[34] Wohlmuth, P. C. (1995). Traveling the highway: Sources of momentum in behavioral regulation. *Journal of Contemporary Legal Issues*, 6, 1-9.

Appendix: Some major categories from the Ontology of Regulatory Devices:

- Coercion (various uses of force, pressure, intimidation)
- Agreements (including contracts, promises, etc.)
- Precedents (guidance from pertinent past cases, decisions, events)
- Plans (recipe-like things, including schedules, designs, forms)
- Standards/norms (benchmark kinds of things—what is contextually acceptable)
- Fashion: (the current style or mode of dress, speech, conduct, etc.)
- Exceptions (various kinds of suspensions of rules, e.g., zoning variances, waivers)
- Commands (orders by an authority)
- Permissions (allowances to conduct certain actions)
- Folkways (including practices, taboos, customs, ceremonies, myths, rituals...)
- Lessons (ways of operating, acting learned from study or, often, unfortunate, experience— "That surely taught me a lesson.")
- Codified rules (things similar to laws)
- Obligations (accrue from positions held or assumed)
- Authorizations (means of granting authority)
- Incentives (which implicates *disincentives,* i.e., classes of enforcement and punishment)
- Design of Affordances (by what we make *hard or easy to do* by design, even physical design, such as the placement of doorways, sidewalks, streets, bridges—cf. Norman [28])
- Transactional utilitarian devices (fast acting, often fluid, regulatory devices that we set for ourselves in the process of conducting everyday affairs—e.g., making promises, agreements, appointments, pledges, gesture, expression…
- "Natural Law:" (for example, rights to life, liberty, property, …rights to punish violations of natural law, etc. "Social Contract" theorists, e.g., Locke [27], Hume, Rousseau [31])
- Physical and physiological law
- Ideology/Belief/philosophy (idea systems of a people, e.g., Confucianism, democracy)
- Honor Codes (including "face," codes of ethics, and duties)