



Frameworks for Supervisory Control: Characterizing Relationships with Uninhabited Vehicles

Christopher A. Miller
Smart Information Flow Technologies LLC.

This article reports efforts to devise frameworks for quickly and easily characterizing 13 different supervisory control applications developed and tested by the international participants in a North Atlantic Treaty Organization working group on supervisory control of multiple uninhabited vehicles (UVs). Multiple prior frameworks for supervisory control are reviewed, and two novel ones (a more complex one involving seven descriptive parameters and a simpler one involving only two) are proposed that build on prior work. These frameworks are used to characterize the group's technology demonstrations. The insights gleaned from each analysis are discussed in the context of the needs and technology usage assumptions shared by the group, along with the cost-benefit tradeoff of applying such frameworks.

Keywords: Supervisory control, human-automation interaction, multiple uninhabited vehicles, framework, taxonomy, Fan Out, intervention demand, polar chart

1 Introduction

As robots, uninhabited vehicles of all types, and automated systems in general proliferate, there is an increasing need to be able to characterize such systems in ways that simplify description and yet accentuate their similarities and differences. Of particular interest to the human-robotics interaction (HRI) community are descriptions of control methods and relationships across potentially highly diverse sets of human-system interactions.

Many modern and future HRI systems will seek to establish some form of supervisory control relationship (Sheridan, 2002; Sheridan & Verplank, 1978)—that is, one in which some human serves as a supervisor who commands or delegates specific services to robot(s) or automation. There are exceptions to this relationship in HRI, of course, such as caregiver robots and robotic coaches or advisors. Nevertheless, supervisory control of automation is of interest in many human-automation interaction applications and is of particular interest to the wide-ranging military applications which are the ultimate goals of North Atlantic Treaty Organization (NATO) researchers who generally seek means of obtaining increased capability from automation without removing the human from the control and decision-making loop. While supervisory control is not a new concept in human interaction with automation, new capabilities in robotics and uninhabited vehicles (UVs) are making it increasingly feasible—and feasible in novel ways, with a larger and broader set of vehicles and other assets. “Control” can now exist along a spectrum of multiple operators, perhaps at multiple levels of an organization, and it can be shared in various ways among them. Machines have different capabilities, and the contexts in which they are used may be more or less benign. This explosion of new capabilities creates a need to investigate and,

Authors retain copyright and grant the Journal of Human-Robot Interaction right of first publication with the work simultaneously licensed under a Creative Commons Attribution License that allows others to share the work with an acknowledgement of the work's authorship and initial publication in this journal.

ultimately, to characterize which forms of control are feasible—and perhaps more importantly, desirable—in applications where many lives may depend on right and wrong behaviors from a human-automation team. Of course, there are many different types of supervisor/subordinate relationships and many different kinds of supervisory control. When and how frequently the control inputs are provided relative to action, and where in a broader organizational hierarchy the supervisor and subordinate automation exist, are among the many parameters that may vary.

The author recently participated in NATO Research and Technology Organization (RTO) Human Factors and Medicine (HFM) Panel 170/217, titled “Supervisory Control of Multiple Uninhabited Systems—Methodologies and Human-Robot Interface Technologies,” in which approximately 25 individuals from eight member countries participated in a multi-year series of discussions about the nature of supervisory control and shared details about (and sometimes collaborated in the development of) more than 13 different HRI applications or “technology demonstrations” (tech demos).

This group consisted of government, university, and industry researchers from some of the NATO member countries who were engaged in research and development of technologies associated with human control of multiple UV systems. By charter, the ultimate focus of this panel was on supervisory control of multi-robotic systems to be used by military personnel. That said, a wide range of systems, authority-sharing concepts, and application domains was considered. While supervisory control is not the only paradigm for human interaction with multiple robots, it offers obvious attractions for domains in which decisions can have life-threatening consequences, where situations are unpredictable and yet the human must operate and retain effective control over as large a range of assets and behaviors as possible. All participants were involved in ongoing research and development of such systems under funding from their parent countries, and the set of systems and concepts considered offers insight not only into how one international community with roughly similar needs and goals regards multi-robot systems, but also into the important dimensions of variation within that community. Since each country funds its own research in these areas, variations in the approaches and applications provide de facto evidence of their different priorities in addressing the shared problem of control of multiple UVs for military purposes.

There was a strong desire within the panel for a framework or taxonomy to describe the different systems under development or consideration by group members. Such a framework was desirable for communication both within the group and with other “outsiders.” Taking a step away from the panel itself, the frameworks created provide a characterization of supervisory control relationships and document the range of alternatives considered within this international community, while also providing the opportunity for contrasting the applications and relationships used in different communities and domains. This article reviews the various frameworks used in HFM 170/217 and their associated strengths and weaknesses, and finally proposes and applies two novel frameworks, developed during the working group, to characterize each of the 13 tech demos.

2 Frameworks and Supervisory Control

Sheridan (2002) defined supervisory control as an arrangement in which “one or more human operators are intermittently programming and continually receiving information from a computer that itself closes an autonomous control loop,” but he also accentuated the human-system relationship underlying the definition: “Supervisory control derives from the close analogy between a supervisor’s interaction with subordinate people in a human organization and a person’s interaction with intelligent automated subsystems” (p. 115). This emphasis on a relationship between humans and technology implies that any framework or taxonomy of supervisory control should be a taxonomy of relationships, not of vehicles, software architectures, context domains, etc.

A framework is an organizing principle or classification scheme that presupposes one or more relevant dimensions and can characterize the similarities and differences of a range of

instances. But it is worth remembering, after George Box (Box and Draper, 1987, p. 24), that ‘all frameworks are wrong, but some frameworks are useful.’ Box was speaking of models rather than frameworks, and he meant that a model will never fully capture the details and intricacies of the real world, but that some models may capture interesting or relevant aspects of it—and, in fact, by eliminating excess detail, some models may even make it easier to see relevant relationships and distinctions. Similarly, there are always multiple ways of parsing any complex phenomenon into frameworks or taxonomies. None of these will capture the full richness of the phenomenon, but some of them—indeed, multiple versions—may be useful because they organize the phenomenon in helpful, insightful ways. There is no single “right” framework for describing supervisory control, but different ones may be more or less useful for different purposes, and frameworks should be evaluated for their utility in making useful distinctions with minimal effort in terms of both application and understanding.

For the purposes of framework discussions associated with HFM 170/217, several objectives were paramount. First, we wanted a simple-to-use and simple-to-understand method of characterizing the space of possible supervisory control relationships, along with the ability to place instances of supervisory control systems within that space, allowing shorthand communication about their similarities and differences. Second, we wanted the framework to reveal what portions of that space had been explored by a given study or studies, allowing mapping of where valuable supervisory control relationship styles might reside and which regions have gone unexplored. Third, we wanted the framework to cover and account for at least the set of technologies being explored, and yet reveal distinctions between them.

3 Prior Frameworks Explored

The models eventually used to characterize the HFM 170/217 tech demos (presented below) evolved through discussions of the strengths and weaknesses of prior frameworks, developed by the author and others, and then attempted to carry the strengths forward into new frameworks while mitigating perceived weaknesses. Brief highlights of this process and of some of the frameworks examined are discussed below. While multiple additional frameworks were examined at least briefly, these represent the core line of our thinking.

Almost concurrently with Sheridan’s defining the term and concept “supervisory control,” he proposed a framework for characterizing it: a spectrum arrayed as a 10-item list of what might be described as authority levels. Its endpoints were full control autonomy for the human with essentially no role for automation and vice versa (Sheridan & Verplank, 1978), while the intermediate levels represented alternative supervisory control relationships wherein the automation has less than full authority; for example, offering alternatives (level 2), recommending an alternative (level 4), or executing an automation-selected alternative with reporting to the human operator (level 7).

Sheridan’s spectrum is very simple and very well known, but it has been criticized (in Parasuraman, Sheridan, and Wickens, 2000, among others) for offering a spectrum of authority levels without an ability to discriminate what the automation is doing (i.e., what types of tasks). In this sense, it is essentially unidimensional. Furthermore, it primarily applies to automation that makes decisions, offers suggestions, and/or executes actions, and is not as applicable to automation that senses or analyzes information without offering advice. Parasuraman et al. (2000) strove to correct these limitations by applying a simple stage model of human information processing to Sheridan’s authority-level spectrum to provide a two-dimensional information processing stage X autonomy level model. The four information processing stages (information acquisition, information analysis, decision and action selection, and action implementation) must be accomplished to perform most tasks. Most human-automation systems can be characterized by a mix of levels of authority (on Sheridan’s initial spectrum) across these four functions, and different systems may show different patterns of autonomy levels across the different functions.

This model has gained extensive acceptance in the research community, and systems

have begun to be developed in accordance with it. Research has shown different effects of automation at the different stages (e.g., Galster, 2003). Nevertheless, the simple division into four information processing stages was regarded by members of HFM 170/217 as too coarse to provide adequate distinctions between the various supervisory control systems being explored—a fact that might be due to the similarity of purpose of the different systems considered by the NATO working group. Furthermore, some group members reported that prior attempts to use this framework to describe and define systems had encountered difficulties in making clean distinctions between the information processing stages.

Miller and Parasuraman (2003) suggested that the two-dimensional model above was overly simple and showed that the four abstract information processing steps could be replaced by a detailed task decomposition model (e.g., Diaper & Stanton, 2004) to provide a more precise picture of what, exactly, the automation and human components were doing at each sequential step of a process. This led to Miller, Goldman, and Funk's (2005) claim that any human-human or human-machine delegation act (required for supervisory control) could be characterized as existing within a three-dimensional “space” consisting of a *level of abstraction* in a task-hierarchical sense (the supervisor delegates the task along with the responsibility to plan and execute all necessary subtasks to the subordinate), a *level of aggregation* of resources (the delegation entails the requirement to make use of some resources, at least the subordinate's own senses and decision-making capabilities, to perform the task), and a *level of authority* (in Sheridan's sense), which need not be complete, over both tasks and resources to accomplish the task delegated.

This model, known as LoA³, provides a rich description of delegation actions, but it had a significant fault as a framework for discriminating supervisory control relationships—it describes only the act of delegating, not the context in which that act occurs. Clearly, context can significantly color the experience and relationship of supervising and should be accounted for in any framework. So, we augmented the LoA³ model by embedding it within contextual dimensions and in performance-impact descriptions to provide the CLAMP³ model (Funk & Miller, 2008). CLAMP³ stands for C³ (three dimensions of Context) + LoA³ + Mapping for Predicting of Personnel Performance. In other words, CLAMP³ takes the three levels describing the delegation interaction from LoA³ and “wraps” them in a description of the context in which that delegation action takes place (the three context dimensions) along with a description of the resulting performance metrics for the human-machine system. The three context dimensions used were (1) a description of the *situation complexity*, (2) a description of the *capabilities of the operator*, and (3) a description of the *capabilities of the uninhabited vehicle* (or other automation). At the “other end” of CLAMP³ is a description of the outcome or effects of a delegation relationship within the context described—that is, performance measures in terms of both mission and human performance.

While CLAMP³ has been used as the framework for a series of supervisory control experiments (Shively, Flaherty, Miller, Fern, & Neiswander, 2012), it retains problems for use as a framework. First, it inherits the problem of LoA³ and of Miller and Parasuraman (2003) in that it is based on a hierarchical task decomposition to describe the delegation relationship. While this provides a rich and detailed depiction of the activities that the human and system perform, it is likely too rich and detailed to be convenient and easily understood in describing systems. More importantly, while CLAMP³ calls attention to the need to situate a description of a supervisory control relationship in a context and to describe its effectiveness, methods of representing these dimensions are underspecified. CLAMP³ tells us, for example, that it is important to consider the complexity of the context, but gives us no metric or even set of factors that might contribute to representing that complexity. The main contribution of CLAMP³ was to remind us to include these dimensions as we moved forward in trying to specify a framework for use by this group.

4 New Frameworks for Supervisory Control Relationships

Review and analysis of the prior frameworks, as sketched above, led us to propose a pair of alternate frameworks for use in characterizing the range of technology demonstrations being created during HFM 170/217. These will be presented below, after a brief discussion of the working group and the tech demos produced.

4.1 NATO RTO HFM 170/217 and its Technology Demonstrations

Unusual for a NATO working group, HFM 170/217 chose to focus on active collaboration among participating nations to foster sharing of ideas, experiences, and even techniques and technologies rather than on a theoretical review and discussion of existing literature. During the course of the working group, collaborations of various sorts occurred between participating individuals on more than 13 ongoing projects. At the end of the working group (in May of 2012), all 13 projects were presented and documented in a final report (NATO RTO HFM-170, forthcoming). These projects, as documented in that report and as observed by the author, were the source of systems and supervisory control relationships to be organized by the frameworks described below. A very brief summary of the 13 projects is provided below. More information about each project can be found in NATO RTO HFM-170 (forthcoming), as well as in the references provided below.

- *CAN-1*— Explored hand off across two crews of long-duration, small uninhabited aerial vehicles (UAVs) during sea lane monitoring missions—one crew was responsible for takeoff, landing, transition, and return, while the other was given sensor and limited maneuvering control when in range to provide surveillance data (Cruck & Lygeros, 2007; Millan & O’Young, 2008).
- *CAN-2*— Explored use of an intelligent adaptive agent to manage a multi-modal display (including tactile and auditory components) to aid in control of a single, simulated Medium Altitude, Long Endurance UAV (Hou, Kobierski, & Brown, 2007; Hou, Zhou, & Arrabito, 2011).
- *FR-1*— Explored human interaction concepts with a simulated UAV swarm to perform area surveillance and intruder detection and tracking. Alternate levels of human involvement were tested, including influence of individual UAV paths and higher level allocation of UAV sets to tasks.
- *FR-2*— Explored human/automation authority sharing via automated, dynamic function allocation in response to contextual cues indicative of higher workload. Involved two dual-operator teams controlling three heterogeneous UAVs in simulation. Scenarios emphasized sensor handoff across teams and automation support focused on trajectory management (Leal, Bouchet, Langlois, & Jourde, 2009; Villaren et al., 2010).
- *GER-1*— Manned-unmanned teaming of tactical UAVs within manned helicopter mission. Both high-fidelity, human-in-the-loop simulation and real UAV flights were demonstrated. Cognitive agents provided intent-based UAV guidance, planning, and decision-making support, and workload- and resource-adaptive interaction in a “cooperative control” arrangement (Schulte & Donath, 2011; Strenzke et al., 2011).
- *NL-1*— Telepresence control of a ground-based robot through headtracking, integrated stereo video, and 3D audio with sensed head position integrated with camera angle and rotation. Military participants performed indoor and outdoor reconnaissance tasks for evaluating telepresence performance (Elliott, Jansen, Redden, & Pettitt, 2012).
- *PT-1*— Explored distribution of workload among two operators controlling up to five UAVs through automated task allocation. Task allocation automation reasoned about operator skills and certifications to perform differing functions, as well as dynamically creating consoles to support tasks to be performed by operator. Most research was done in simulation, but some involved real, small UAVs (de Sousa, Pereira, & Gonçalves, 2010; Gonçalves, Sousa, Ferreira, Pinto, & Gonçalves, 2011).
- *SWE-1*— Explored operator control of 1–3 uninhabited ground vehicles (UGVs) in a

simulated urban environment to analyze workload and attention splitting effects. Attention to vehicles was manually switched on a cyclic basis on the operator console, while the only automation consisted of maintaining commanded heading and velocity when unattended by the operator (Lif, Hedström, & Svenmarck, 2007).

- *UK-1*— A series of integrated trials with multiple simulated and real heterogeneous UAVs (fast fixed wing and rotorcraft). Interactions were through a shared pool of UV resources drawn on by multiple service requesters. Emphasis was on realistic levels of organizational approvals and resource requests.
- *US-1*— Technology demonstration of single or paired operators controlling multiple (up to four) small UAVs in intelligence, surveillance, and reconnaissance (ISR) point, route, and area surveillance missions. Emulation of realistic coordination with other operational actors. Human operator role included target detection from sensor images. Both simulated and real deployment and testing.
- *US-2*— Play-based delegation with drill-down and play refinement to control multiple, heterogeneous small fixed wing and rotary wing UAVs and a UGV to perform integrated surveillance, attack, and resupply missions. Single operator investigations included equipment failures and both simulated and partial real deployments (Shively et al., 2012).
- *US-3*— Explored interaction with an intelligent agent making decisions about how to allocate, manage, and control another tier of subordinate UGV robots. Human interaction experiments involved up to eight UGVs in simulation (Chen & Barnes, 2012).
- *US-4*— Explored human interaction with one or two uninhabited surface vehicles (USVs), primarily using vehicle simulation with limited deployment. Focused on USV transit to/from sensor deployment operations for anti-mine or anti-submarine warfare under realistic, high-workload conditions, including emergencies, system failures, etc. Multiple levels of control were possible (Osga et al., forthcoming).

Note that these systems, while all under development for military applications, nevertheless represent a wide range of application domains and purposes and a variety of solution approaches. These reflect the priorities, resources, and needs of their respective governments, as well as the creativity and interests of the individual researchers. The goal of the framework development process was to characterize both the similarities and the range of differences among these systems.

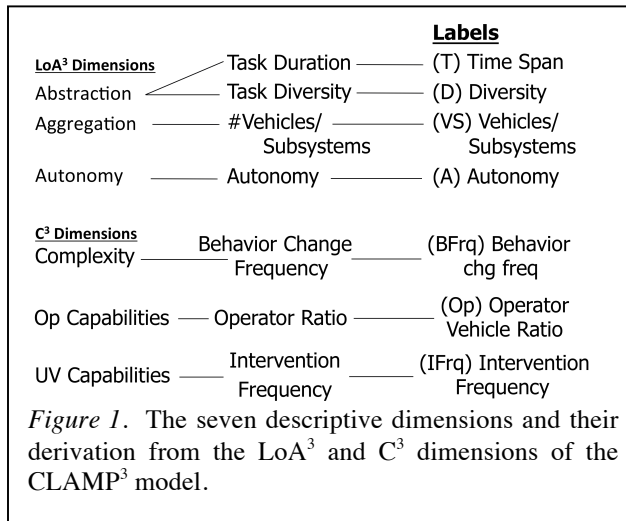
4.2 Framework #1—A Seven-Parameter Description

4.2.1 Seven-Parameter Framework Definition

The first framework developed specifically for the workshop attempted to simplify and make more rigorous the dimensions defined for the CLAMP³ model described above. In brief, we attempted to provide a scale for each of the dimensions from the CLAMP³ model by reasoning about what was important or significant for the supervisory control relationship for each of the dimensions. The scales were created to capture the range of variation, yet show interesting degrees of difference along the dimensions we saw in the supervisory control systems under consideration by the group. A secondary motivation was an attempt to synchronize the length and scalar values on each of the scales to facilitate later visualizations. To achieve this goal, we developed seven-point scales for each dimension and arrayed each of the scales from “less” to “more” supervisor-like. The specific divisions are, of course, somewhat arbitrary and debatable, but the intention was to separate each of the dimensions into seven significantly different categories—where each different region or “chunk” of the scale represented a difference that “made a difference” to the experience of the operators. The set of dimensions identified are depicted in Figure 1 (which also illustrates how they were derived from the LoA³ and C³ dimensions), while the scales for each of them are shown in the survey questions in Table 1. Each dimension is discussed below.

Abstraction. Abstraction, in a task hierarchy sense, captures the number, types, and

relations of tasks/behaviors the UV is designed for. If the top-level task in a hierarchy for a given UV can be thought of as “perform mission,” then a complete decomposition will represent all the possible tasks that the system can perform. The “size” of that hierarchy tells us important things about the mission(s) and capabilities of the UV(s) and is therefore an important dimension for inclusion in a framework. The more “abstract” the tasks and the more different types of tasks that can be delegated to a subordinate, the more “supervisory” the system might be said to be. These insights led to the proposal of two different subdimensions:



- **Mission/Task Duration (T):** Duration of missions is a reasonable stand-in for “size” of the hierarchy—what is the time span that a typical mission or task being analyzed covers? This is not the task that the operator delegates, but rather the operational window for the UVs themselves. Length/duration is a simple dimension ranging from seconds to days or weeks. A scale of interestingly different levels on this dimension (for the set of UV systems we were considering) was proposed as shown in Table 1.
- **Task Diversity (D):** Mission duration, while an interesting and simple gauge of task diversity (as Sheridan, 2002, proposed), is imperfect. I can “delegate” extremely long-duration tasks to my thermostat, but because the range and depth of tasks it performs is limited, my experience with it is not overly “supervisory.” Therefore, it is necessary to define a second dimension to identify the “complexity” of missions and of UV roles. How many different *types* of tasks is the UV involved in? How many conceptually distinct functions¹ are performable by the vehicle(s)? Table 1 presents our scale of interestingly different levels for Task Diversity.

Aggregation. Aggregation, in our initial LoA³ model, referred to the number of vehicles (or vehicle “parts” or sub-functions) being controlled by the human-machine system in an application to be characterized by the framework. Some supervisory control systems are being designed to control multiple UVs simultaneously, while others are controlling at most a single subsystem. This gave rise to the **Vehicles/Subsystems (VS)** dimension, which captures how many UVs and/or UV subsystems are typically involved in a mission (in the analyst’s focus of interest).

Autonomy. Autonomy refers to who is in charge—who is leading and who is following. For the mission as a whole, what is the relationship between human and automation? To characterize this dimension, **Autonomy (A)**, we relied on an abbreviation of Sheridan’s initial autonomy scale, as shown in Table 1.

(Mission) Complexity. We argued that contextual factors associated with the complexity of the mission could be captured by noting how often the UV has to change its behaviors (where “behaviors” are significant variations within the tasks or functions defined for task diversity above). This led us to score a dimension called **Behavioral Change Frequency (BFrq)**, defined as the average duration between required changes in vehicle behaviors (either user- or system-

¹ Of course, determining what a “conceptually distinct function” or task type is will be subject to individual judgment and to the needs and focus of the analysis. This is largely irrelevant as long as the selected level is kept approximately constant across systems to be compared.

initiated) in a typical mission of interest, with the associated scale in Table 1.

Operator Capabilities. We felt that for this dimension of context, all the required operator capabilities—while significant in their own right for training and selection of military operators, etc.—could be rolled up and reflected in how many operators are required to control the vehicle(s) in a typical mission on which the analyst is focusing—hence, **Operator to Vehicle Ratio (Op)** as characterized by the scale in Table 1.

UV Capabilities. Here, we argued the raw capabilities of the UV were not as important (and were too diversified for good abstraction in a model) as its capabilities to perform its functions without operator intervention. This, in turn, gave rise to a focus on the frequency with which the operator had to intervene in the functioning of the UV(s). This concept, proposed initially by Goodrich and Olsen (2003) and refined by Crandall and Cummings (2007), among others, was defined for our framework as **Operator Intervention Frequency (IFrq)**. This dimension was captured in a scale tied to the required frequency with which the operator had to interact with the system to achieve successful mission behavior, though note (as in the second framework below) that the percentage or ratio of clock time which the operator must spend in interaction with the vehicle might be a more accurate metric—and closer to the references listed above.

4.2.2 Seven-Parameter Framework Application and Results

Table 1. Survey used to assess responses on each of the 7 descriptive dimensions, along with associated scales.

1. <i>Mission Duration (T)</i> --What is the duration of a typical "mission" (period of intentional activity) with the UV(s)?	f. Every 5-10 sec
a. Seconds	g. Once per second or faster
b. 1-5 minutes	5. <i>Operator Intervention Frequency (IFrq)</i> --What is the average duration between required OPERATOR intervention in vehicle behaviors in a typical mission of interest?
c. 5-30 minutes	a. Once per second or faster
d. 30-90 minutes	b. Every 5-10 sec
e. 1.5-6 hours	c. Every 10-60 sec
f. 6-24 hours	d. Every 1-5 min
g. Multiple days	e. Every 5-20 min
2. <i>Task Diversity (D)</i> --How many significantly different (i.e., conceptually distinct—where the difference "makes a difference") functions can be commanded for the UV(s)?	f. Every 20-60 min
a. 1 only	g. Longer than 1 per hour
b. 2-3	6. <i>Operator-Vehicle Ratio (Op)</i> --How many operators are used to control the vehicle(s) in a typical mission?
c. 4-6	a. 4+ operators to 1 UV
d. 7-10	b. 2-3 Ops to 1 UV
e. 11-15	c. 1 to 1
f. 16-25	d. 1 Op to 2 UVs
g. 25+	e. 1 Op to 3-4 UVs
3. <i>Vehicles/Subsystems (V/S)</i> --How many UVs (or significant UV subsystems) are typically involved in the user's attention and direct command in a mission?	f. 1 Op to 5-10 UVs
a. Single subsystem	g. 1 Op to 10+ UVs
b. Multiple (2-4) subsystems, but not whole vehicle	7. <i>Autonomy (A)</i> --Overall, what is the relationship between human and automation, who is in charge?
c. One whole vehicle or 4+ subsystems	a. Human is in charge and commands specific, limited, non-integrated functions from Automation
d. 2 whole vehicles (or parts thereof)	b. Human sets overall goals, dictates tasks, but delegates moderate decision authority within isolated functions to Automation, while retaining monitoring and intervention authority
e. 3-4 vehicles	c. Human responsible for overall goals, but Automation is given large tasks which may integrate across functions. Automation may initiate actions within its functions
f. 4-12 vehicles	d. Balanced responsibilities between Human and Automation
g. Swarms (12+)	e. As for c, but Human and Automation roles reversed
4. <i>Behavioral Change Frequency (BFrq)</i> --What is the average duration between required changes in conceptually-distinct behaviors (EITHER user or system-initiated) in a typical mission?	f. As for b, but Human and Automation roles reversed
a. Longer than 1 per hour	g. As for a, but Human and Automation roles reversed
b. Every 20-60 min	
c. Every 5-20 min	
d. Every 1-5 min	
e. Every 10-60 sec	

The seven descriptive parameters were presented in the survey form shown in Table 1 to representatives of each of the 13 tech demos for HFM 170/217. Responses were provided for 10 of the 13 demos, while the surveys were completed for the remaining three by the author based on his own observations during the working group meetings and the information contained in NATO RTO HFM-170 (forthcoming). Participants were asked to answer the questions based on what they regarded as the most capable implemented version of their system—generally the most fully realized, though not (as will be seen below) necessarily the most “autonomous.” In cases where system design or adaptive automation enabled multiple possible answers, an average value was used in the figures below, though note that it would have been possible to present many of the graphs so as to show a range covered by the system rather than a single value on each dimension. Figure 3 and Figure 4 convey the results of these analyses and will be explained below, following an explanation of the visualization format in Figure 2 below.

One of the reasons for normalizing the parameter scales to seven significantly different increments and orienting them all in terms of increasingly supervisory relations was so that the results could be visualized via “radar” or polar graphs, also known as spider charts, as illustrated in Figure 2. Since each of the seven axes of the chart is partitioned and oriented similarly, a seven-sided polygon is produced by graphing the results for a single system (a single survey’s responses), providing a unique shape that can be rapidly and visually compared to other shapes from other responses. Furthermore, since all axes have been oriented so that increasing numbers convey more supervisory relationships, larger shapes (or shapes that are larger in some directions) convey a sense of more complete or complex supervisory control relationships. Finally, we have organized the seven descriptive parameters in clusters roughly affiliated with task characteristics, platform or vehicle characteristics, and operator characteristics as shown in Figure 2. Thus, expansion of the shape in any of the “quadrants” of the polar charts implies more supervisory relationships in that dimension.

Figure 3 presents a polar graph of each of the 13 tech demo survey scores. Several conclusions are immediately derivable from these figures. First, it is readily apparent that the 13 tech demos were different. At least in the terms accentuated by the framework, they explored different regions of the “space” of possible supervisory control relationships. Second, some similarities and differences are readily apparent. NL-1 and SWE-1 produced remarkably similar

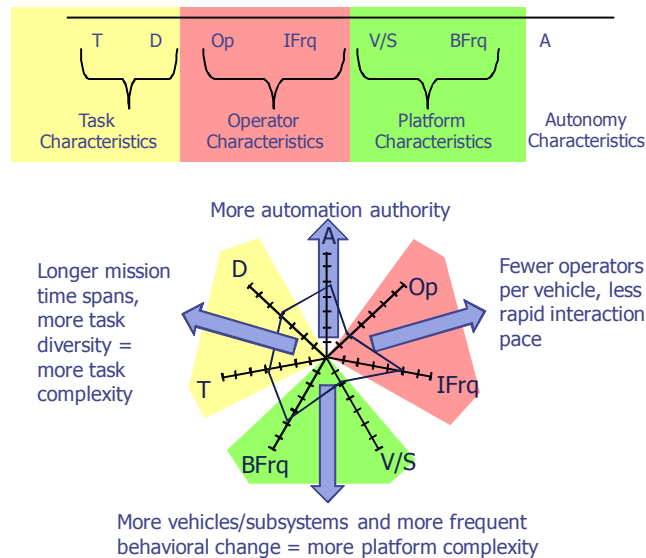


Figure 2. The layout of the polar charts used to report the 13 tech demo survey scores in Figure 3 and Figure 4.

Miller, Frameworks for Supervisory Control

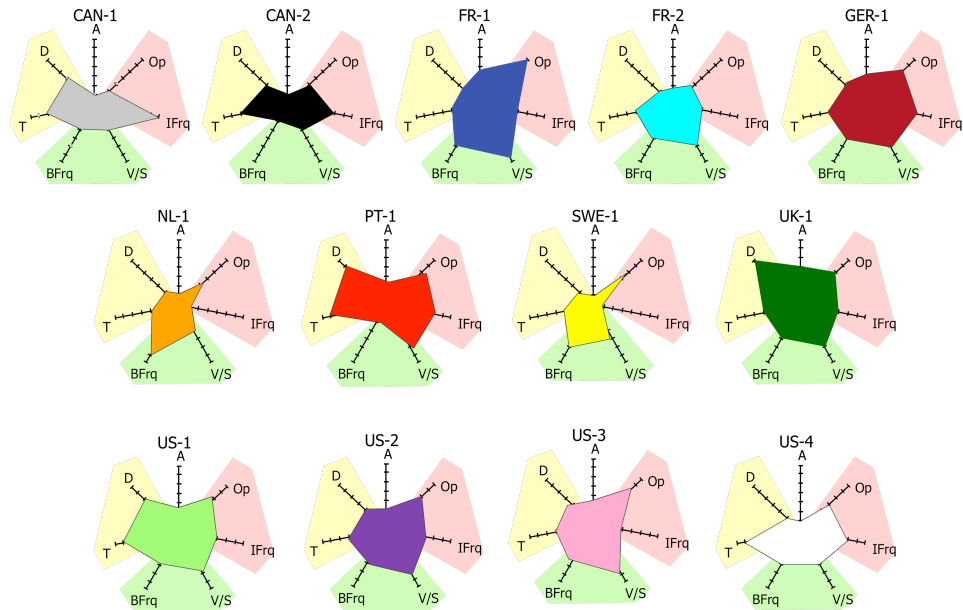


Figure 3. Polar charts graphing survey question responses for each of the 13 tech demos.

shapes, which implies that the user experience of supervisory control in those systems should be similar. In fact, the systems themselves are very dissimilar in their functionality (NL-1 primarily emphasized stereoscopic telepresence via vehicle sensor control, while SWE-1 explored automated attention switching in simulated control of 1–3 ground robots), but are arguably similar in factors that influence user experience and the supervisory control relationship: They both involved limited mission duration, frequent behavioral changes and user intervention, low amounts of vehicle/system control, and low operator:vehicle ratios. Similarly, there was a cluster of more comprehensive systems that were striving for higher forms of supervisory control, which produced similar shapes: particularly US-1 and US-2 and, to a lesser degree, GER-1, UK-1, US-3, and PT-1. Furthermore, it is worth noting that each of this group of systems produced very different survey results and corresponding polar graph shapes from NL-1 and SWE-1, whose user experience was radically different from that of this group of systems.

These results also illustrate interesting variations in the style of HRI being explored in each tech demo. For example, both NL-1 and SWE-1 embraced an approach that involved very low task diversity and low autonomy for machine systems, combined with very frequent changes required in system behaviors and frequent operator interventions—all in all, a strategy characteristic of a special purpose system with low, but frequent, operator demands. By contrast, most US systems, as well as UK-1 and GER-1, were exploring more expansive automation capabilities: greater task diversity and mission durations, more vehicles per operator but longer durations between required behavioral changes in the vehicles, and somewhat longer operator intervention intervals. This might well be taken as evidence of greater resources available or greater needs for such applications in these countries, but note that FR-1 and PT-1 represent approaches that are similar in several of these dimensions as well.

The left hand portion of Figure 4 graphs each data point on the polar chart and then overlays a bar showing the range of coverage on each of the dimensions, while the right hand portion presents a simple histogram showing the same data as a frequency of response on each of the seven parameters. Together, these two figures provide a sense of how thoroughly each of the dimensions in the framework was explored by the 13 tech demos. We can readily see, for example, that none of these projects explored higher levels of autonomy relationships (those in which automation had more control than the human(s)) or the lower end of the vehicles/subsystems dimension (where only one or a few subsystems, rather than whole vehicles, were under control of

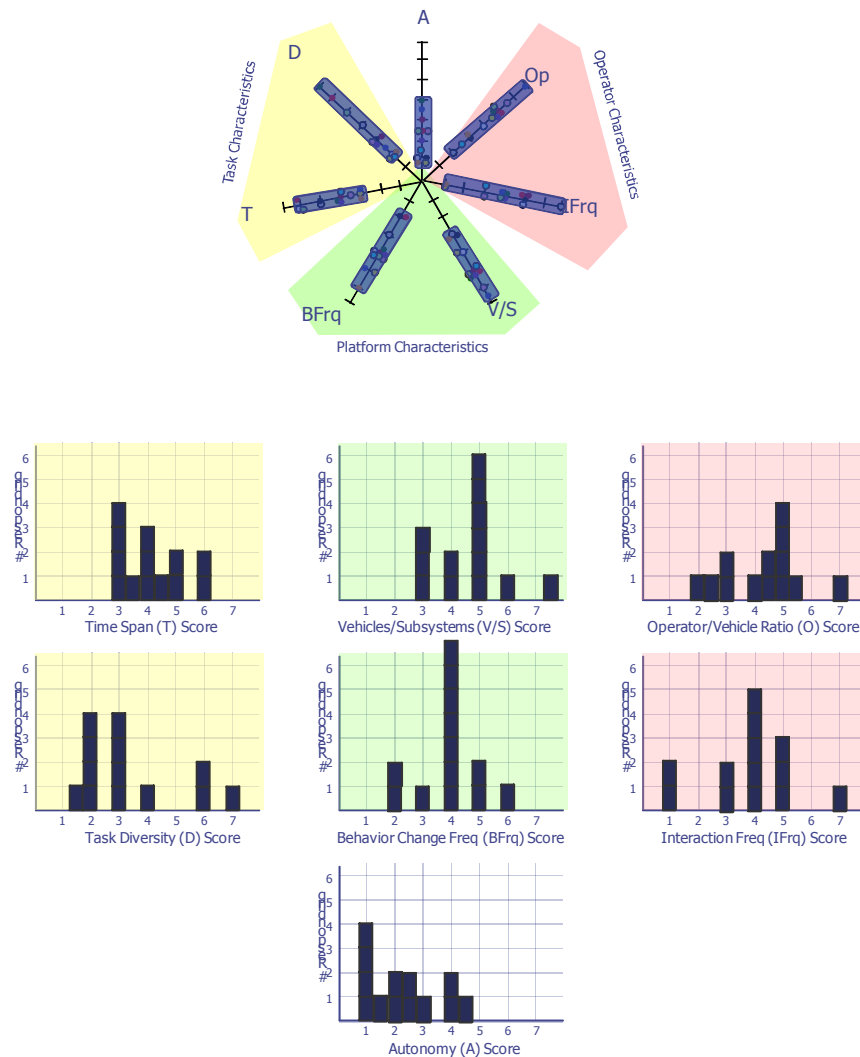


Figure 4. Frequency of responses on each of the seven parameters from the 13 tech demo responses.

the automation). Similarly, although there was reasonable coverage of the behavioral change frequency dimension across the 13 studies, it is clear that there was a strong clustering of studies involving behavioral changes slightly more than once per minute. It might be worth examining whether this is a controllability optimum or a simple coincidence of selection across the studies.

Of course, many other analyses (and visualizations) are possible, as are alternate dimensions and scales. We note, for example, that while these scales are adequate for the types of UV systems studied by this group of NATO researchers, there are certainly other forms of robotics which would fall outside these ranges. For example, Mars rover systems have mission durations (and, possibly, change frequencies) outside the scale ranges provided. Similarly, a social robot for use by autistic children could possibly be characterized by these dimensions and scales, but they would likely miss aspects of the relationship important for those applications, which are much less focused on supervisory interactions. Much greater rigor could be achieved in scale definitions, though this might well undercut the simplicity of application and use. At any rate, this approach

seems to be providing us with insights in the form of an ability to make discriminations between alternate systems, to provide indications of their similarities and differences along the specified dimensions, and to provide overall indications of which regions of the space of supervisory control alternatives have been explored and which have not.

4.3 Framework #2—A Simplified Two-Parameter Description

4.3.1 Two-Parameter Framework Definition

Initial presentations of the seven-parameter approach were met with some skepticism by members of the NATO working group. The primary objection was that the seven dimensions were too complex for the framework to be easily used and understood. Therefore, we attempted to simplify the dimensions still further to provide a two-parameter framework that would concentrate even more on the supervisory control interaction, rather than on specific attributes of the vehicles, operators, or context. We sought to roll up the impact of interaction dynamics, such as how frequently the operator has to intervene with the system, how explicit or expressive intervention must be, and what kinds of behaviors can be requested, into two covering dimensions. The following two dimensions were proposed:

- *Intervention Demand*—How much time or attention is required of the supervisor interacting with the system in order for it to achieve useful work?
- *Behavioral Scope*—What is the “scope” or range of functions and capabilities that the subordinate(s) provide(s) for performing useful work at that level of attentional demand?

Operationalizing these dimensions occupied a significant portion of our effort. While in the survey results reported below we simply asked experts in the systems to provide a scalar rating, we experimented with refining the metrics for each dimension to improve consistency as follows:

Intervention Demand. This dimension was meant to capture the amount and frequency of time and attention required by the supervisor to manage the system and achieve the work desired. Our initial thought was that Goodrich and Olsen’s (2003) Fan Out metric was a close fit to what was needed. Olsen and Goodrich actually labeled their metric “Robot Attention Demand” (RAD) and defined it by the formula $IE / (IE + NT)$, where

- IE is “interaction effort”—the time (or effort) required to interact with the robot, and
- NT is “neglect tolerance”—the robot’s effective performance time without intervention.
- $IE + NT =$ total time.

Thus, RAD is the proportion of time/effort during operation that the supervisor must devote to interacting with the automation. Based on RAD, we proposed “SAD” (System Attention Demand)—the proportion of supervisor time/effort required to interact with the system in order to perform desired work. SAD is simply extending the RAD concept to a system of robots and human supervisors. SAD is a unitless metric that ranges from 0 (for completely autonomous automation that requires no supervisory input) to 1 (no effectively “free” human time—the supervisor spends 100% of his/her time interacting to achieve the desired work). Note that there is no assumption that a rating of 1 means the human is doing everything; it merely means that the human has no spare attention or performance capability to do anything other than monitor and instruct the system.

To compare multiple systems with SAD, it is important to maintain consistent assumptions and scoring. Important considerations include the following:

- What set-up time should be included? The RAD definition was unclear (and, in fact, has been criticized by Crandall and Cummings, 2007) in that it failed to consider pre-mission planning and configuration time, as well as engineering and design time. To use SAD or RAD to compare multiple systems, any set of practices with regard to these non-execution time parameters may be used, but it is important that they be applied consistently across systems that are analyzed.
- What performance context assumptions are used? Again, when assigning time and effort

used to control and task the automation, it is also important to maintain consistency in assumptions across different systems rated if the goal is to compare those systems. For example, does the scenario of use represent a “sunny day” where nothing goes wrong or a worst case or factored error assumptions, etc.? Is the user considered a novice, an expert, or somewhere in between? What error rates are assumed for the user’s inputs? Again, the SAD metric can accommodate a wide range of different assumptions, but it is important that the assumptions be applied consistently across systems rated if comparisons are to be valid.

Behavioral Scope. There is a problem with using only SAD or RAD as the basis of comparison across supervisory control systems, however. Essentially, both compare systems *only* on the percentage of supervisor time they require; there is no explicit notion of the level of system effectiveness or work accomplished for a given level of SAD. Comparing system functionality or effectiveness using SAD requires an assumption of homogenous tasks and performance targets—even Goodrich and Olsen’s (2003) Fan Out application of RAD presumed a homogenous task: “fanning out” a set of robots to conduct a search. That said, there is no explicit notion of the domain or task included in RAD or SAD. Using SAD alone, each of the following examples would have the same “attention demand” value:

- telling a fleet of 100 UAVs to “stay put” on the tarmac (that is, to do nothing)
- telling a very highly autonomous UAV to “execute” its trip around the world
- telling an efficient secretary to plan your next month’s trips (assuming s/he already has access to your required trips and times and knowledge of your needs and preferences)

In each case, the supervisor’s attention demand is identical—one short, verbal interaction—but the examples differ radically in the *scope of work performed*. Hence, we felt a second dimension was needed to reflect the variety of tasks or functions the subordinate automation can perform. Aspects of this dimension were included in the seven-parameter model described above, primarily task diversity and behavioral change frequency. The problem is that tasks are inherently hierarchically decomposable and characterizing them across systems and domains is notoriously difficult. Therefore, to maintain consistency in comparing different applications, we would need a *common* task model for the domain of interest that is shared by the applications/relationships. This is not to say that all the systems must perform exactly the same tasks in the same way, but some basis for comparison across tasks is necessary—otherwise we would be stuck simply saying that the systems did different things. One way to provide this consistency might be to require that the systems all accomplish a shared function or goal, though perhaps through different methods. If a shared reference model could be provided for this shared function, which might necessarily be fairly abstract (perhaps akin to the high-level “Aviate, Navigate, Communicate, and Manage Systems” model for much of aviation, as in Billings, 1997), we could perhaps simply count the tasks (at a given decomposition level) that the proposed system performs, and such a count would itself provide a metric for performance scope. While we worked on various shared examples in simple, non-UV domains during the work group, we ultimately abandoned this approach in favor of the simplicity of a simple rating system in the survey results reported below. That alternate approach retains promise, however, especially for restricted domains of relatively homogenous systems, and would no doubt improve consistency in rating behavioral scope.

4.3.2 Two-Parameter Framework Application and Results

For characterizing the HFM 170/217 tech demos, we simply composed a visual scale rating question for each of the two dimensions as shown in Table 2. Again, representatives of all 13 tech demos were invited to provide ratings for their system. Ten of the thirteen did, and the remaining three ratings were provided by the author given his observations of the systems and the reported capabilities in NATO RTO HFM-170 (forthcoming). Once responses to these questions were acquired, it was trivial to graph them on a two-dimensional grid, as shown in Figure 5.

Table 2. Questions and visual scales used to characterize the tech demos for the two parameter model.

1. How much human attention and intervention does your system require in order to achieve effective work in its application domain? Please provide a rating based on the scale below:

0	1	2	3	4	5	6	7	8	9	10
Completely autonomous system; no operator intervention/attention required				Human must monitor/intervene moderately frequently but can perform tasks other than managing the system						No effective work accomplished by system without operator management and oversight/intervention

2. Of the range of tasks/functions in your operational domain, what proportion is your system involved in? Please provide a rating based on the scale below (Note that Zero (0) would be reflective of a useless system...):

(0)	1	2	3	4	5	6	7	8	9	10
(Zero/None) Function	A single, isolated Function			Many independent functions or several integrated and spanning functions						All aspects of the mission or task domain

While providing less detail than the seven-parameter approach described above, this framework nevertheless conveys some of the same conclusions—and does so in an easier-to-use and more readily perceptible format. Again, we can see the clustering of NL-1 and SWE-1 in the upper left corner of the grid, reflecting their low scope and high intervention demand. Now, however, it is more apparent that FR-2 and CAN-1 share these attributes—as might be expected for systems whose primary focus was transitioning sensor control between teams of operators with otherwise somewhat limited automation capabilities. Similarly, the cluster of US-1, US-2, and UK-1 is now readily apparent in the high-scope, high-frequency quadrant, as is GER-1 and FR-1 in the high-scope, low intervention frequency demand quadrant. This approach also conveys which portions of the supervisory control “space” (as defined by the framework) are underexplored. It is interesting to note that none of the tech demos fell into the low-scope, low intervention demand quadrant—though this would seem to be a profitable relationship, akin to a

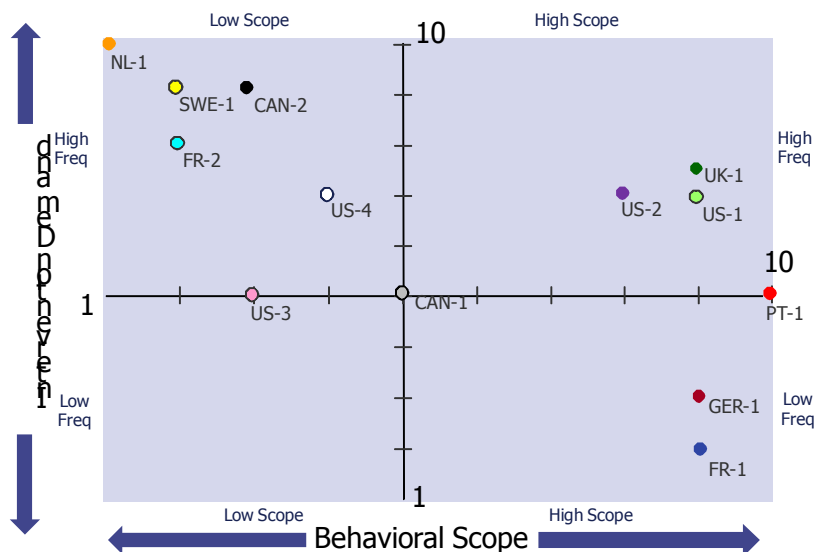


Figure 5. A graph of the 13 responses to the two parameter survey.

thermostat that can be set and forgotten, performing its restricted set of functions reliably over a long period of time.

5 General Discussion and Conclusions

Clearly, this discussion of frameworks for supervisory control has illustrated that a very wide diversity of such models is possible, each with different strengths and weaknesses. While no single model emerged as a preferred standard, we did identify several dimensions that seem relevant to discriminating between the supervisory control approaches being examined by the HFM 170/217 group, and we proposed methods for identifying and characterizing supervisory control relationships—particularly in the seven- and two-parameter models described above.

It is generally difficult for a framework, appropriately applied, to be “wrong” in the sense that it conveys false information. It may, however, be more or less useful. Furthermore, whatever utility is achieved by the framework may come at greater or lesser cost in terms of ease of application and ease of understanding its implications.

The frameworks discussed here strive for ease of use and focus on describing the supervisory control relationship. These frameworks clearly omit many important aspects of the systems we studied, not to mention aspects that may be important to other relationships or HRI goals or domains. For example, the degree of “realism,” maturity, or “technology readiness level” is not captured in any of the above frameworks, yet it is of critical importance to anyone procuring a supervisory control system. Incidentally, it would serve to discriminate many of the tech demos examined—including the otherwise similar NL-1 (which has already been deployed on real hardware and tested in realistic operational environments) and SWE-1 (which exists only as a simulation prototype).

But such omissions or information compressions are an inevitable and even desirable aspect of a framework or taxonomy. If assessing technology readiness is an important aspect of why a user wishes to apply a framework to a set of candidate technologies, there are other, well-established frameworks which may be used, and/or such data could be readily incorporated as an additional or substitute parameter in one of our multi-parameter frameworks.

Both our seven-parameter model and especially the two-parameter one are clearly easy to use. Data for the parameters were gathered via a simple survey instrument that involved seven questions for one framework and two for the other—taking a total of less than 15 minutes to fill out for most participants. Furthermore, the accuracy of the reported data used to graph results cannot be criticized without faulting the system experts themselves, since they were the ones who (for the most part) filled out the surveys. On the other hand, the consistency of application of the scales may well be questioned. Attributes such as intervention demand, the notion of a task or behavior for task duration and diversity and behavioral scope, and especially the autonomy attribute, are not crisply defined and may well have been applied differently by different respondents. As discussed above for behavioral scope, it would be possible to further refine these parameters, but that might reduce the ability to apply them across diverse platforms and applications and would certainly increase the complexity (and training) required to use the framework. This may be an appropriate direction for future work and would improve our ability to crisply define and compare alternate supervisory control systems, but it was somewhat at odds with the simplicity goals of this effort.

There is an inevitable tradeoff between expressiveness and ease of use in frameworks and other organizing structures. Our examination of the LoA³ model (and, to a lesser degree, the two-parameter model) showed that, even though the term “supervisory control” arguably defines a relationship between supervisor and subordinate, any framework which concerns itself exclusively with this relationship and does not concurrently capture aspects of the operator, system, and environment or task domain of usage is likely to be seen as insufficient. Instead, frameworks that seek to provide a basis for comparing and representing a set of alternate systems or approaches should also capture aspects of the equipment, personnel, and context of usage—especially when those aspects vary in interesting ways from system to system. Note, however, that those

dimensions need not be explicitly represented (as in the vehicles/subsystems or operator/vehicle ratio dimensions of the seven-parameter model described above). Instead, contextual aspects may be usefully captured in terms of the ways they impact the supervisory control relationship. Goodrich and Olsen's (2003) Fan Out metric, like our own intervention frequency, intervention demand, and SAD metrics, does this by reflecting how system, operator, and environmental characteristics impact the time or attention demand the operator must expend with the system in order to achieve useful work. This is part of how we achieved greater simplicity and ease of understanding in our two-parameter vs. our seven-parameter model. The tradeoff, however, is that compressing these sources of variance into a parameter that is sensitive to them makes them more implicit and difficult to track or understand. The Fan Out metric provides an adequate basis for comparison as long as aspects of the context of use are held constant, but reported Fan Out-like metrics would afford distorted comparisons of their respective systems if, for example, those systems were used by experts in some cases and novices in others. The comparison would not be *wrong* insofar as the metric was calculated correctly—after all, novices generally do require more interaction effort and achieve less neglect time—but failing to include the expertise level of the operator as a parameter in a summary framework for conveying those results would make it all too easy to forget that this variance, rather than some aspect of the automated system itself, is the source of the increased RAD score. An implication of this tradeoff is that, while a very simple framework like our two-parameter model may be attractive because it saves effort in both application and understanding, it also poses risks by suppressing potentially important data.

Finally, we need to ask whether the frameworks defined were useful. Did they make appropriate distinctions and convey insights that would not have been obtained otherwise? The answer to this question seems to be a qualified yes. As noted in the results of the seven- and two-parameter frameworks above, systems were grouped and discriminated in ways that the workshop participants found appropriate in spite of the fact that they involved very different technologies and application domains. Furthermore, some of the similarities (such as between NL-1 and SWE-1) were not ones that had been noted prior to these analyses. Similarly, the ability of the frameworks to identify thoroughly explored vs. underexplored regions on the parameters examined was deemed very useful to an overall research effort such as exploring supervisory control for multiple uninhabited military vehicles.

On the other hand, a thorough test should have two parts: (A) whether these frameworks provided useful insights for the amount of effort expended, and (B) whether there is another or different framework which would provide better insight for similar effort. With regard to part A, the extremely low effort required to apply the frameworks, combined with the novel insights obtained, seems to make the answer a clear affirmative. With regard to part B, the answer is unknown. Clearly, there are other dimensions of interest, and understanding, and determining which are most important to selecting, designing, and evaluating supervisory control systems can remain a challenge for future work.

It is worth noting that the frameworks discussed and applied in this work were both derived from and meant to apply to the types of systems under development by the members of HFM 170/217—a group explicitly chartered to discuss supervisory control of military UVs. It is quite possible that the resulting frameworks will not be adequate for other types of robotic systems and human-robot interactions. That said, supervisory control as conceptual relationship between humans and automation has had a long and productive life as a conceptual structure and remains the paradigm for a wide variety of human-robot interactions both within and far beyond military applications. Arguably, even caregiver and social interaction robots exist within supervisor-subordinate relationships—perhaps not to those for whom care is being given or with whom social relationships are being formed and extended, but for “supervisors” who program, field, and evaluate the success of such robots. Whether and how these frameworks should be applied to other types of human-robot interactions remains a question for future work, but the prevalence of supervisory control relationships indicates that there should be crossover applicability.

Acknowledgments

All 25 members of HFM 170/217 deserve thanks for participating in this effort, not just because they provided survey responses and shared details of their systems with me, but also for actively participating in a series of discussions about frameworks and their strengths and weaknesses. Particular thanks are due to Dr. Mark Draper of the U.S. Air Force Research Laboratory, who consistently championed this framework analysis effort as a part of HFM 170/217. This work is published courtesy of the NATO/RTO/Human Factors and Medicine Panel.

References

- Billings, C. (1997). *Aviation automation*. Mahweh, NJ: Erlbaum.
- Box, G. E. P., & Draper, N. (1987). *Empirical model-building and response surfaces*. Oxford: Wiley.
- Chen, J. Y. C., & Barnes, M. J. (2012). Supervisory control of multiple robots: Effects of imperfect automation and individual differences. *Human Factors*, *54*, 157–174. <http://dx.doi.org/10.1177/0018720811435843>
- Crandall, J., & Cummings, M. (2007). Developing performance metrics for the supervisory control of multiple robots. In *Proceedings of the 2007 ACM/IEEE Conference on Human-Robot Interaction* (pp. 33–40). New York: ACM. <http://dx.doi.org/10.1145/1228716.1228722>
- Cruck, E., & Lygeros, J. (2007). Sense and avoid system for a MALE UAV. In *AIAA Proceeding of Guidance, Navigation and Control Conference: Vol. 3* (pp. 2750–2765). Hilton Head, SC: AIAA. <http://dx.doi.org/10.2514/6.2007-6611>
- de Sousa, J., Pereira, F., & Gonçalves, G. (2010). Unmanned vehicle systems for environmental data collection. *Clean Technologies and Environmental Policy*, *13*(2), 360–380.
- Diaper, D., & Stanton, N. (Eds). (2004). *The handbook of task analysis for human-computer interaction*. Mahweh, NJ: Erlbaum.
- Elliott, L. R., Jansen, C., Redden, E. S., & Pettitt, R. A. (2012). *Robotic telepresence: Perception, performance, and user experience* (Technical Report ARL-TR-5928). Aberdeen, MD: US Army Research Lab. Retrieved from <http://www.arl.army.mil/www/default.cfm?page=515&id=6405>
- Funk, H., & Miller, C. (2008). *Delegation in LoA³ space* (Technical Report SIFT TR 08-LOA3-01). Minneapolis: Smart Information Flow Technologies. Retrieved from <http://www.sift.net/publications/2008>
- Galster, S. (2003). *An examination of complex human-machine system performance under multiple levels and stages of automation* (Doctoral dissertation, Catholic University). Retrieved from <http://www.dtic.mil/docs/citations/ADA420328>.
- Gonçalves, R., Sousa, J., Ferreira, S., Pinto, J., & Gonçalves, G. (2011). Authority sharing in mixed initiative control of multiple uninhabited aerial vehicles. In *Engineering Psychology and Cognitive Ergonomics, Lecture Notes in Computer Science*, *6781*, 530–539. http://dx.doi.org/10.1007/978-3-642-21741-8_56
- Goodrich, M., & Olsen, D. (2003). Seven principles of efficient interaction. In *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics* (pp. 3943–3948). Washington, DC: IEEE. <http://dx.doi.org/10.1109/ICSMC.2003.1244504>
- Hou, M., Kobierski, R., & Brown, M. (2007). Intelligent adaptive interfaces for the control of multiple UAVs. *Journal of Cognitive Engineering and Decision Making*, *1*, 327–362. <http://dx.doi.org/10.1518/155534307X255654>
- Hou, M., Zhou, M., & Arrabito, G. R. (2011). Optimizing operator-agent interaction in intelligent adaptive interface design: A conceptual framework. *IEEE Transactions on Systems, Man and Cybernetics—Part C: Applications and Reviews*, *41*, 161–178. <http://dx.doi.org/10.1109/TSMCC.2010.2052041>
- Leal, A., Bouchet, J., Langlois, G., & Jourde, F. (2009). Conception de l'interaction homme-machine et partage d'autorité : Application aux systèmes de drones [Human-machine

- interaction design and authority sharing: Application to a UAV system]. In *Proceedings IHM'09, 2ème Conférence Francophone sur l'Interaction Homme-Machine* (pp. 283–290). New York: ACM. <http://dx.doi.org/10.1145/1629826.1629872>
- Lif, P., Hedström, J., & Svenmarck, P. (2007). Operating multiple semi-autonomous UGVs: Navigation, strategies, and instantaneous performance. *Engineering Psychology and Cognitive Ergonomics, Lecture Notes in Computer Science*, 4562, 731–740. http://dx.doi.org/10.1007/978-3-540-73331-7_80
- Millan, J., & O'Young, S. (2008). Online discrete event supervisory control of hybrid dynamical systems using embedded simulation. *IFAC Journal of Control Engineering Practice*, 16, 1004–1021. <http://dx.doi.org/10.1016/j.conengprac.2007.11.007>
- Miller, C., Goldman, R., & Funk, H. (2005). *Using delegation as an architecture for adaptive automation* (Technical Report AFRL-HE-WP-TP-2005-0029). Dayton, OH: Air Force Research Laboratory. Retrieved from <http://www.dtic.mil/dtic/tr/fulltext/u2/a443116.pdf>
- Miller, C., & Parasuraman, R. (2003). Beyond levels of automation: An architecture for more flexible human-automation collaboration. In *Proceedings of the 47th Annual Meeting of the Human Factors and Ergonomics Society* (pp. 182–186). Santa Monica, CA: HFES. <http://dx.doi.org/10.1177/154193120304700138>
- NATO RTO HFM-170 (forthcoming). *Supervisory control of multiple uninhabited systems: Methodologies and enabling human-robot interface technologies* (Technical Report RTO-TR-HFM-170). Neuilly-sur-Seine, France: NATO RTO.
- Osga, G., McWilliams, M., Powell, D., Kellmeyer, D., Kaiwi, J., & Ahumada, A. (forthcoming). *Unmanned surface vehicle human-computer interface for amphibious operations* (Technical Report forthcoming). San Diego, CA: Space and Naval Warfare Systems Center Pacific.
- Parasuraman, R., Sheridan, T., & Wickens, C. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man, and Cybernetics—Part A: Systems and Humans*, 30, 286–297. <http://dx.doi.org/10.1109/3468.844354>
- Schulte, A., & Donath, D. (2011). Measuring self-adaptive UAV operators' load-shedding strategies under high workload. *Engineering Psychology and Cognitive Ergonomics, Lecture Notes in Computer Science*, 6781, 342–351. http://dx.doi.org/10.1007/978-3-642-21741-8_37
- Sheridan, T., & Verplank, W. (1978). *Human and computer control of undersea teleoperators* (Technical Report, accession number ADA057655). Cambridge: Massachusetts Institute of Technology, Man-Machine Systems Laboratory. Retrieved from <http://www.dtic.mil/cgi-bin/GetTRDoc?AD=ADA057655>
- Sheridan, T. (2002). *Humans and automation: System design and research issues*. Santa Monica, CA: HFES/Wiley.
- Shively, J., Flaherty, S., Miller, C., Fern, L., & Neiswander, G. (2012). Delegation control in control of unmanned aerial systems (UAS). In *Proceedings of the 2012 AIAA Infotech and Aerospace Conference*. Reston, VA: AIAA. <http://dx.doi.org/10.2514/6.2012-2458>
- Strenzke, R., Uhrmann, J., Benzler, A., Maiwald, F., Rauschert, A., & Schulte, A. (2011). Managing cockpit crew excess task load in military manned-unmanned teaming missions by dual-mode cognitive automation approaches. In *Proceedings of AIAA Guidance, Navigation, and Control (GNC) Conference*. Reston, VA: AIAA. <http://dx.doi.org/10.2514/6.2011-6237>
- Villaren, T., Madier, C., Legras, F., Leal, A., Kovacs, B., & Coppin, G. (2010). Towards a method for context-dependant allocation of functions. In *Proceedings of HUMOUS 2010: Humans Operating Unmanned Systems*. Retrieved from <http://websites.isae.fr/humous-10?lang=fr>

Authors' names and contact information: C. A. Miller, Smart Information Flow Technologies, LLC., Minneapolis, MN, USA. Email: cmiller@sift.net