Norms as a Basis for Governing Sociotechnical Systems: Extended Abstract*

Munindar P. Singh
Department of Computer Science
North Carolina State University
Raleigh, NC 27695-8206, USA

singh@ncsu.edu

Abstract

We understand a sociotechnical system as a microsociety in which autonomous parties interact with and about technical objects. We define *governance* as the administration of such a system by its participants. We develop an approach for governance based on a computational representation of norms. Our approach has the benefit of capturing stakeholder needs precisely while yielding adaptive resource allocation in the face of changes both in stakeholder needs and the environment. In current work, we are extending this approach to tackle some challenges in cybersecurity.

1 Challenges in Sociotechnical Systems

We define a *sociotechnical system (STS)* as a microsociety in which autonomous stakeholders interact. How can these stakeholders collaborate effectively even though their interests may only be imperfectly aligned? Administering STSs is made complicated by complexity and change not only in the resources being shared but also the nature of the interactions among the stakeholders. We address the challenge of enabling stakeholders to administer or *(self-)govern* such systems in a manner that respects their autonomy. A particular benefit are the resulting gains in adaptability in accommodating any exceptions and opportunities.

The Ocean Observatories Initiative (OOI) [Arrott *et al.*, 2009], which facilitates scientists and research institutions in acquiring, storing, analyzing, and sharing information from the world's oceans, is a paradigmatic STS. Its stakeholders include oceanographers, educators, and members of the public as well as research laboratories and universities. The OOI has three key features. One, autonomy: the stakeholders own and share resources such as Underwater Autonomous Vehicles (UAVs), buoys, ocean sensors, and research databases. Thus, the OOI would support collaborations in which it would not own all resources involved. Two, lifetime: a requirement for the OOI was to sustain operation for decades. Thus we must accommodate changes in stakeholder needs without relying upon any specific technology to be available throughout the lifetime of the system. Three, scale: the OOI could end up with thousands of stakeholders, tens of thousands of physical resources such as ocean gliders, and potentially millions of cyber resources such as datasets. At those scales, automation and adaptation are essential for administering resources according to the preferences of the stakeholders.

How can we accommodate continually changing stakeholder needs? How can multiple stakeholders collaborate in a sustainable, efficient manner? How can individual ownership and control be respected as autonomous parties interoperate? How can resources be added or dropped dynamically at runtime? How can coalitions be constructed and enacted to dynamically share resources while entertaining challenges such the stakeholders' needs changing unexpectedly, as in an emergency? How may we accomplish all of the above *adaptations* over a wide range of resource granularities and timescales?

1.1 Governance: Norms and Organizations

We term dealing with the above challenges *governance*. Governance contrasts with traditional management, which presumes authority (superior to subordinate) relationships. In STSs, the collaborating parties are autonomous peers and none has authority over the others. Today, governance is carried out "by phone call"—by ad hoc negotiations among humans. However, the STSs of interest involve large numbers of resources and require decision making at fast timescales, so manual negotiations would simply not be feasible.

From the perspective of governance, the stakeholders of an STS are themselves *participants*. Recognizing their autonomy, we observe that we cannot prescribe a decision-making strategy for each participant. Instead, each STS can prescribe its rules of encounter via a set of norms. Informally, a *norm* characterizes sound or "normal" interactions among members of a social group, reflecting their mutual expectations. We emphasize *interactions*: behavior that has no effect on others is irrelevant for our purposes. Two examples of norms in a scientific setting are putting an instrument in power-save mode at the end of an experiment and closing unneeded datastreams from sensors. Norms may arise through top-down legislation or bottom-up conventions emerging from norms implicit in participant strategies [Savarimuthu *et al.*, 2009]. We restrict ourselves to norms that carry contractual force, so that their satisfaction or violation is significant.

 This paper is an extended abstract of [Singh, 2013], which appears in the *ACM Transactions on Intelligent Systems and Technology*.

Based on the above intuition, we formalize an STS as an *organization* that involves two or more roles, each specified in terms of the norms applying to it. To this end, we formalize norms not as amorphous properties of the "system"—whatever that might be—but as directed normative relationships between participants in the context of an organization. Our formal model reflects this essential duality of organizations and norms: an organization is defined via norms and a norm is defined in an organization. Importantly, our approach accommodates *open* settings where a party may live and act outside the scope of a sociotechnical system while remaining subject to the norms defined in the system.

1.2 Adaptation in Sociotechnical Systems

Our approach seeks to engineer a sociotechnical system in such a manner as to support adaptation, both (1) in its configuration and implementation and (2) in its enactments, as realized through the interactions of its participants. The twin challenges of ensuring adaptation and achieving rigor lead us to adopt the following main principles.

The first principle is the *Centrality of Norms:* A normative, as opposed to an operational, characterization of acceptable interactions is minimally constraining and thus essential for capturing the "invariants" of a long-lived system. The second principle is *Autonomy and Policies:* The participants are autonomous, though subject to applicable norms. Each participant applies its internal *policies* to decide how to interact given the norms; its policies reflect its autonomy.

The foregoing emphasis on autonomy and adaptation suggests that our computational system must incorporate *agents*: active computational entities that represent individual participants and organizations, and whose interactions (subject to norms) are relevant to governance. The agents are only partially regimented. Where appropriate, we prefer to develop agents that respect the applicable norms, but the autonomy of the participants means that any agent may violate a norm. Therefore, norms provide a rigorous basis for *coherence*, which we view as a relaxed notion of correctness that accommodates restoring a "good" state after a violation.

1.3 Contributions and Claims

We develop a novel approach for governance that is computationally realized and deals well with complexity and dynamism. Our first contribution is a *formal model* for governance that incorporates a rich set of normative clauses promoting adaptability and reuse. This model provides a natural mapping to computations and provides a standard of correctness over those computations. In addition, it supports analyzing particular organizations and norms, e.g., with regard to consistency. Our second contribution is an *architecture* that realizes the above model. We claim that our approach (1) enables the construction of a *flexible* STS that can both adapt in (2) its *configuration* to accommodate changes in stakeholder needs by reconstituting its rules of encounter and (3) its *operations* to accommodate changes in its environment.

2 Modeling a Sociotechnical System

An effective approach must be adaptive (to accommodate change) and rigorous (to provide assurance of appropriate outcomes despite complexity). A normative approach can address both needs. Our technical development proceeds as follows. We begin from a general organizational model for STSs. We refine this model to introduce a small set of norm types. (This abstract elides a corresponding vocabulary for expressing norms and an agent architecture based on policies expressed using the vocabulary and additional predicates.)

2.1 Conceptual Model of a Sociotechnical System

Figure 1 shows our conceptual model. An STS maps to an Org; its participants to members of the Org, i.e., principals who play roles in the Org. An Org is crucial in formulating interactions in terms of norms. All norms arise with an Org as a backdrop. An Org is recursively constructed: its members are principals that could themselves be Orgs. A principal may be a member of more than one Org: thus Orgs can have overlapping memberships. We assume that the membership relation between Orgs and principals is well-founded.

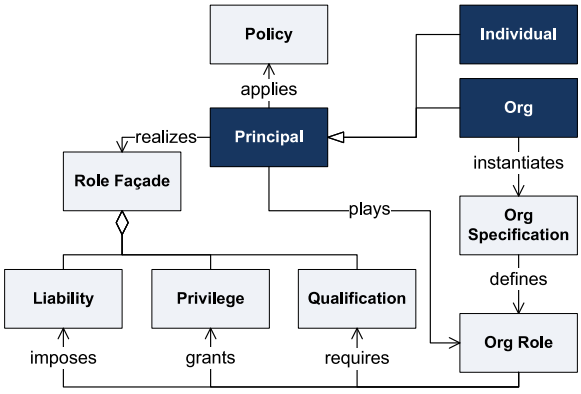

Figure 1: Simple conceptual model of an STS. (White text indicates active entities; black text indicates representations.) A *principal* corresponds to a participant in a system and may be an *individual* or an *Org*. A principal is a locus of autonomy and accordingly chooses its own *policies*.

Principals collaborate within the scope of an Org of which they are members. An Org systematizes the norms among its members and potentially provides an authority to which the members may complain regarding norm violations by others. An Org may apply appropriate sanctions on its members; such sanctions typically include canceling the membership of, or further escalating a complaint against, a principal it judges malfeasant.

Orgs are finely structured through the notion of a *role*, which codifies a set of related interactions that a member of an Org may enact. To be a member of an Org means to play at least one role in that Org. A principal may concurrently play more than one role in the same or different Orgs. Each Org is specified by defining rules of encounter for each of its roles: we can think of these rules as forming a multiparty contract. For each role, we collect the elements that concern it directly into the role's *façade*. Each façade comprises three major components.

Qualification A prerequisite or eligibility requirement for a principal to play the specified role. Example: Only a

credentialed and currently employed teacher may participate as an *educator* in a continuing education Org for school teachers.

Privilege A liberty, broadly understood, accorded to a principal who plays the specified role. Example: A teacher admitted as an *educator* to a summer camp Org is authorized to access all camp datasets and is empowered to admit a student as a *pupil* in the camp.

Liability A demand imposed on a principal who plays the specified role. Example: A teacher who becomes an *educator* must entertain help requests from a student who is a *pupil*. A *pupil* who introduces a virus into the camp's computers risks sanctions, including expulsion.

Importantly, privileges and liabilities map to the normative relationships that the principals enter into, some of which accord liberties and some of which impose demands on the principal who adopts the specified role. Adopting a role in an Org is thus a path for a principal to enter into norms with other principals. Although principals may negotiate additional norms through negotiation, such norms are governed by the Org. For example, a teacher as an *educator* gains access to datasets but not to instruments. To use an instrument owned by a scientist, a teacher may agree on additional terms and conditions, such as that he would not reboot the instrument. Such agreements would arise in the scope of the same Org, and their violation could have consequences such as the imposition of sanctions based on the *educator* façade.

The model of Figure 1 posits that an Org is a principal and can thus play a role in another Org. We further posit that an Org *qua* principal may interact with, and enter into norms with, its own members. For example, when researcher Ryzard joins OOI, not only is he subject to OOI's norms but he may also expect OOI to keep his private information safe. We capture the above intuition by postulating a distinct *self* role for each Org to be played by exactly one principal, the Org itself, and to be instantiated along with the Org. An Org as *self* interacts with all its members, handles their requests to discover other members and resources, entertains their complaints about each other, adjudicates on the norms between them, and enforces any applicable sanctions upon them.

2.2 Normative Concepts

Based on an analysis of STSs, we postulate the following normative concepts as the key elements of a role façade. We expect their familiarity to people would facilitate eliciting requirements and explaining how principal's interact.

A norm codifies desired properties of interactions among principals: it captures how an interaction *ought* to proceed and thus regulates the interactions of the principals involved. By providing a rich set of constructs with which to express the norms, we enable encoding the essential properties of interactions in a manner that is flexible (any enactment that satisfies the norms is acceptable) yet rigorous (there is a precise computational notion of when a norm is violated). The flexibility helps ensure correctness while supporting adaptation in configuration (to accommodate changes in stakeholder requirements) and during enactment by the principals. The norms progress because of the principals' interactions and may be activated, satisfied, or violated as an enactment proceeds. A snapshot of the norms taken together constitutes the *social state* of the STS, in Baldoni et al.'s [2010] terminology.

Each norm involves a *subject* (the principal on whom the norm is focused), an *object* (the principal with respect to whom the norm arises), a *context* (the Org within whose scope the norm arises), an *antecedent* (expressing the conditions under which the norm is fully activated and brought into force), and a *consequent* (expressing the conditions under which the norm is fully satisfied and deactivated). An antecedent of true indicates an unconditional norm. Further, the context could be distinct from or the same as the subject or the object, whereas the subject and object are always distinct.

The norm representation enhances modularity by supporting patterns over norms that have interrelated conditions or where one norm references whether another norm is satisfied, violated, and so on. Examples of these are reciprocal commitments and sanctions for violations of prohibitions. By placing norms in an organizational *context* and supporting their *manipulation*, we combine the benefits of (1) a precise declarative characterization of social state with (2) a clear statement of institutional actions. We consider five types of norms.

Commitment The subject (i.e., debtor) commits *to* the object (i.e., creditor) that if the antecedent holds, the debtor will bring about the consequent [Singh *et al.*, 2009]. When the consequent holds, the commitment is satisfied and deactivated. Example: A researcher who borrows an instrument commits to returning it within one hour of being requested to do so.

Authorization The object authorizes (i.e., permits) the subject to bring about the consequent when the antecedent holds. Bringing about the consequent if the antecedent is false is a violation. Example: An instrument owner authorizes a colleague to use it between 7:00PM and 9:00PM today.

Prohibition The object prohibits (i.e., forbids) the subject from bringing about the consequent provided the antecedent holds. Bringing about the consequent if the antecedent holds is a violation. Examples: An instrument owner prohibits a borrower from changing the firmware on the instrument. A dataset curator prohibits a reader from publishing any of the data on an external web site.

Sanction The object would sanction the subject by bringing about the consequent provided the antecedent holds. Examples: An instrument owner sanctions a borrower who illicitly changes the firmware on a borrowed instrument by giving the borrower a poor rating. A dataset curator sanctions a reader who publishes any of the data externally by complaining to the relevant resource sharing Org. The Org sanctions a reader who publishes any of the data externally by ejecting him from the Org.

Power The object empowers the subject to bring about the consequent by bringing about the antecedent. A power is the ability to alter the norms between two or more principals [Hohfeld, 1919], usually those playing specific roles. In addition, a power exemplifies the *counts-as* relation between a low-level (physical) ability and a high-

level (institutional) action [Jones and Sergot, 1996]. In our setting, each physical action is a communication: thus when the antecedent holds, the subject need only "say so" to bring about the consequent. A principal may be empowered to do something while being prohibited from exercising that power. Examples: The Chesapeake Bay Org is empowered to admit or eject its members by declaring so. A system administrator is empowered to admit new people into OOI by creating their accounts, but is—crucially—prohibited from creating accounts (and thus effectively admitting members) without approval from the membership department. However, because the administrator has the power, her creation of a new account succeeds, though it might later be deemed illicit and revoked and the administrator sanctioned for exercising the power illicitly. Here, the power is misused and the prohibition is violated.

A commitment, prohibition, or sanction is a liability for its subject since it can only lead to the subject investing effort or having its freedom curtailed or suffering a penalty. In the same spirit, an authorization or a power is a privilege for its subject since the subject obtains an option to perform additional actions without being required to do so. Liability and privilege are two faces of the same coin: a norm that is liability for its subject is a privilege for its object and vice versa. Qualifications do not feature as privileges or liabilities because they are formed of credentials of the principals, such as their participation in specified Orgs in specified roles.

An STS is inherently open in that its autonomous participants have an external existence. No Org can regiment all the actions of its participants. In general, a principal ought to perform only those actions for which it is authorized and not prohibited. We adopt the following "organizational design" pattern. Authorizations apply exclusively to interactions that are architecturally regimented and never occur unless authorized. Prohibitions apply exclusively to interactions that are socially regulated: a principal can perform a prohibited action but may invite sanctions by doing so.

2.3 Ongoing Work: Norms for Cybersecurity

Although the conception of Orgs and norms introduced above has broad applicability, we particularly focus on cybersecurity as an ideal setting for the exploration of normative concepts [Singh, 2015]. Of the challenges in cybersecurity, some of the most insidious ones arise in how people interact with each other and how their actions undermine security by creating and exploiting vulnerabilities through careless or malicious behavior. Such actions can arise in how software modules are developed and maintained and how inadequate governance in the social architecture leads to vulnerabilities by letting users acquire privileges they should not hold and providing insufficient incentives for good behavior or ineffective sanctions for bad behavior. The proposed approach can help in dealing with the challenges of *human behavior* and *secure collaboration* [Williams *et al.*, 2015], in particular.

One ongoing theme is a normative approach for accountability that captures accountability as a relationship between autonomous parties: the first is accountable to the second and the second has standing to hold the first accountable. Current approaches, e.g., [Haeberlen, 2010], confuse accountability with traceability, which is merely a possible mechanism for enforcement. For example, if Alice and Bob perform some actions in the reverse of some desired order (Alice goes first), we would not be able to determine who is accountable for violating the ordering requirement. It could be that Alice was late or that Bob was early or both. Perfect traceability would not solve this problem because the performance of the actions is not contested.

Given scenarios such as the above, it is important to show how the social state of an STS can be captured in information stores and queried as the participants enact their interactions. For commitments, Chopra and Singh [2015] propose an approach, Cupid, that maps commitment specifications to the relational algebra. Cupid relates the specification and progress of commitments to underlying events. In essence, the social state can be computed from event logs. Although Cupid's language is limited (e.g., it does not support aggregation), with greater expressiveness and coverage of other types of norms, such an approach could be used to help capture and improve accountability in an STS.

Another ongoing direction is the study of norms related to security-related actions, such as keeping one's computer up to date with patches and antivirus software [Du *et al.*, 2015]. How can such norms emerge in an organization? And, given different forms of monitoring and sanctioning, how effectively can an organization maintain a secure state while enabling its members to progress efficiently on their main (non-security) tasks? If some members violate these norms, how difficult is it for an adversary to compromise the organization and how easily does it revert to its ideal functioning state?

2.4 Directions for Future Work

We need rich methodologies for building STSs that accommodate stakeholders' goals [Bresciani *et al.*, 2004]. Chopra et al. [2014] propose a design process for STSs that produces a normative specification (just commitments, in their case). Challenges include (1) developing systematic approaches for handling conflicts among stakeholders while eliciting norms and (2) handling conflicts among principals while enacting an STS: a well-designed Org should handle these, but how?

How can a principal decide whether to participate in an Org? In one approach, it would compute the utility derived from participation including the risk—to be determined from the norms defined in an Org and the enforcement and conflict adjudication it offers.

How can we verify if a set of norms is consistent, whether for one role (at design time) or for multiple roles (at time of adoption)? How can we author policies and verify them with respect to norms? A related challenge is model checking an operational description such as a sequence diagram to determine whether it supports specified norms, as Telang and Singh [2012] do for commitments.

Acknowledgments

The original work was partially supported by NSF contract OCE-0957938 and OCE-0964093. This abstract and the work on cybersecurity are supported by the US Department of Defense through a Science of Security Lablet.

References

[Arrott *et al.*, 2009] Matthew Arrott, Alan D. Chave, Claudiu Farcas, Emilia Farcas, Jack E. Kleinert, Ingolf Krueger, Michael Meisinger, John A. Orcutt, Cheryl Peach, Oscar Schofield, Munindar P. Singh, and Frank L. Vernon. Integrating marine observatories into a system-of-systems: Messaging in the US Ocean Observatories Initiative. In *Proceedings of Oceans, the MTS-IEEE Conference on Marine Technology for our Future: Global and Local Challenges*, pages 1–9, Biloxi, Mississippi, October 2009. IEEE Computer Society.

[Baldoni *et al.*, 2010] Matteo Baldoni, Cristina Baroglio, and Elisa Marengo. Behavior-oriented commitment-based protocols. In *Proceedings of the 19th European Conference on Artificial Intelligence (ECAI)*, pages 137–142, August 2010.

[Bresciani *et al.*, 2004] Paolo Bresciani, Anna Perini, Paolo Giorgini, Fausto Giunchiglia, and John Mylopoulos. Tropos: An agent-oriented software development methodology. *Journal of Autonomous Agents and Multi-Agent Systems (JAAMAS)*, 8(3):203–236, May 2004.

[Chopra and Singh, 2015] Amit K. Chopra and Munindar P. Singh. Cupid: Commitments in relational algebra. In *Proceedings of the 23rd Conference on Artificial Intelligence (AAAI)*, pages 1–8, Austin, Texas, January 2015. AAAI Press.

[Chopra *et al.*, 2014] Amit K. Chopra, Fabiano Dalpiaz, F. Başak Aydemir, Paolo Giorgini, John Mylopoulos, and Munindar P. Singh. Protos: Foundations for engineering innovative sociotechnical systems. In *Proceedings of the 18th IEEE International Requirements Engineering Conference (RE)*, pages 53–62, Karlskrona, Sweden, August 2014. IEEE Computer Society.

[Du *et al.*, 2015] Hongying Du, Bennett Y. Narron, Nirav Ajmeri, Emily Berglund, Jon Doyle, and Munindar P. Singh. Understanding sanction under variable observability in a secure, collaborative environment. In *Proceedings of the International Symposium and Bootcamp on the Science of Security (HotSoS)*, Urbana, Illinois, April 2015. ACM. In press.

[Haeberlen, 2010] Andreas Haeberlen. A case for the accountable cloud. *ACM SIGOPS Operating Systems Review*, 44(2):52–57, April 2010.

[Hohfeld, 1919] Wesley Newcomb Hohfeld. *Fundamental Legal Conceptions as Applied in Judicial Reasoning and other Legal Essays*. Yale University Press, New Haven, Connecticut, 1919. A 1919 printing of articles from 1913.

[Jones and Sergot, 1996] Andrew J. I. Jones and Marek J. Sergot. A formal characterisation of institutionalised power. *Logic Journal of the IGPL*, 4(3):427–443, June 1996.

[Savarimuthu *et al.*, 2009] Bastin Tony Roy Savarimuthu, Stephen Cranefield, Martin K. Purvis, and Maryam A. Purvis. Norm emergence in agent societies formed by dynamically changing networks. *Web Intelligence and Agent Systems*, 7(3):223–232, September 2009.

[Singh *et al.*, 2009] Munindar P. Singh, Amit K. Chopra, and Nirmit Desai. Commitment-based service-oriented architecture. *IEEE Computer*, 42(11):72–79, November 2009.

[Singh, 2013] Munindar P. Singh. Norms as a basis for governing sociotechnical systems. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 5(1):21:1–21:23, December 2013.

[Singh, 2015] Munindar P. Singh. Cybersecurity as an application domain for multiagent systems. In *Proceedings of the 14th International Conference on Autonomous Agents and MultiAgent Systems (AAMAS)*, pages 1–4, Istanbul, May 2015. IFAAMAS. Blue Sky Ideas Track.

[Telang and Singh, 2012] Pankaj R. Telang and Munindar P. Singh. Specifying and verifying cross-organizational business models: An agent-oriented approach. *IEEE Transactions on Services Computing*, 5(3):305–318, July 2012. Appendix pages 1–5.

[Williams *et al.*, 2015] Laurie Williams, William Scherlis, William Sanders, David M. Nicol, Jonathan Katz, and Munindar P. Singh. Collaborating on the science of security: Active research partnership through the Lablet model. Unpublished manuscript; available from the authors, 2015.