

Trust and Delegation: Achieving Robust Interactions with Subordinates

Christopher A. Miller

Smart Information Flow Technologies
211 First St. N., Ste. 300 Minneapolis, MN 55401
cmiller@sift.net

Abstract

This paper presents arguments for the role of delegation interactions in achieving and tuning accurate trust between supervisors and subordinates (whether human or machine). The act of delegation defines an explicit frame within which trust can be more accurately assessed since it defines what both supervisor and subordinate define as the delegated instructions. It also provides a focusing effect, defining expected outcomes and processes, against which information flow and attention can be more effectively organized. Yet delegation instructions cannot be exhaustive if they are to be efficient. Thus, some authority must be ceded to the intelligence of the subordinate, and some instructions left unspecified. Shared “cultural” backplanes of assumptions are one form of organizing this intelligence for shared understanding and expectations.

Introduction

For many years, we have been advocating human-human delegation as a model for human-machine supervisory control (Miller and Parasuraman, 2007; Miller, Goldman, Pelican, 1999; Miller, et al., 2011). This led to substantial research and repeated implementations of a concept for human-machine delegation called Playbook[®] based on the metaphor of a sports teams’ “book” of shared task templates. The concept of plays, playbooks and delegation offers an interesting and, we believe, instructive perspective on human trust in automation, which we are beginning to investigate. This short paper articulates some of those thoughts.

Delegation, Playbooks & Supervisory Control

Delegation is inherently a hierarchical (rather than a truly networked) relationship. Delegation implies a supervisor who has decision-making authority over some resources and some task or goal domain. Delegation involves hand-

ing over a portion of one’s authority to subordinate(s) who are then allowed/required to make decisions and perform actions within the task, resource and goal “space” delegated. That delegation need not be complete (the supervisor may require further reviews or authorizations). It may be done by posing goals, partial methods/plans, constraints or stipulations and state-based priorities, or any combination thereof. In any event, some authority to make decisions, take actions and use resources (at least cognitive resources) must be delegated if there is to be any workload efficiency gains. Our focus has been on this supervisor-subordinate relationship since it is the way many human organizations are structured and the way many humans expect and want to interact with complex machines and automation.

Sports teams’ playbooks are just highly efficient methods of achieving delegation. A “playbook” represents pre-compiled packets of goals and partially constrained methods or plans under a simple label. A play, such as a “Hail Mary” pass in American football (where many players run as far down field as possible while the quarterback throws a pass to one of them to gain a much yardage), embodies both a goal and a constrained but not exhaustively enumerated range of methods for accomplishing it. Almost any suite of behaviors that meets the criteria will be recognized as an acceptable instance of a “Hail Mary” play. But this leaves a literally infinite variety of behaviors that would be a recognizable Hail Mary instance—a variable number of players can run downfield in variable patterns, the quarterback can throw to any one of them at various times, and they all can jink, dodge and block in infinite variations.

This potential variability, especially if expressed in a compositional hierarchy as is implicit in many human task domains (Kirwan and Ainsworth, 1999) and as we have done using Hierarchical Task Networks in Playbook implementations (Miller & Parasuraman, 2007; Miller, et al., 2011), provides two desirable features that, in turn, have implications for trust. First, it supports an ability for the supervisor to “tune” the specific behavior delegated or to negotiate it with the subordinate. The play can be con-

strued as a goal that is achievable by an infinite number of alternate task decompositions when expanded to leaf nodes. There are also methods which could achieve the play's goals but would not be seen as "legal" methods fitting the definition of the play (e.g., a running play, even though it might gain lots of yardage, would not be an instance of a Hail Mary pass). By "calling the play" with no further tuning or stipulation (or, in American football, signal calling), the supervisor essentially says "do whatever you think best as long as it is an instance of the play." But the supervisor can also provide more specific instructions by either further constraining the range of acceptable methods (e.g., "For this Hail Mary, I want you 4 guys to run in these patterns") or can use the baseline play as a reference for creating new plays (e.g., "Do a Hail Mary, but I'm going to fake the pass and run instead").

This aspect of tuning the delegation act provides flexibility to the supervisor as to specifically what is delegated, but it also allows flexibility in the effort expended in the delegation act itself (cf. Miller and Parasuraman, 2007). Calling a play, usually via customized jargon or gestures, is extremely rapid and efficient, while issuing more detailed instructions takes increasing time and effort. Negotiating or reviewing plans takes still more time but will result in increased understanding of what the subordinate will do and may result in a superior final behavior product.

Implications for Trust

The act of delegation is an expression of intent from the supervisor and forms a contract between superior and subordinate of the form "perform within the space I have delegated to you and I won't blame you." As might be expected, the notions of intent, expectations and blame are intimately bound up with trust, so it's worth thinking about the impact of delegation on trust and vice versa.

If trust is "the attitude that an agent will help achieve an individual's goals in a situation characterized by uncertainty..." (Lee and See, 2004, p. 51), then one important impact delegation will have is to make it clear(er) to both parties what those goals are. *Delegation provides a frame within which trust can be assessed and judged.* If I declare a set of intentions as instructions to you, then I can evaluate whether you have performed appropriately within them and, perhaps, even diagnose why not (Did you hear me? Did we share an understanding of the terms? Are you incapable?) Furthermore, an understanding of the requirements of the task (i.e., its subtasks and the capabilities they entail) can lead to improved assessment of the subordinate's reliability for delegated task both over time, as the supervisor learns about individual subordinates' capabilities, and at delegation time, when the supervisor uses that knowledge to select delegates.

Delegation serves to manage one of the primary sources of mismatch and misunderstanding in human-human trust formation, especially among strangers: lack of knowledge of attitude and intention (Lee and See, 2004). The supervisor-subordinate relationship establishes the responsibility of the subordinate in general (to follow the instructions provided by the supervisor) and the specific delegatory act provides instructions to be followed. In less well-defined relationships, humans must make educated guesses (based on affective or analogic cues, Lee and See, 2004) as to whether another intends to support or further their goals; but the supervisory relationship defines the attitudes which all adherents should adopt. Moreover, in human-machine delegation (with the human in a supervisory role), one of the main sources of human-human mistrust is defused: machines (to date at least) never try to intentionally mislead a human supervisor into belief that they are complying with delegated instructions while trying to pursue other ends.

While there are certainly edge cases in which a supervisor may not accurately express what s/he intends, or may include instructions which produce adverse consequences, detecting and averting these situations is not the *responsibility* of the subordinate (though it may well be quite helpful, if possible). Thus, failure to detect these situations is not a matter of disrupting trust of the supervisor for the subordinate (though it might well affect trust of the subordinate for the superior).

Another important impact of delegation, stemming from the framing described above, is that they also provide a *focusing effect which can be used to manage information flow*. The act of delegation frames a set of expected actions and outcomes which can focus information exchange on confirmation of and, more importantly, violations of those expectations. When a supervising anesthesiologist manages multiple subordinates, s/he wants to review initial plans, and then hear periodic updates on progress, but significant deviations from those expectations are allowed to interrupt ongoing activities (Turrentine, et al., 2003) because they signal unexpected conditions. In both cases, the fact of the pre-existing delegation act focuses the information exchange on a few specific actions or ranges of parameter values.

In short, an explicit delegation relationship and interaction serves to bound the set of expectations which should serve to determine trust. In Lee and See's (2004) framework for human-machine trust formation, this provides a some (though not all) of knowledge required for analytic trust—that is, deep structure understanding of how and why agent behaves the way it does which enables accurate prediction of future behavior—and thus serves to speed and tune trust formation.

Trust & “Backplane” Behaviors in Delegation

In practice, however, we are becoming aware that the above picture of trust-related delegation benefits paints a bit too rosy a picture. The reason is bound up in the nature and differences of human-human vs. human-machine delegation. As described above (and in Miller and Parasuraman, 2007), delegation is a form of instructing leaving some autonomy and work to be performed by the subordinate. At the extreme, instruction *becomes* programming, with the programmer’s need to foresee all possible interactions, contingencies, and potential failure modes. The reason humans can be tasked to perform complex duties and behave with “robust intelligence” in contexts where machines cannot stems from what has too frequently and too loosely been called “common sense”. In practice, this is essentially a shared set of goals and priorities extending beyond the explicit instructions conveyed in delegation. Consider the Hail Mary pass play again. The reason a captain doesn’t have to explicitly team players not to take out a gun and shoot opposing players, even though that might be an effective method of achieving the play, is that they all share a common framework of goals and constraints bound up not just in the rules of football, but in the rules and conventions of their society. That’s why even football rule books generally don’t contain explicit prohibitions of guns on the field.

We have begun to refer to this deeper set of rules, conventions, priorities and expectations as a “backplane” against which delegated instructions are evaluated and execution is checked. Delegation instructions are foregrounded and explicit, but unless they are “programs”, they are made within the assumptions of a backplane. Any plan or execution action which entails the violation of a “backplane” constraint, rule or goal is suspect. While especially military supervisors may delegate plans that involve the complete destruction of a subordinate, these are precisely the types of circumstances which we usually want to trigger a verification check or alert: “I can do that but I may not come back. Are you sure you want it?”

The ability to delegate within a set of such shared conventions both greatly streamlines what needs to be conveyed in instruction and ensures more reasonable behaviors from subordinates when expectations break down. The quarterback who cannot get off a Hail Mary pass may nevertheless improvise a running play to gain yardage. We would not call this a Hail Mary, but it is a good, useful behavior nonetheless. The quarterback can take this innovative action without having been explicitly told to do so, because the priority to gain yardage is a part of his “backplane” knowledge of his duties.

Of course, getting machines to exhibit “common sense” has been a nearly eternal challenge, and encoding a full and consistent set of such background knowledge is just as

challenging. Interestingly, though, some of this backplane knowledge already seems to exist in many human-machine systems. Oil refineries and nuclear power plants run complex multivariate control schemes in response to a controller’s commands (roughly analogous to “calling a play”) but these typically exist and are managed within a set of “fail-safe” behaviors built into the physical plant and the operating control system. Similarly, much of our recent work (Miller, et al., 2013) has involved “grafting” play calling onto existing control systems which include default behaviors such as returning to a UAV to a designated point when loss of control signal occurs or issuing alerts when “bingo” fuel levels are reached.

Such representations of “backplane” priorities and reasoning should be expanded and better integrated with delegation interactions for better trust relationships. For example, a play that will cause bingo fuel levels to be reached should be flagged as such and that concern raised to the supervising pilot before it is enacted. But ultimately, and interestingly, the need to submerge some degree of the priorities and reasoning of a subordinate into a “backplane” for efficiency in instructing brings us full circle to the issues of shaping and managing accurate trust. Backplane reasoning must be presumed if it is to make tasking efficient. Therefore, it can be either accurately or inaccurately understood, leading to behaviors deemed either expected and reasonable or not.

Delegation relationships help by framing the interactions and removing otherwise free variables that would be sources of mistuned ‘attitudes that the agent will help achieve my goals’. Anyone who enters into a delegation relationship is expected to adopt a set of priorities, among them being the adherence to instructions provided by a superior. But delegation is not a panacea since, for any even moderately complex system, the interaction between what is explicitly delegated and what is left to backplane reasoning will produce potentially unexpected results. Thus, the standard methods of trust formation through analytic knowledge, analogy and affect (Lee and See, 2004) still apply, but with delegation relationships to formalize, restrict and sharpen focus for the expected variations, the task of gathering and interpreting relevant data about the relationship should be speeded.

References

- Kirwan, B. & Ainsworth, L. 1999. *A Guide to Task Analysis*. London: Taylor & Francis.
- Lee, J. D. & See, K. A. 2004. Trust in Computer Technology: Designing for Appropriate Reliance. *Human Factors*, 46(1):50-80.
- Miller, C., Draper, D., Hamell, J., Calhoun, G., Barry, T., and Ruff, H. 2013. Enabling Dynamic Delegation Interactions with

Multiple Unmanned Vehicles; Flexibility from Top to Bottom. *Lecture Notes in Computer Science* 8020:282-291.

Miller, C., Shaw, T., Musliner, D., Hamell, J., Emfield, A. and Parasuraman, R., 2011. Delegation to Automation: Performance and Implications in Off-Nominal Situations. *Lecture Notes in Computer Science* 6781: 322-331.

Miller, C. & Parasuraman, R. 2007. Designing for Flexible Interaction Between Humans and Automation: Delegation Interfaces for Supervisory Control." *Human Factors*, 49(1): 57-75.

Miller, C., Pelican, M. and Goldman, R. 1999. High Level 'Tasking Interfaces' for Uninhabited Combat Air Vehicles. In Proceedings of the International Conference on Intelligent User Interfaces, 109-116. New York, NY.: Association for Computing Machinery.

Turrentine, B., Calland, J. F., Adams, R., Shin, T., & Guerlain, S. 2003. Studying communication patterns during surgery. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 1488-1492. Santa Monica, CA.: Human Factors and Ergonomics Society.