Computers in Human Behavior 26 (2010) 840–856

Contents lists available at ScienceDirect

Computers in Human Behavior

journal homepage: www.elsevier.com/locate/comphumbeh

# Managing workload in human–robot interaction: A review of empirical studies

Matthew S. Prewett [a,*], Ryan C. Johnson [a], Kristin N. Saboe [a], Linda R. Elliott [b], Michael D. Coovert [a]

[a] Department of Psychology, University of South Florida, Tampa, FL, USA
[b] Army Research Laboratory, United States Army, Fort Benning, GA, USA

ARTICLE INFO

*Article history:*
Available online 14 April 2010

*Keywords:*
Performance
Teleoperation
Automation
Perception
Display

ABSTRACT

Working with artificial agents is a challenging endeavor, often imposing high levels of workload on human operators who work within these socio-technical systems. We seek to understand these workload demands through examining the literature in major content areas of human–robot interaction. As research on HRI continues to explore a host of issues with operator workload, there is a need to synthesize the extant literature to determine its current state and to guide future research. Within HRI socio-technical systems, we reviewed the empirical literature on operator information processing and action execution. Using multiple resource theory (MRT; Wickens, 2002) as a guiding framework, we organized this review by the operator perceptual and responding demands which are routinely manipulated in HRI studies. We also reviewed the utility of different interventions for reducing the strain on the perceptual system (e.g., multimodal displays) and responses (e.g., automation). Our synthesis of the literature demonstrates that much is known about how to decrease operator workload, but there are specific gaps in knowledge due to study operations and methodology. This work furthers our understanding of workload in complex environments such as those found when working with robots. Principles and propositions are provided for those interested in decreasing operator workload in applied settings and also for future research.



## 1. Introduction

The successful teleoperation of robots occurs at the interface of socio-technical systems. Human–robot interaction (HRI) has become an essential process for a myriad of applications, most notably in military operations and tasks that occur in extreme environments (e.g., space and oceanic exploration, disaster search-and-rescue). Through the use of unmanned aerial (UAV) and ground (UGV) vehicles, personnel can carry out tasks previously thought impossible or life-threatening. In recognition of the utility for robots, there has been an increased interest in understanding and improving HRI to improve performance in teleoperation tasks. From a human factors perspective, operator workload remains a central concern in determining successful teleoperation. Regardless of the sophistication of the technology, a robot is operated – with different levels of intervention and control – by humans. It is critical to understand this interaction of individuals and technologies. As an analogy, consider the history of accidents associated with commercial aircraft. Although many generations of technological evolution have occurred over the past 60 plus years, the cause of more than 80% of crashes is attributed to preventable human error (Wier, 2004). Even with exceedingly sophisticated and highly evolved technologies, it is not yet possible to engineer human error out of the system; so we must come to grips with understanding the limits of effective human behavior in complex technological systems.

Existing research has examined a multitude of manipulations and outcomes that outline the cognitive sources of teleoperator strain. Individual studies vary by many characteristics, including the type of workload manipulation, the apparatus used, task characteristics, and/or type of outcome measures. Due to the variability between studies, achieving a general consensus on HRI workload and performance is difficult without a comprehensive review of the literature. The current paper addresses this need by synthesizing the empirical literature on HRI workload manipulations as they relate to operator task performance. We also review several proposed solutions towards mitigating this workload (e.g., display design, platform autonomy) and provide propositions to guide future research.

Although a previous review has been conducted on workload in HRI (Chen, Haas, & Barnes, 2007), their work is limited to perceptual factors in teleoperator performance. The current paper, by contrast, provides a comprehensive review of human workload in HRI which addresses a broad range of socio-technical factors that affect operator strain as well as task performance. These factors include the number of platforms controlled, display characteristics

* Corresponding author. Tel.: +1 813 951 1162; fax: +1 813 974 4617.
*E-mail addresses:* mprewett@mail.usf.edu, mprewett1027@gmail.com (M.S. Prewett).

0747-5632/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.chb.2010.03.010

that affect operator perception, task difficulty (or demands), and the level/reliability of automation available. Based upon our summary of the literature, we draw guiding principles and propositions for reducing operator workload in HRI.

### 1.1. Information processing and response in HRI

Controlling a platform or interacting with an artificial agent consists of many tasks. Examples include executing menu functions, navigating to waypoints, manipulating a foreign object, processing information from data links, communicating with team members, and in some cases, physically moving or interacting with the platform. We describe the processes underlying human interaction with artificial agents using multiple resource theory (MRT), as described by Wickens and colleagues (2002, 2008). This model is deemed appropriate for the current review because it provides an organized and comprehensive account of the myriad of workload demands imposed by HRI tasks. MRT posits a model of time-sharing performance based upon multiple cognitive resources (vs. a single resource or task-based theory of workload). The first dimension of the model, the work process, is divided into three stages: perception, cognition, and responding. Wickens (2002) theorized that the perception and cognition stages would involve the same comprehension resources (e.g., working memory, language comprehension), whereas responding involves functionally distinct cognitive resources, such that responding to one task demand should not interfere with perceiving stimuli for another task demand. As an example of this functional separation, verbally confirming a command should produce little interference with visually tracking the environment.

The second dimension of MRT, perceptual modalities, refers to the sensory mechanisms utilized. Theoretically, tasks providing information in the same sensory modality are more likely to cause interference (or overload) than tasks using different modalities. That is, perceptual demands may be affected by the modalities in which they receive information. Based on this theory, time-sharing performance should be stronger with cross-modal cues between tasks (e.g., visual and audio) than intra-modal cues (visual and visual). The visual channel is further broken down into focal and ambient vision, based on the different cognitive structures associated with the use of each. Focal vision provides pattern recognition and processing of fine detail (e.g., reading text). Ambient vision, in contrast, guides the visual processing of movement and self-orientation.

The final dimension of resources refers to processing codes. This dimension describes separate cognitive systems involved with spatial and verbal comprehension. Processing codes are also applied in responding, through either manual or verbal actions. Given that processing codes occur across both perceptual and response stages, we expect these demands with coding resources to be associated with specific tasks, task type, and criteria. For example, responding to text alerts may interfere with team communication, as both tasks require symbolic processing of linguistic patterns. Furthermore, this interference may not even be detected if operators do not explicitly measure team communication performance, or response times to text alerts. Although processing code demands are expected to be reflected by specific task and criterion measures, these variables are infrequently manipulated in HRI studies. Thus, our framework confines the review of HRI studies to sensory modalities and work stage.

Our review of HRI categorizes workload manipulations as primarily affecting the demands placed on the operator during either visual perception or while making a response. This classification is based on the method used to increase task demands. For example, manipulations of visual display designs directly affect perception and interpretation of task stimuli. Similarly, manipulations of a performance goal (or the number of platforms), are classified as manipulations of response demands. These manipulations produce a need for either more frequent or more efficient responses by the user, whether it is engaging more targets, issuing additional commands (e.g., from multi-robot control), or increasing the tempo of providing commands. Given that perception and responses both affect task performance; we note that some overlap exists between response manipulations and sensory manipulations presented in our framework. For example, adding more robots to control may also affect perceptual demands due to additional display information. The key question to distinguish these categories, however, is operational: did the study directly manipulate features of the visual display (perceptual demands) or the performance/management requirements of the operator (response demands)?

Stemming from the distinction between perceptual and response demands, the reduction (or offloading) of tasks should vary by the type of resource requested for task accomplishment. Automation, for example, is explicitly designed to reduce the number of operator actions by offloading demands to an artificial agent. Therefore, the benefit of automation is likely to be realized when manipulating responses more so than perceptual demands. Perceptual demands, by comparison, should be reduced more effectively by new display or task designs that provide additional or effective sensory cues. Finally, it is important to acknowledge that while MRT provides predictions of operator workload, operator behavior occurs in a much broader social, organizational, and socio-technical milieu. Socio-technical factors consider the available resources for personnel and devices, the task purpose, the desired criteria, and the psycho-social characteristics of the work team. These factors should affect operator workload processes, as well as operator performance outcomes. For example, the task mission in HRI (e.g., to find survivors) will likely impact the desired criteria (e.g., overall efficiency) and the optimal device configuration to achieve those criteria (e.g., multiple robots). Thus, different socio-technical systems may yield different HRI guiding principals depending upon the task, devices configuration, and the social context.

In summary, the current study organized a review of empirical studies within the HRI workload literature using an MRT model of workload (Wickens, 2002), based within a socio-technical context. This framework is presented in Fig. 1. We separated the review by workload manipulations affecting visual or response demands. We also reviewed the evidence for several methods of mitigating these demands in HRI tasks. Display designs (e.g., visual changes, multimodal displays) are expected to affect perceptual demands more so than response demands, whereas automation processes impact response demands more so than perception. Next, we describe the literature search and study coding procedures, as well as the summary findings for HRI studies.

### 1.2. Review of the HRI literature

The literature search included several scientific and military databases, including: Academy of Computing Machinery (ACM), Defense Technical Information Center (DTIC), Google Scholar, and Institute of Electrical and Electronics Engineers (IEEE). References found in other reviews (c.f., Chen et al., 2007) were checked for eligibility. Finally, a hand search was conducted on the following journals and proceedings for the past five years: human factors, presence, human–computer interaction (HCI), and journals of the IEEE.

To be selected for inclusion in our work an article was required to report a study that compared human performance or operator attitudes/perceptions between experimental conditions designed to affect HRI. Study task and apparatus were also screened for HRI relevance. Independent variables were selected if they related theoretically to HRI workload and were examined by enough

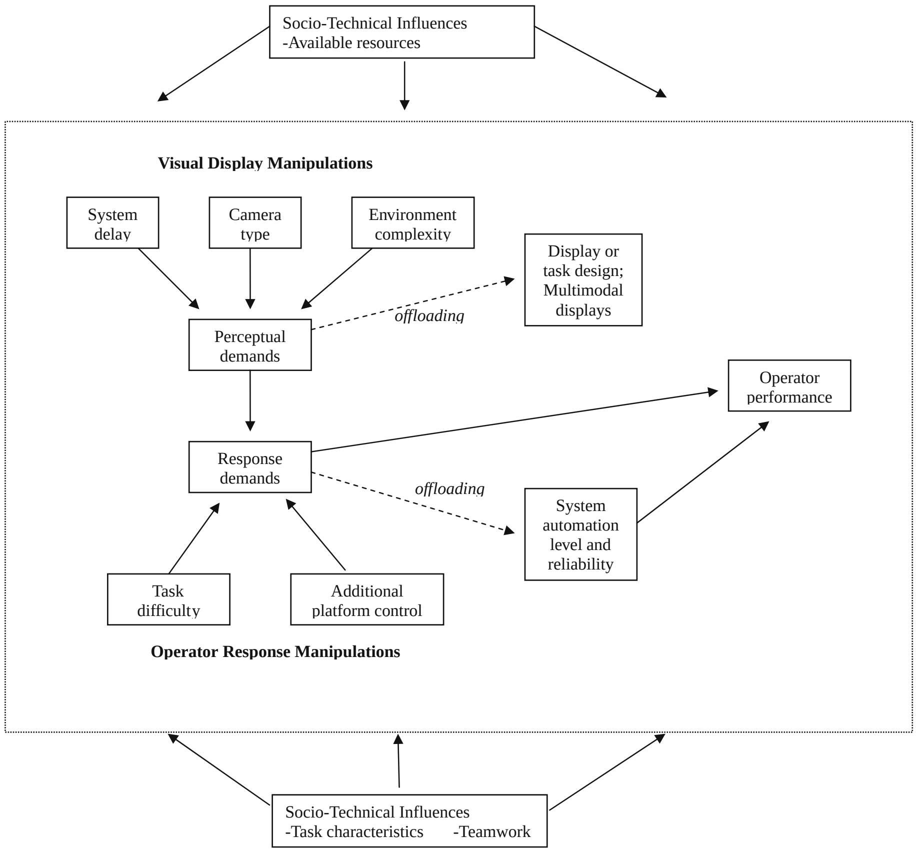


**Fig. 1.** A socio-technical multiple resource framework for workload in HRI.

studies to permit a review (e.g., Coovert & Elliott, 2009). Studies with tasks employing virtual environments (VE), artificial agents, or teleoperation were included, whereas studies using equipment for non-HRI tasks (e.g., motor vehicle simulation) were excluded.

One out of total of 4 coders placed studies in the following ten categories based upon the experimental manipulation: (1) Frame rate (FR), (2) latency, (3) field of vision (FOV), (4) camera perspective, (5) depth cues, (6) environmental complexity, (7) performance standard, (8) number of platforms controlled, (9) level of autonomy (LOA), and (10) automation reliability. Dependent variables were coded into one of the following categories: (1) task errors (e.g., incorrect actions), (2) reaction time (RT), (3) operator efficiency (e.g., time to task completion), (4) perceived workload (e.g., NASA-TLX scores), (5) situational awareness (SA), (6) usability, or (7) operator well-being (usually stress or motion sickness). Finally, study characteristics including the design (e.g., repeated measures), sample type/size, task type, and device (e.g., UAV) were noted.

## 2. Manipulations of visual demands

Teleoperation is an inherently visual task, one which uses ambient vision to guide platform navigation, and focal vision to detect critical objects in the environment or to interpret system text data. Thus, one would expect that HRI demands would primarily strain visual channels when affecting user perception. This expectation is supported by the multitude of studies investigating visual displays and visual cues. From the socio-technical systems perspective, the developments in the visual demands area are attempting to accomplish two distinct goals. The first is to ensure the camera system is capable of providing a veridical perspective to the operator. This can be seen by research in the areas of camera perspective, field of vision, and environmental cues. The second is to facilitate processing of the information by the user's perceptual system. This is provided by such factors as frame rate, response delay and depth cues. These two classifications, however, are not mutually exclusive. For example, a correct camera perspective will facilitate both an accurate presentation of the system as well perceptual processing by the user. Our review found six prominent manipulations of visual demands: frame rate (FR), response delay, field of vision (FOV), camera perspective, depth cues, and environmental detail. This review organizes these manipulations into three higher-order dimensions due to conceptual overlap: system delay (FR and latency), camera type (FOV and perspective/orientation), and environmental detail (depth cues, number of visual objects).

### 2.1. System delay

System delay refers to lags in computer image processing (e.g., to reflect updating task situations or user actions). In many cases, system delay is unavoidable due to the nature of the task or the type of resources available. For example, space exploration with artificial agents contains an inherent lag due to the distance between the operator and the robot. Thus, it is important to understand the impact of delay on operator effectiveness and error. The most commonly studied manipulations of delay are FR and response latency. FR is defined as the number of screen shots displayed over time, or the image refresh rate of a system (typically measured as frames per second). Latency refers to the temporal discrepancy between an actual event and when the event is viewed on a display or console. FR and latency are frequently addressed simultaneously by experimental methodology and defined as system responsiveness (Chen & Thropp, 2007; Darken, Kempster, &

Peterson, 2003). Existing research has also varied the consistency of the system delay as well as the extent of delay. For example, Luck and colleagues (2006) manipulated two forms of system resources: the time delay between a camera display and its operator's teleoperation of an unmanned ground vehicle (UGV) along with whether the latency was variable or consistent over trials. Delays from system processing should affect an operator's ability to visually integrate multiple screen views over time, limiting the interpretation of visual stimuli and negatively influencing SA.

Table 1 provides the study summaries of system delay manipulations of FR and system latency. Fourteen studies, reported in 10 articles (see upper panel of Table 1), address FR manipulations. Of these studies, 11 measured efficiency and errors, eight usability, three situation awareness, and two examined workload. Not surprisingly, overall findings suggest that higher FR increases efficiency, reduces errors and improves usability, amongst other criteria.

System latency/time delay was manipulated in eight studies contained in eight articles (see lower panel of Table 1). Six of these studies measured errors and efficiency. Usability and reaction time were each assessed in one study. Findings suggest that increased latency/time delays between an operating system and its operator results in decreased efficiency and increased errors rates. All but one of the studies examining fixed latency versus variable delays reported that fixed latency delays ameliorate operator efficiency and error rate.

Generally, higher FR and decreased latencies benefitted user performance. Frequently, a consistent FR was used throughout studies. Though methodologically consistent, this approach lacks external validity because FR does vary within and across HRI tasks (e.g., Darken et al., 2003). Thus, experimental studies of FR often require less operator attention since conditions are predictable.

Another concern was the impact of learning effects upon the task criterion. Most studies took two approaches to learning effects. Either participants completed practice trials prior to a study's data collection to minimize effects or the study included a measure of learning effects as part of the experiment. Several authors reported that task relevant learning led to significant increases in performance criteria in system delay conditions. Given this finding, researchers and practitioners should embrace practice and learning as a method of overcoming latency issues. When operators were aware and trained on latency issues, they were more likely to adapt to its presence (Ellis, Mania, Adelstein, & Hill, 2004; Watson, Walker, Woytiuk, & Ribarsky, 2003). Thus, pre-task awareness and training should mitigate the deleterious effects of latency on performance measures.

**Table 1**
Summary of studies manipulating system delay.

| Study | Manipulation | Criteria (by task type) | Results |
|---|---|---|---|
| *Studies manipulating frame rate* | | | |
| Calhoun, Draper, Nelson, Lefebvre, and Ruff, (2006) | 7 update rates: .5–24 Hz | Efficiency, SA, usability, and workload on UAV targeting | – Higher update rates improved subjective performance ratings – No difference on efficiency between FR conditions |
| Chen, Durlach, Sloan, and Bowers, (2008) | Normal vs. degrading: from 25 to 5 frames per second (fps) | Errors, efficiency, usability, workload, and sickness on UAV and UGV navigation and targeting | – No significant differences between presence or lack of<br>– Usability decreased with presence of latency |
| Darken et al. (2003) | 4 Update rates: 1.5–22 fps | Errors, SA, and usability during building navigation (with camera) | – No significant differences found between FR video conditions; no significant learning effects |
| Fisher, McDermott, and Fagan, (2009) | Resolution-FR combination | Usability (FR/resolution combination preference) | – Combination of high resolution/low frame rate was used most often (5 combinations from high res/low FR to low res/high FR) |
| Lion (1993) | 33 vs. 22 Hz | Errors on a tracking task using 3D computer interface | – Higher FR related to better performance; learning effects present |
| Massimino and Sheridan, (1994) | 3 fps vs. 5 fps vs. 30 fps | Efficiency in moving mechanical arm to target via camera view | – Increased FR significantly improved efficiency; the addition of force feedback improved efficiency for all FR conditions |
| Reddy (1997) | A: 2.3 vs. 11.5 Hz<br>B: 6.7 vs. 14.2 Hz | Errors and efficiency in completing a VE navigation task | – Errors and efficiency decreased with lower FR |
| Richard et al. (1996) | 6 Update rates: 1–25 fps | Efficiency in tracking and grasping 3-D moving virtual target | – Higher FR coupled with MS compensated for a lack of SS visual cues; learning effects were significant |
| Watson, Walker, Ribarsky, and Spaulding (1998) | 3 studies: 9 Hz vs. 13 Hz vs. 17 Hz | Efficiency, errors, RT, and usability on tracking and grasping of virtual object using HMD | – With lower FR, RT increased, usability decreased and efficiency was reduced; errors were not significantly effected |
| Watson et al. (2003) | 35, 75, 115 ms | Errors, efficiency, and usability on virtual object placement (HMD) | – Efficiency decreased and errors and task difficulty increased as FR decreased |
| *Studies manipulating latency* | | | |
| Adelstein, Thomas, and Ellis (2003) | Latency, Constant or random head motion rates | RT to stimuli in VE using HMD | – Only interactions were significant – changes in motion patterns resulted in a decrease in operators' discrimination abilities and latency detection |
| Allison, Zacher, Wang, and Shu (2004) | Latency delay between 2 workstations | Errors, efficiency | – Greater system latency delays reduced efficiency, increased error rates and increase the time spent making errors |
| Chen et al. (2008) | Normal vs. 250 ms delay | Errors, efficiency, usability, workload, and sickness on UAV and UGV navigation and targeting | – No significant differences between FR conditions for UAV;<br>– For UGVs, performance (hit rates) decreased with reduced FR |
| Ellis et al. (2004) | Latency detection | Errors and efficiency in latency detection of VE with a HMD | Complexity of environment failed to effect operator errors; learning effects reported |
| Lane et al. (2002) | Time delay between input and robot action | Efficiency in tracking and grabbing using UGV simulator | – Increased time delays led to a decrease in efficiency |
| Luck et al. (2006) | Study A and B: Latency rates, variable and fixed latency lengths | Errors, efficiency, and usability in navigation on UGV simulator | – Increased latency/time delay let to a reduction in efficiency and more errors; efficiency improved when time delay was fixed as opposed to variable |
| Shreik-Nainar, Kaber, and Chow (2003) | Constant or random time delay | Errors and efficiency in navigation of VE with a HMD | – When time delay was constant, as opposed to variable, errors increased and efficiency decreased |
| Watson et al. (2003) | Image latency, system responsiveness | Errors and efficiency in VE navigation using HMD | – Significant learning effects for impact of system latency |

### 2.2. Type of camera

Camera manipulations are distinguished by studies that change the range, perspective, or orientation of the viewpoints provided by the platform. These manipulations alter the environmental perspective to holistically adjust the extent to which operators are able to visually perceive their surroundings. Thus, the operator's visible range of sight is physically altered via the grounding and/or positioning of a map or camera view. For example, Darken and Cervik (1999) manipulated a virtual map to either orient "up" as north or in the direction of forward movement. The manipulations reviewed here include field of view (FOV), camera perspective/orientation, and environmental detail.

FOV describes the physical dimensions of the operator's visual screen. A typical manipulation contrasts a wide-panoramic perspective with a narrow viewpoint. Camera perspective is characterized by the immersion level of the camera in reference to a target object. Manipulations often compare a third-person, or exocentric, camera perspective, with a first-person, or egocentric, perspective. The latter is a fully immersed viewpoint. For tasks which allow for 3-axes of movement (e.g., left–right/yaw, forward–backward/roll, up–down/pitch), perspective also refers to whether the camera view is gravity- or vehicle-based.

Studies involving manipulations of camera type are presented in Table 2. FOV was examined in 10 studies (across nine articles – see upper panel of Table 2); nine measured efficiency, eight looked at errors, four examined workload, three addressed situation awareness and stress, and two accounted for self-reported motion sickness and usability. As the type of independent variables used across studies varied quite a bit, the results on FOV are mixed, but do suggest higher levels of performance with wide to moderate FOV over one more narrow. A potential downside, however, with a wider FOV are increased rates of motion sickness (Scribner & Gombash, 1998). Another finding of interest is that narrow FOV's tended to negatively affect self-reported workload more so than objective performance indices (Parasuraman, Galster, & Miller, 2003; Parasuraman, Gaslter, Squire, Furukawa, & Miller, 2005).

Ten studies addressed camera perspective (see lower panel of Table 2); nine reported measures of error, six assessed efficiency, five usability, two reaction time and one situation awareness. Overall, performance is maximized when the camera perspective is either an exocentric, third-person view of the environment or

**Table 2**
Summary of studies manipulating type of camera.

| Study | Manipulation | Criteria (by task type) | Results |
|---|---|---|---|
| *Field of view (FOV)* | | | |
| Draper, Calhoun, and Nelson (2006) | Narrow vs. Wide | Efficiency, errors, and usability on UGV search task | – Completion times were faster with a wider FOV; efficiency is incrementally improved when both wide FOV and warning are present |
| Parasuraman et al. (2003) | Visual range of camera | Efficiency and workload in virtual UGV navigation | – FOV showed no effects on criteria |
| Parasuraman et al. (2005) | FOV at 3 levels (Narrow–Wide) | Efficiency, workload, and SA in UGV navigation of VE | – Workload increased as FOV decreased; no significant difference was present for efficiency |
| Pazuchanics (2006) | Narrow vs. Wide | Efficiency, errors, and usability in UGV navigation | – Widening FOV resulted in improved performance compared to narrower FOV |
| Reddy (1997) | 2 Studies: 8 levels of FOV (.25°–32°) | Efficiency and errors on navigation task in VE | – Errors and efficiency were reduced with wider FOV |
| Scribner and Gombash (1998) | Narrow vs. Wide | Errors, efficiency, stress and motion sickness in UAV navigation | – Motion sickness was reported more frequently in wide FOV condition; no interaction was present between FOV and depth cues |
| Smyth et al. (2001) | Direct vs. 3 indirect view types (unity, wide, extended) | Errors, efficiency, workload, stress, and sickness on UGV navigation | – Wider FOV was desired for navigation but the FOV closest to typical vision was preferred for steering |
| Smyth (2002) | Indirect vs. natural vs. unity | Errors, efficiency, workload, stress and sickness on UGV navigation | – Indirect FOV resulted in decreased driving speed and more errors compared to the baseline natural vision condition |
| Wang and Milgram (2003) | 6 Comparisons of FOV | Errors and SA in navigation of UGV | – SA increased as FOV extended outward from robot; the moderate<br>– FOV condition provided the best local SA and error rate |
| *Camera perspective* | | | |
| Darken and Cervik (1999) | Map direction orientation | Errors and efficiency in UGV navigation task using camera/map | – Forward-up map alignment was best for targeted searches but north-up alignment was best for naïve and primed searches |
| Draper et al. (2006) | Camera view vs. picture-in-picture | Efficiency, errors, and usability on UGV search task | – Usability was reduced when camera perspective is placed within the virtual environment display (picture-in-picture) |
| Drury, Keyes, and Yanco (2007) | Map-based vs. video-based display | Errors, efficiency, SA, and usability for UGV search and navigation | – Video-based displays provided better performance indices, but map-based displays yielded better location and status awareness |
| Heath-Pastore (1994) | Gravity-based vs. vehicle-based | Errors in navigation of UGV simulator | – Operators reported greater confidence and SA for gravity-referenced view; gravity-based perspective also yielded fewer errors |
| Hughes and Lewis (2005) | Camera alignment and # of cameras | Errors and usability in UGV navigation and target identification | – Operator controlled cameras best for usability |
| Lewis, Wang, Hughes, and Liu (2003) | Gravity-based vs. vehicle-based | Errors, efficiency, and usability in navigation of UGV | – Efficiency and usability were significantly better for gravity-fixed display |
| Murray (1995) | Fixed vs. mobile vehicle-based view | Efficiency on target detection using camera views | – Efficiency was reduced with mobile camera views versus fixed-position cameras |
| Nielson and Goodrich (2006) | Video-only, map-only, or video-map | Errors and efficiency in UAV navigation | – Video-only displays yielded slower completion times than the other two conditions, particularly when display was 2-D |
| Olmos et al. (2000) | Exocentric vs. split-screen display | Error, Efficiency, and RT for navigation of VR terrain | – Split-screen, when displays were made visually consistent, yielded stronger performance indices than 2D and 3D exocentric displays |
| Thomas and Wickens (2000) | Third person view vs. first person | Errors, RT, and usability for navigation of UGV simulator | – Third person view yielded faster RT, fewer errors and operators reported higher levels of confidence (usability) compared to the first person view |

gravity-referenced (as opposed to being referenced towards the camera's physical direction of movement or tilt). Additionally, when a split-screen display is present (e.g., either a third-person perspective or three-dimensional image is viewed alongside a first person or two-dimensional image, respectively), performance is maximized compared to single perspective conditions (Olmos, Wickens, & Chudy, 2000).

### 2.3. Image dimensionality and environmental complexity

Image and environmental complexity studies are listed in Table 3. A summary of main findings in each area is now provided.

#### 2.3.1. Depth cues

HRI studies examining the effectiveness of depth cues (see top panel of Table 3) tend to compare monoscopic (MS) to stereoscopic (SS) displays. MS visual displays consist of a two-dimensional (2-D) image presented to both eyes which provides visual cues such as object size, shadows and the interposition of objects (Draper, Handel, Hood, & Kring, 1991). SS visual displays present a three-dimensional (3-D) image representation to both eyes allowing for greater perceived realism and, importantly for cognitive processing, retinal disparity. Retinal disparity, as in typical viewing conditions, allows for richer visual cues, complex depth cues and enhanced visual acuity. Based on Wickens' (2002) description of visual channel resources, MS displays capitalize on peripheral vision perceptual resources whereas SS displays primarily assist focal vision perceptual resources.

SS and MS visual cues were examined by nine studies within eight articles (see upper panel of Table 3). Seven reported errors and efficiency while the other criteria such as workload, usability, and self-reported stress, were each assessed within a study. A consistent finding across studies is that efficiency increased and errors decreased with a SS visual perspective. This trend should be tempered as Richards and colleagues (1996) found that when other modalities (e.g., tactile) provide additional cues for the operator or when visual conditions are optimal (e.g., high FR), MS displays perform on par with SS displays.

#### 2.3.2. Environmental detail

Environmental detail is defined as the level of visual complexity, or the number of task-irrelevant objects, within a virtual environment. This research comes at the perceptual problem from a different perspective than those studies we just reviewed. Here, the quality of operator perception depends upon the quantity of the stimuli for the teleoperator to process and discriminate. Example manipulations in this category include altering the complexity of the terrain (e.g., forest vs. desert) or changing the number of irrelevant or "distractor" targets. Consistent with Wickens' (2002) model, manipulations of environmental detail are likely to strain focal vision, as they primarily affect background detail in the virtual environment. This detail, in turn, is more likely to affect

**Table 3**
Summary of Studies Manipulating Environment Complexity.

| Study | Manipulation | Criteria (by task type) | Results |
|---|---|---|---|
| *Depth cues (SS and MS displays)* | | | |
| Drascic and Grodski (1993) | SS vs. MS | Navigation errors with robot arm | – SS display significantly reduced errors compared to MS display |
| Draper et al. (1991) | 3 Studies: SS vs. MS | Errors and efficiency during placement task using robot arm | – SS displays provided better performance indices than MS displays in difficult conditions only |
| Lion (1993) | SS vs. MS | Production and errors on 3D tracking task | – SS display was significantly related to enhanced performance and a reduction in errors |
| Nielson and Goodrich (2006) | 2-D vs. 3-D cues across display types | Errors and efficiency in UAV navigation | – Map-only display had slower completion times than map-video (2D) and video-only (3D); learning effects were detected |
| Olmos et al. (2000) | 2-D vs. exocentric 3-D and split-screen 3-D displays | Error, efficiency, & RT for navigation of VR terrain | – 2D display was detrimental to vertical maneuver performance, 3D display showed greatest deficits during lateral maneuvers |
| Park and Woldstad (2000) | 2-D vs. 3-D MS vs. 3-D SS | Errors, efficiency, and workload on placement task using robotic arm | – No significant difference between 3D MS and 3D SS; 2D display outperformed both 3D displays |
| Richard et al., 1996 | 2 studies: SS vs. MS | Efficiency in estimating virtual distances (using haptic glove) | – In baseline conditions, users were more efficient with SS than MS<br>– With high FR and multimodal cues, however, the displays yielded similar performances |
| Scribner and Gombash (1998) | SS vs. MS | Errors, efficiency, stress, & usability on UAV navigation task | – SS resulted in fewer errors, reduced stress scores, and was preferred by users (usability) over MS |
| *Environmental detail* | | | |
| Chen and Joyner (2009) | Dense vs. sparse targeting area | Targeting errors | – Errors increased with more distractor objects around the target<br>– In difficult conditions, manual control outperformed semi-autonomy |
| Darken and Cervik (1999) | Ocean vs. urban virtual environments | Efficiency in navigation | – Users had stronger performance in visually sparse ocean environments than in complex urban environments, regardless of the type of camera |
| Fisher et al. (2009) | Display image color (color vs. grayscale) | Efficiency, accuracy | – Color image enabled greater efficiency and increased accuracy for target identification compared to grayscale |
| Folds and Gerth (1994) | Dense vs. sparse targeting area | RT to identify new threat in in virtual tracking task | – RT to emerging threat was slower in dense environment<br>– Auditory warnings improved RT more so in dense environments |
| Hardin and Goodrich (2009) | 200 vs. 400 Distractor targets | Efficiency and errors in VE search and rescue | – # of distractors had a significant effect on efficiency, but not on errors<br>– Introducing autonomy did not mitigate this impact |
| Murray (1995) | Target images were complex vs. simple | Efficiency in monitoring and tracking targets in VE | – Increasing image complexity increased target detection time<br>-Automated mobility improved user performance in complex stimuli conditions |
| Schipani (2003) | Difficult vs. easy terrain | Workload ratings of UGV navigation | – Workload increased with greater terrain complexity, whereas platform speed and line of sight with the operator did not impact workload |
| Sellner, Hiatt, Simmons, and Singh (2006) | Simple vs. complex display images | Efficiency and errors on task decision-making (on stimuli) | – Simple displays decreased decision time, but also increased errors<br>– Integrative presentations reduced the time penalty in complex displays |
| Witmer and Kline (1998) (2 studies) | Dense vs. sparse virtual environment | Errors in distance estimation for Virtual environment | – More complex environments did not impact virtual distance estimation |
| Yeh and Wickens (2001) | Dense vs. sparse virtual environment | Errors, workload, and trust on target detection | – Users had better performance with low (vs. high) environmental detail<br>– With reliably cued targets, the impact of visual detail was reduced |

pattern recognition (e.g., target detection) than tasks involving platform movement and orientation (e.g., navigation).

Ten studies investigated manipulating environmental detail in a virtual environment (see lower panel of Table 3). Most of these compared targeting efficiency and errors between detail conditions, though a few also measured navigation outcomes and workload. Consistent differences involving conditions emerged from the available studies. In the case of environmental complexity, simpler was better. Across most studies, users were able to identify targets more quickly with low detail in the surrounding environment (Yeh & Wickens, 2001), few distractor targets (Chen & Joyner, 2009), or terrain that is easy to judge and navigate (Darken & Cervik, 1999). This finding is not surprising, as environments for teleoperation tasks are complex, making targets more difficult to locate through increasing demands on the visual system. But this is not the end of the story. In a demonstration that operator efficiency and effectiveness are often separate aspects of performance, studies reveal that environmental detail does not affect accuracy to the same degree it affects operator efficiency (Hardin & Goodrich, 2009; Witmer & Kline, 1998). In short, increasing environmental detail may lengthen visual search times, but it does not decrease the hit rate of critical targets.

Because HRI tasks are limited to interface and camera views, the visual channel will inherently receive greater strain than the other resource channels. Based on the evidence presented here, one may attenuate these demands, however, by reducing visual information (e.g., using integrative displays or lower environmental detail) or by offloading information to other sensory channels (e.g., tactile, auditory).

## 3. Improving perception through display design

Several common themes from the literature highlight the importance of the visual channel in determining HRI task performance. First, users have better functioning in visually sparse or simple environments (e.g., Chen & Joyner, 2009; Darken & Cervik, 1999). Second, studies that manipulated visual features to mitigate workload report a positive impact from the interventions (e.g., Park & Woldstad, 2000; Yeh & Wickens, 2001). Third, as task demands are increased, auditory and tactile feedback facilitates operator performance (e.g., Folds & Gerth, 1994). Using MRT as the framework, we can conclude that the performance effects from task demands are dependent on the types of resource channels being strained. Specifically, the evidence suggests that the demand on the visual sensory channel is typically the limiting factor on user performance. What follows are some guidelines for reducing these visual demands to the benefit of operator workload.

### 3.1. Displays with improved visual features

#### 3.1.1. System latency and FR

Generally, higher FR and decreased latencies benefit operators and lead to increased performance. These results are consistent with the notion that a more realistic image will result in less discrepancy between typical visual processing and the visual processing of technologically-altered stimuli. Technologically-altered stimuli are those either partially or wholly constructed – as in augmented or virtual environments. Thus, it appears that relatively straightforward guiding principles exist for delay issues. First, increase frame rate to a level optimal for human information processing. Second, if one is unable to minimize system delays (e.g., as in the great distances involved with teleoperation in space missions), keep the delay constant. Third, learning will occur, so provide operator training for both latency adjustment and task awareness.

#### 3.1.2. Camera perspective and FOV

Despite a wide-range of methodologies and manipulations, the study of contextual resources all indicate moderation (i.e., FOV within typical visual range) and integration (i.e., perspective and FOV presenting multiple visual displays) as a superior strategy. For example, when combined with another workload reduction method (e.g., increasing contextual information), an FOV design that allowed an operator to switch between a manual and an automated operating system was beneficial for performance (Pazuchanics, 2006). This suggests that integrating contextual resources with other interface features can decrease operator workload. In addition, some differences were noted among study tasks and criteria. Specifically, workload (Parasuraman et al., 2005) and motion sickness (Scribner & Gombash, 1998) outcomes favored a different FOV condition than task criteria. Task type also affected which FOV users preferred (Smyth, Gombash, & Burcham, 2001). This would suggest that practitioners should measure and identify the tasks and criteria relevant for their purposes to determine an optimal level of FOV.

In the related area of visual perspective, research suggests that a third-person view or a stable, gravity-based orientation facilitates performance (e.g., Thomas & Wickens, 2000). Results underscore the utility of an operator's natural spatial ability when it comes to decreasing workload and increasing performance on camera-based tasks (e.g., Darken & Cervik, 1999). We caution that the available number of studies for each type of camera perspective manipulation is small. As a result, a variety of camera perspectives warrant greater attention in order to verify these conclusions.

Guiding principles from camera studies suggest employing a moderate to wide FOV and/or a third person or gravity-referenced perspective of the task for the operator. Researchers should also monitor multiple task outcomes, including self-reported workload, motion sickness, and usability in addition to performance indices.

#### 3.1.3. Depth cues and environmental detail

The benefits of SS displays over MS displays are observable, but not overwhelming as many researchers had hypothesized. In baseline conditions, the added realism and depth cues provided by SS displays did benefit operator performance. However, in the presence of auditory alerts, MS displays mostly fared as well as SS displays. The guiding principles documented in image dimensionality and environmental complexity studies should promote a higher level of performance. First, provide SS systems if possible. When providing MS systems, have the highest possible frame rate and augment the system with cues to another sensory modality (e.g., hearing, tactile). If speed is important, eliminate as much background complexity as possible to ensure target saliency.

### 3.2. Use of multimodal displays/cues

Socio-technical systems may be constrained by a variety of factors, such as a limitation in computer hardware or inherently difficult tasks. These constraints can create visual demands beyond the control of visual display interventions. In such cases, workload may be mitigated by transferring task demands to other sensory mechanisms. Multimodal displays accomplish this goal by providing task information in alternative sensory modalities (e.g., audio, tactile). As a result, multimodal displays may frequently provide a positive solution to the workload issue in HRI. Theoretically, use of multimodal displays should mitigate workload by offloading visual demands onto cognitive resources for other senses (Wickens, 2002).

Research on multimodal displays has produced a heterogeneous and extensive body of literature. The benefit of multimodal displays across tasks has already been summarized in a number of reviews (Burke et al., 2006; Chen et al., 2007; Coovert, Walvoord, Elliott, & Redden, 2008; Prewett et al., 2006). These reviews have generally concluded that the addition and/or substitution of audio and tactile feedback provide an empirical benefit to human performance. This effect occurs across a variety of tasks and outcomes. However, several reviews have noted some differences in utility between audio and tactile feedback. Burke and colleagues (2006) found tactile feedback improved performance more so than audio when task demands were high. Other research has indicated that audio cues increase situation awareness and grab attention, whereas tactile cues can aid orientation, navigation (via direction cues), and alert responses (via tactile warnings; Chen et al., 2007). Additionally, providing feedback in multiple modalities in a complementary method appears to promote performance more so than modality substitution (Chen et al., 2007; Elliott et al., 2009). The appropriate use of visual, audio, and tactile cues, for example, should improve performance more so than visual and audio cues alone.

Within our own review of the HRI literature, multimodal displays were a viable solution to visual demands. Multimodal feedback was useful when visual conditions were poor, such as a low FR (Massimino & Sheridan, 1994) and a 2-D (MS) display (Richard et al., 1996). Audio feedback was particularly effective in improving reaction time to system alerts across a variety of workload manipulations (Dixon & Wickens, 2003; Folds & Gerth, 1994; Wickens, Dixon, & Chang, 2003). This is not surprising, given the attention-capturing qualities of audio stimuli. In summary, integrating multimodal feedback into a socio-technical system for HRI should mitigate operator performance, but implementation should follow existing guidelines for multimodal research (Coovert et al., 2008).

### 3.3. Unresolved issues in device design

A principal weakness of existing FR and latency studies is that a wide variety of delay rates have been used on different systems. Thus, it is difficult to ascertain an acceptable threshold for delay as it concerns operator performance, or if delay thresholds may vary by the type of system. Furthermore, existing research has only examined linear relationships between delay and performance through the use of ANOVA or other general linear models. Future research in these areas should seek to compare a multitude of common operations of FR and latency to determine non-linear relationships with user performance. This is important for a couple of reasons. First, it will allow the field to assess any complex effects of different FR and latency rates on learning. Second, it will help determine the threshold for cognitive processing of a realistic/real environment in contrast to an augmented or virtual one. Identification of such a threshold is critical for systems where frame rate or latency delay may not be eliminated.

*Proposition 1*: System delay variables have a non-linear relationship with performance, in which performance remains relatively constant with delay values lower than the threshold value, but degrades rapidly with delay values beyond the threshold.

For MS and SS comparisons, the current review's findings may be biased by the small number of studies, the specificity of task manipulations, and a variety of task purposes and operator instructions. As an example of these differences, several studies stress speed over accuracy, and vice versa. As a result, overall results are inconsistent regarding the advantages of SS over MS displays, although there is a consistent trend favoring SS in high difficulty situations requiring greater visual acuity. Thus, the advantages of each are highly contingent on the task difficulty and the presence of multimodal cuing.

*Proposition 2*: Task difficulty and the presence of auditory or tactile feedback interact with display type (MS or SS) to predict operator performance.

Surprisingly, relatively few studies examined the benefit of haptic (force) feedback from human–robot interfaces. Tactile and force feedback have benefited displays for many types of tasks, including aviation, motor vehicle simulations and gaming interfaces. Although haptic interfaces should theoretically assist robot control, few studies have validated such a setup for robot interfaces. We expect haptic feedback would specifically ease responding demands in robot operators, as the feedback is targeted towards manual executions of task actions.

*Proposition 3*: Haptic or force feedback for robot interfaces improves operator performance by reducing manual response demands.

Existing multimodal research has focused mainly upon the feedback or cues provided by other modalities (audio or tactile). However, multimodal *inputs* may also mitigate operator workload by offloading the demands required in manual responses. For example, some existing research has examined verbal vs. manual execution of actions (e.g., Draper, Calhoun, Ruff, Williamson, & Barry, 2003), but additional studies are needed to draw firm conclusions on this manipulation. For a preliminary review on the effect of multimodal inputs, see the review by Chen and her colleagues (2007). Based upon multiple resource theory, tasks which stress verbal processing and communication should benefit from manual execution of actions, whereas manually taxing tasks should benefit from verbal responses.

*Proposition 4*: Manual responding facilitates operator performance in communication intensive tasks, whereas verbal responding promotes performance in manual tasks.

Finally, visual display features have rarely been manipulated in conjunction with other modality features to determine additive or interactive effects. It is important that such effects are investigated so as to inform optimal design for HRI tasks. A positive example of such a result is found in the review of MS and SS displays, which provide high performance levels when both visual conditions are optimal and multimodal feedback is provided. However, the available research on such comprehensive displays remains relatively scant. Rather than simply manipulating visual information or adding auditory/tactile feedback to a baseline condition, future research should investigate such manipulations performed together. Based upon the positive results documented in modal and multimodal studies, we posit that optimizing the combination of visual, auditory, and tactile feedback would benefit operators most.

*Proposition 5*: Optimal visual displays (delay, detail, camera perspective) in combination with appropriate audio and tactile feedback will produce better operator performance than such features applied individually.

Even though display designs may be improved, it is recognized that any system will likely operate at a suboptimal level (Wier, 2004). Such complex systems invariably suffer from process loss as operators coordinate with devices, robots, and other team members. Even with intuitive visual and multimodal displays, operator performance may suffer from high demands for executing task functions within artificial agents. We now review studies that directly manipulate response demands by requiring greater efficiency and accuracy in operator actions.

## 4. Manipulations of response demands

Interaction with artificial agents imposes considerable task demands which may require continuous actions and rapid responses to external conditions during a mission. For example, successful performance during emergency search-and-rescue situations requires frequent actions for navigation and quick responses to environmental stimuli. Given the many response demands in different HRI tasks, human performance may suffer due to divided attention with multiple tasks and limited resources to compensate for this division. Researchers have realized the benefit of manipulating responding demands in order to examine the human limits in commanding artificial agents. Changing response demands can also gauge the performance limits of operators for specific tasks and situations. We now consider two frequently applied manipulations of teleoperators responding: task performance standards and the number of operator-controlled platforms.

### 4.1. Performance standards

Manipulations of task performance standards alter the desired criterion levels (e.g., changing the number of targets to hit) or increase the difficulty of responding to a task-critical object (e.g., making target radius smaller to affect accuracy). An example is provided by Galster, Knott, and Brown (2006), who manipulated the number of targets for UAV operators.

Twelve articles were identified in the literature that manipulated performance standards. Table 4 presents the study citations, type of performance manipulation, criterion and tasks measured affected, as well as key findings for studies examining task demands. The types of devices used had more variability in this sample than in multi-platform control samples. Devices ranged from a robotic arm interface (Park & Woldstad, 2000), an air-traffic controller decision-making system (Hendy, Lao, & Milgram, 1997), to flight and UAV simulations (e.g., Draper et al., 2003) and virtual environment exploration (Schipani, 2003).

Results from these studies indicated that increasing performance standards leads to reduced performance outcomes. Given the demanding tasks in HRI, it is not surprising that requesting additional operator responses will have a negative impact, as reported across the studies. Studies that manipulated performance standard also examined a wide variety of moderator variables as methods to mitigate the strain on responding. Existing evidence indicated that optimal visual conditions can reduce the impact of high performance standards (Park & Woldstad, 2000; Watson et al., 2003), whereas the response modality (verbal vs. manual) did not have an effect (Draper et al., 2003). The type of manipulation and criterion measured also appears to affect the relationships between performance standard and workload. Providing personnel with less time to complete the task improved user efficiency, but it also increased workload and task error rate (Hendy et al., 1997; Mosier, Sethi, McCauley, Khoo, & Orasanu, 2007).

Manipulations of performance standards have demonstrated that device and criterion play an integral role in HRI workload. Based on the guiding principles from our review, we suggest optimizing visual displays when tasks are anticipated to be difficult. Furthermore, the desired task criteria must be considered. A socio-technical system which values accuracy must take care to monitor and mitigate the negative impact of high performance demands. If user efficiency or overall production is desired, however, high performance standards may serve to improve operator performance. Finally, the current review combined task difficulty resulting from either task complexity or from task goals, primarily because these distinctions were not made in the extant literature. Future research protocol, however, may be well served to explicitly distinguish these two task characteristics. Based upon the goal-setting literature in applied psychology (Locke & Latham, 1990), more difficult task goals should improve operator performance for similarly complex tasks, as these goals by encourage attention, effort, and persistence. Task difficulty arising from complexity, on the other hand, should hinder operator performance simply due to the higher level of perceptual and responding demands.

*Proposition 6*: HRI task complexity and task goal difficulty will bear different relationships with task criterion, with a positive

**Table 4**
Summary of studies manipulating task performance standards.

| Study | Manipulation | Criteria (by task type) | Results |
|---|---|---|---|
| Cosenzo, Parasuraman, Novak, and Barnes (2006) | # Of targets to photo | Errors in targeting, RT to navigational decisions | – As # targets increased, targeting errors and reaction time to navigational stimuli increased |
| Draper et al. (2003) | # Of alerts needing responses | Errors and reaction time in responding to UAV alerts | – Performance degraded as system alerts were more frequent; no interaction between condition and form of responses (manual vs. verbal) |
| Galster et al. (2006) | # Of targets to process | Errors, efficiency, and workload in processing targets; RT to probes | – Workload differences emerged favoring the low target condition<br>– 4 UAVs yielded better performance with more targets than 6 or 8 UAVs |
| Hendy et al. (1997) | Low, medium, and high degrees of time pressure | Efficiency, error, and workload in air-traffic control | – Performance dropped only at high levels of time pressure<br>– Workload indices increased sharply beyond low time pressure |
| Mosier et al. (2007) | Low or high levels of time pressure | Errors and efficiency in diagnosing system problem in flight simulator | – Adding time pressure increased pilot efficiency, but also increased diagnosis errors; this was worsened by system information conflicts |
| Park and Woldstad (2000) | Size of destination for placement | Efficiency and workload in object transfer with robotic arm | – Less efficiency and higher workload in conditions with smaller targets<br>– 3D displays helped performance in with small targets |
| Schipani (2003) | Navigation distance | Workload ratings in VE navigation | – Workload increased with greater distance to travel<br>– Line of sight with the operator did not impact workload |
| Wang, Wang, and Lewis (2008) | Robot coordination demands | Region explored, victims located, and coordination demands | – Tasks with fewer coordination demands yielded higher productivity in exploration and victim location<br>– The level of coordination demands varied by the type of robot used (explorer vs. inspector) |
| Wang, Lewis, Velagapudi, Scerri, and Sycara (2009a) | # Of tasks assigned | Victims saved, area explored, efficiency, and workload in search and rescue task | – Users covered more surface area, switched between robots more frequently, and reported less workload with simple exploration task<br>– Users with search and locate tasks had worst production, but this was mitigated with control of 8 UGVs (vs. 4 or 12 UGVs) |
| Wang et al. (2009b) | Individual vs. shared robot control | Victims located, region explored, and team process measures | – Individually controlling a robot led to slightly more victims located and significantly more surface area explored<br>– Sharing control of a pool of robots introduced some process loss from team communication and coordination requirements |
| Watson et al. (2003) | Distance in 3-D placement | Errors, efficiency, and usability on virtual object placement (HMD) | – Placement errors increased with greater distances in addition to task completion time; poor FR worsened this effect |
| Yi, Song, Ji, and Yu (2006) | # Of targets to photo | Errors and SA in targeting with UAV | – Accuracy and SA decreased with more mission targets<br>– Amount of practice affected task performance positively |

relationship between goal difficulty and performance, and a negative relationship between task complexity and performance.

### 4.2. Multi-robot control

The control of multiple platforms affect response demands by increasing the number of sub-tasks requiring actions, such as navigation, alarm responses, and target acquisition. Other things being equal, providing an operator with more than one platform to control will certainly cause an increase in workload; the burning question that must be answered is if the additional strain outweighs the benefit of having more platforms to accomplish tasks. Addressing this issue requires a look at the impact that different numbers of platforms have on diverse performance criteria. We suspect that error and reaction time measures will likely degrade from the additional attention and time required in controlling an extra platform. However, measures of overall production may reflect the benefit from having an additional platform to accomplish the work.

Table 5 presents the summary information for the research in the area of multi-platform control. We note that differences are determined by statistical significance within a particular design. A total of 19 studies are examined. In general, coding revealed that most studies used counterbalanced, repeated measures designs in laboratory conditions. Populations studied ranged from students to aviation and HRI professionals. Tasks predominantly included: (a) navigating platforms to targets or areas of interest, (b) executing an action (e.g., inspection, manipulation), and (c) monitoring and responding to system displays and alerts.

When examining results by the task performance measures, we observe an emerging trade-off between overall efficiency and other measures. In several studies, users could execute more total actions and navigate to more overall waypoints as more platforms were controlled (Crandall & Cummings, 2007; Lif, Jander, & Borgwall, 2007; Squire, Trafton, & Parasuraman, 2006). However, increasing the number of platforms does have negative consequences. For example, controlling more robots increases error rates in targeting and navigation (e.g., Dixon & Wickens, 2003; Galster et al., 2006), as well reaction times to system alerts (e.g., Chadwick, 2006; Levinthal & Wickens, 2006). These results suggest that the control of multiple platforms allows the user to accomplish more tasks overall because more resources are available. However, this added productivity comes at the cost of accuracy and timely attention. Although control of one robot was optimal for task errors and reaction time across studies, control of two robots did not inhibit performance to nearly the same degree as control of four or more robots (Adams, 2009; Chadwick, 2006; Ruff, Narayan, & Draper, 2002). These studies suggest that control of two platforms might provide an optimal fit for maximizing both speeded performances as well as error rate.

A couple of variables were examined to determine if they can be utilized to lessen these negative consequences – audio feedback and increased automation (which varies in terms of level and

**Table 5**
Summary of studies manipulating the number of robots controlled.

| Study | Manipulation | Criteria (by task type) | Results |
|---|---|---|---|
| Adams (2009) | 1 vs. 2 vs. 4 UGVs | # Of actions, efficiency, and workload for search and transfer | – Slight differences between 1 and 2 UGVs, but efficiency and perceived workload were worse with 4 robots |
| Chadwick (2005) | 1 vs. 2 UGVs | Errors and perceived workload in targeting, and navigation | – No significant differences between groups |
| Chadwick (2006) | 1 vs. 2 vs. 4 UGVs | RT in target responding and navigational correction | – RT was similar between 1 and 2 UGVs but degraded from 2 to 4 UGVs |
| Chen et al. (2008) | 1 vs. 3 UGV and/ or UAVs | Errors, efficiency, SA, and workload in targeting (with navigation) | – Targeting errors were equal between 3 platforms and single UAV or UGV, but perceived workload and efficiency suffered |
| Crandall and Cummings (2007) | 2 vs. 4 vs. 6 vs. 8 UGVs for team | Errors and efficiency in navigation and target detection/transfer | – 4 and 2 UGV conditions exhibited fewest lost robots<br>– 6 and 8 UGV condition yielded highest # of target successes |
| Dixon and Wickens (2003) | 1 vs. 2 UAVs | Errors in tracking and targeting, RT to system alerts | – 1 UAV users had slightly better performance indices than 2 UAVs<br>– Adding auditory feedback improved performance across conditions |
| Galster et al. (2006) | 4 vs. 6 vs. 8 UAVs | Errors, efficiency, and workload in processing targets; RT to probes | – 4 UAV users had better accuracy and RT, but equal times |
| Hill and Bodt (2007) | 1 vs. 2 UGVs | Perceived workload in navigation and image processing | – Perceived workload was higher with 2 UGVs<br>– Operators reported different levels of impact from adding a robot |
| Humphrey, Henck, Sewell, Williamson, and Adams (2007) | 6 vs. 9 UGVs | Efficiency, workload, and SA in bomb disabling simulation | – # platforms also coincided with # of bombs to diffuse (difficulty)<br>– Performance and workload indices were similar between conditions |
| Levinthal and Wickens (2006) | 2 vs. 4 UAVs | Efficiency in UAV navigation, RT to system alerts | – Users were less efficient when controlling 4 UAVs<br>– False alarms in automation hurt performance more than false misses |
| Lif et al. (2007) | 1 vs. 2 vs. 3 UGVs | Efficiency in navigation (# of waypoints) | – 2 or 3 UGVs had equal efficiency (# of waypoints) than 1 UGV |
| Parasuraman et al. (2005) | 4 vs. 8 UGVs | Completion time for game, # of games won, workload | – Completion time and win rate deteriorated from 4 to 8 UGVs<br>– As workload increased, automation features had a greater impact |
| Squire et al. (2006) | 4, 6, or 8 UAVs | Efficiency in navigation and control (total # of actions) | – Users performed increasingly more actions with more platforms |
| Ruff et al. (2002) | 1 vs. 2 vs. 4 UAVs | Errors and workload for targeting and decision-making | – 1 UAV users had the fewest rejection errors, 2 UAV users had the best targeting accuracy, and 4 UAV users reported the most workload |
| Ruff et al. (2004) | 2 vs. 4 UAVs | Efficiency and workload in targeting; RT to system alerts | – All performance indices were better in 2 UAV condition than 4<br>– Reliability of automation, rather than level-of, had greatest impact |
| Trouvain and Wolf (2003) | 2 vs. 4 vs. 8 UGVs | Efficiency and perceived workload in navigation and target processing | – Users performed more overall inspections with 4 and 8 UGVs, but also had more idling time and efficiency loss |
| Trouvain, Schlick, and Mervert (2005) | 1 vs. 2 vs. 4 UGVs | Errors and efficiency in navigation | – Users of 1 UGV had optimal navigation performance<br>– 2 and 4 UGV users were equal in performance |
| Wang et al. (2009a) | 4 vs. 8. vs. 12 UGVs | Victims saved, area explored, efficiency, and workload in search and rescue task | – Use of 8 UGVs provided optimal production, though effect strength was affected by # of tasks assigned (more tasks yielded a stronger effect)<br>– Users of 4 UGVs reported low workload but also had little production |
| Wickens et al. (2003) | 1 vs. 2 UAVs | Errors and RT in tracking, targeting, and system monitoring | – 1 UAV users demonstrated faster RT and efficiency<br>– Errors in tracking and system failure detections were equivalent |

reliability). Multimodal feedback was not expected to improve performance with multi-robot control, primarily because multi-robot controls strains responding, rather than perceptual, processes. However, studies using multiple displays found that audio feedback facilitated faster reaction times in responding to system alerts (Wickens et al., 2003; Dixon & Wickens, 2003). This finding suggests that multimodal (e.g., audio) alerts are primarily useful for directing operator attention when it is divided between multiple robots. Another solution to the issue of divided attention is the use of integrated displays for multiple robots, as exhibited in the work of Wang and colleagues (2009a, 2009b). In the case of automation, it was reliability that made a much greater impact than the power or even type of automation (Levinthal & Wickens, 2006; Ruff, Calhoun, Draper, Fontejon, & Abbott, 2004). If automation is consistently reliable, it will be utilized to a greater extent. If it is unreliable, it does not matter how powerful the automation is, it will not be utilized. More will be said about automation below in the section on reducing workload in response demands via automation.

These results yield several principles for managing operator performance in multi-robot tasks. First, the production benefit of controlling multiple platforms should be explicitly weighed against the deterioration of other performance indices, including reaction time and errors. Researchers and practitioners need to determine which criterion is more essential to task success, and acknowledge it may be a moving target, varying according to the situation. For example, rescuing the most individuals possible is the critical outcome for search-and-rescue operations, whereas operators disabling explosives are more concerned with correct actions for each and every explosive device. Second, workload from multi-platform management may be alleviated through several techniques. In particular, audio feedback is beneficial for improving reaction time and can facilitate responses to system alerts during multi-robot control. Another potential intervention includes the use of practical and reliable automation, discussed next.

## 5. Reducing response demands through automation

In many applications, human execution of tasks has been slowly replaced by automated systems. The goal of increasing the level of automation is to lower workload by responding for the operator whenever possible. Empirical research in the area of HRI and automated systems, however, has revealed more complex relationships between the human operator, an automated agent, and their combined performance. We review the efficacy of automation based on the two prominent streams of research: level of autonomy/control (LOA), and automation aid reliability. The first, research on LOA, focuses on investigating outcomes when the balance of control between the human and autonomous agent is manipulated. The second, automation reliability research, focuses on manipulating the accuracy and frequency of automation aids in the control of robots or complex semi-autonomous systems. The impact of each on performance is now more fully considered.

### 5.1. Level of autonomy

Advances in technology increasingly allow for human operators to simply monitor a process or be minimally involved, such as through safety checks or the press of a single button. A multitude of situations exist in which humans and semi-autonomous systems or robots must work together in a more cooperative fashion. In some instances this cooperation stems from the inability of technology to fully subsume a human operator's role (e.g., air-traffic control). In other situations, an autonomous system is technologically capable of fully performing a task but legal or safety restrictions exist that require a human operator (e.g., hazardous materials handling).

Research in LOA focuses on manipulating either the amount of control a human operator has over an automatic process, or the amount of autonomy a robotic entity or system has. The LOA may either be inherent, as in an expert system, or may be 'allocated' by a human operator. For our purposes, studies in this area assess one of two task types: human teleoperation of one or more robots and human supervision and control of semi-autonomous systems.

Researchers have long noted that the most common implementation of automation in an applied setting involves allocating as much responsibility to an automated system as is technologically possible (Kaber, Onal, & Endsley, 2000). If multiple tasks can be automated and supervised by a single operator, this configuration often results in workers who observe the process and are unable to intervene. Operators are essentially left out of the loop. Since most automation is inherently imperfect – see again the arguments presented at the beginning of this article concerning socio-technical systems – failures of automation or unsuccessful collaboration can lead to performance decrements worse than if the operator was acting solely and without the use of an autonomous aid (Endsley & Kaber, 1999; Muthard & Wickens, 2003).

Table 6 presents the studies reporting research on the topic of LOA. One third of the studies utilized a version of Endsley and Kaber's (1999) 10-level LOA taxonomy. This representation separates tasks into four roles: monitoring, generating, selecting, and implementing. Each of the 10 levels in the taxonomy assigns either a human operator, computer (autonomous agent) or both to control each role. Across this work it is clear that some amount of automation does increase overall performance for primary tasks. This is true for novice robot operators (e.g., Hughes & Lewis, 2005), UGV and UAV operators (e.g., Wang & Lewis, 2007), as well as performance on targeting simulations (Kaber & Endsley, 2003). In certain conditions, however, automation can lead to significant problems, especially if the operator is unable to access raw data (Rovira, McGarry, & Parasuraman, 2007) or does not know how to regain control of a robot (Krotkov, Simmons, Cozman, & Koenig, 1996). In essence, it is important the operator be able to disengage and override the automation, taking it out of the loop. Once again this is consistent with the broader socio-technical perspective (Beer, 1966; Wier, 2004) that no system can operate at full performance, and at some point errors are likely.

The main guiding principle for LOA is to allow the human operator to generate or select potential actions and have the action subsequently implemented by the system (e.g., Kaber & Endsley, 2003). In other words, human cognition should remain part of the work process, but automation can reduce responding demands by executing tasks for the operator. This is consistent with work reported in the area of expert systems (Coovert, Ramakrishna, & Salas, 1989), whereby users preferred those that kept the user central in the decision-action chain. So the outcomes appear clear; an increase in task or process automation reduces subjective workload and situation awareness of the operator (Kaber et al., 2000). It seems sensible that operators should use all available technology for their task. A review of the literature, however, does not fully support this belief. Although modest levels of automation may be helpful, automation cannot replace the operator in the overall work process. This is especially true given that automation is an imperfect decision-making system, discussed next.

### 5.2. Automation reliability

While research on LOA tends to focus on system level automation, automation does not always occur in every aspect of a given task. Much research exists exploring the use of automated aids

**Table 6**
Summary of studies examining level of autonomy (LOA).

| Study | Manipulation (IV) and automation design | Criteria (by task type) | Results |
|---|---|---|---|
| Bruemmer et al. (2004) | Manual robot control vs. shared control with robot navigating and operator focused on targets | Efficiency and errors in targeting | – For novice robot operators, performance was increased with the use of a semi-autonomous (shared control) navigation aid |
| Chen and Joyner (2009) | Manual UGV control vs. semi-autonomy (monitor UGV actions) | Targeting errors | – Users performed gunnery tasks in addition to teleoperation<br>– Manual control improved robot task performance over semi-autonomy, but at the expense of gunnery task performance |
| Endsley and Kaber (1999) | Ten LOAs in monitoring, generating, selecting, and implementing between human operator and automated system | Efficiency and errors in decision-making | – LOAs which combine human generation of options and automated implementation produced superior results<br>-Joint decision making (human/system collaboration) was detrimental to performance |
| Hardin and Goodrich (2009) | Search and rescue mission with with varying levels of autonomy: adaptive, adjustable, or mixed initiative | Efficiency and workload | – Mixed initiative (MI), where operator and UGVs jointly decide on LOA for situation performed better than operator in complete control (adjustable) and complete UGV control (adaptive) |
| Hughes and Lewis (2005) | User-controlled vs. sensor-driven control of secondary independent UGV camera | Efficiency in searching and targeting | – Sensor-driven control was better;<br>– Automatic gaze control of a UGV camera helped in object ID |
| Kaber and Endsley (2003) | 5 LOAs and 5 schedules of automation (automation on, then off for a specified time) | Errors, workload, and SA in system control task (decision-making and targeting) | – When automation was cycled on and off, performance was best when the human operator implemented a corresponding strategy<br>– Workload correlated with secondary task performance |
| Kaber et al. (2000) | 5 LOAs range from simple support to full automation | Errors, efficiency, workload, and SA for systems control and decision-making | – Increased automation led to performance improvements and reduces subjective workload, but also reduced SA for some system functions |
| Krotkov et al. (1996) | None, veto-only (e.g., to avoid damage), or semi-autonomous aid (adjusts course) | Usability in UGV navigation | – Users struggled to adapt strategies around autonomous agent control and steering/navigation trouble may arise if the operator is unable to adjust |
| Luck et al. (2006) | 3 LOAs: manual control, veto-only, and autonomous waypoint navigation | Errors, efficiency, and usability for UGV search and rescue | – Increased automation led to performance improvements in both errors and time as well as a buffer from the negative effects of control latency |
| Schermerhorn and Schultz (2009) | Exploration/search task with autonomous or non-autonomous robot | Efficiency and satisfaction | – When using autonomous robot participants were more accurate, but not faster<br>– Participants seemed to ignore "disobedience" and preferred working with the autonomous vs. normal robot |
| Wang and Lewis (2007) | 3 levels of LOA for team of 3 UGVs: full autonomy, mixed control, full control | Efficiency and usability for UGV search and rescue | – With multiple UGVs, mixed control paradigm (manual control and cooperative automation) provided best performance<br>– Switching attention between robots more frequently performed better in manual and mixed control scenarios |
| Wickens et al. (2003) | Single or dual UAV control with no aid, auditory aid, or flight path tracking automation | Errors and RT in tracking, targeting, and system monitoring | – Automation aid helped improve target identification task more when operating multiple UAVs versus single UAV control |

and decision-making support systems that augment and assist a human operator controlled task.

Automation aids typically are used to alert a human to important information that is either necessary for task completion or helpful in completing a task more efficiently or effectively. Some aids simply present the user with raw information in a more salient form, such as an auditory warning (Wickens et al., 2003). Other automated aids are more sophisticated and aggregate different sources of information to make a recommendation or alert to the user by way of complex computer algorithms (Wickens, Dixon, Goh, & Hammer, 2005). Existing research in this area falls in one of three general design categories: production systems, targeting tasks, and diagnostics monitoring.

More complex aids aggregate raw data and present recommendation or alerts to operators in an aggregated or fused format. For these types of aids, imperfect calculations can lead to misleading information or incorrect decisions. These automation imperfections can take the form of either false-alarms or misses (Dixon & Wickens, 2006). While these imperfections can be attributed to a myriad of causes (e.g., low quality video feed, raw data inaccuracy), they are commonly associated with thresholds set in the decision-making computer algorithms that calculate the raw data and produce the alerts and cues. In many cases, these thresholds can be adjusted to make an automated aid more or less prone to false-alarms or misses (Levinthal & Wickens, 2006; Yeh & Wickens, 2001).

Table 7 presents the summarized information for studies examining automation reliability. Across all studies, reliability and accuracy of automated aids has a significant effect on performance. Automation with a high tendency for false alarms results in the greatest detriment to performance. When operators are given automated aids with a high level of false alarms, they rely upon and take actions in response to the devices recommendation less frequently and are more likely to ignore raw data in targeting tasks (Dixon & Wickens, 2006). In a scenario where operators were required to make a response to imperfect automated diagnostic aids, responses were slower to all automation aids if false alarms were common. Raw data became relied upon more frequently, reducing the overall efficiency provided by the automated aid (Wickens et al., 2005). If an operator is working in imperfect automation conditions, complacency leads to further decreases in performance (Rovira et al., 2007). In nearly all cases, when workload is increased, the overall detrimental effects of imperfect automation are polarized (e.g., Levinthal & Wickens, 2006).

Imperfect automation aids also influence performance through the reallocation of attention. This can occur in several ways, the simplest being when an incorrectly activated alert or cued target is attended to by an operator while an actual target or event goes unnoticed (e.g., Yeh & Wickens, 2001). Additionally, automation can lead operators to ignore raw data for a portion of a task that has become automated (Muthard, 2003), essentially assuring a problematic situation will arise should automation fail. A couple of studies also suggest that it is useful to provide accurate information of automation reliability to the operator, particularly when automation is unreliable (Cassidy, 2009; Wang, Jamieson, & Hollands, 2009). Operators that are aware of potential automation

**Table 7**
Summary of studies examining automated aid reliability.

| Study | Manipulation (IV) and reliability design | Criteria (by task type) | Results |
|---|---|---|---|
| Cassidy (2009) | Different information (3 groups) about automated aid reliability for target identification: none, accurate, inaccurate | Trust and reliance on automation, and mental model accuracy | – Participants who received no information about reliability relied more on the automation aid than those who were given correct and incorrect information about the aid's reliability |
| Chen (2009) | Targeting aids with imperfect reliability (false-alarm or miss-prone); spatial ability and attentional control | Errors and workload for communication and gunnery tasks | – More automation led to higher performance and reduced workload<br>– High attentional control led to false-alarm-prone alerts being more detrimental; low attentional control participants did worse with miss-prone automation |
| Dixon and Wickens (2006) | Automated alerts were100% reliable, 67% with false alarms, and 67% with misses | Errors, RT, and SA in UAV targeting and system monitoring | – False-alarm prone automation decreased the use of aids encouraged operators to ignore raw data<br>– Imperfect automation lea to better detection of a target miss |
| Goodrich, McLain, Anderson, Sun, and Crandall (2007) | Manual robot teleoperation vs. semi-autonomous navigation via waypoints with or without failure warning | Reaction time | – Autonomy results in less idle time to recognize problems, but without automation aid, this benefit turns into a major obstacle<br>– Automation led to dependence when engaged in secondary tasks |
| Kaber et al. (2000) | Normal operation vs. unexpected automation failure | Errors, efficiency, workload, and SA for systems control and decision-making | – In automation failure, lower level LOAs with more human control resulted in the best performance due to increased SA |
| Levinthal and Wickens, (2006) | No automation, 90% reliable, 60% reliable but prone to false alarms, or 60% reliable but prone to true misses | Efficiency in UAV navigation, RT to system alerts | – Aids prone to false alarms were inhibited performance more than 90% reliable or 60% reliable aids prone to misses |
| Meyer, Feinshreiber, and Parmet (2003) | Automated cuing agent for: 45% vs. 80% reliable; High vs. low overall automation | Errors in quality control decision-making task | – Higher levels of automation resulted in more reliance on cues – No performance differences between LOA conditions for valid cues, but low LOA outperformed high LOA for unreliable cues |
| Muthard (2003) | Flight simulation with or without reliable automation (for route selection only) | Errors, efficiency, and confidence in route selection and implementation | – When flight plan selection was automated, pilots were more likely to ignore environmental changes that made flight unsafe<br>– Automation was best in selection, but not inv implementation |
| Rovira et al. (2007) | 60% vs. 80% decision reliability in automation aid | Errors, RT, workload, and trust on command and control decision-making task | – Imperfect decision-making automation was detrimental to performance, explained by operator complacency with automation and lack of access to raw data |
| Ruff et al. (2002) | 95% or 100% accurate automated or by-consent decision-making aid | Errors and workload for UAV targeting and decisions | – Management-by-consent automation aid resulted in best performance as it left operators in the loop but was scalable to increases in workload (more UAVs) |
| Wang, Jamieson, and Hollands (2009) | Target identification task with no aid, 67% reliable aid, or 80% reliable aid which was either disclosed to participants or not | Trust and reliance on automation, errors | – 80% reliable aid improved performance compared to 67% reliable and no aid<br>– Trust mediated relationship between belief and reliance on feedback, thus disclosing reliability information led to more appropriate reliance on aids |
| Wickens et al. (2005) | Automated diagnostics information: none, 100% accurate, 60% reliable w/false-alarms, 60% reliable w/misses | Errors and efficiency for UAV navigation, targeting, systems monitoring | – Automation prone to misses decreased concurrent task performance, whereas automation prone to false alarms led to slower RT to all auto-alerts and decreased efficiency, accuracy |
| Wickens, Rice, Keller, Hutchins, Hughes, and Clayton (2009) | Air-traffic controller data on conflict alerts and controller behavior | Responses to alerts, reaction time, reliance on alerts | – False alarms were related to more non-responses, but not to true alerts, and no RT delay was found (no "cry wolf" effect)<br>– Anticipatory behavior before alerts was common, and reliance on alerting system increased with hard to visualize conflicts |
| Yeh and Wickens (2001) | 75% vs. 100% reliable cuing for some targets | Errors, workload, and trust on UAV targeting | – Partially reliable cuing increases false alarms and eliminates overall performance benefits of cuing; Cuing draws attention towards cued target results in other targets being overlooked |

failures should be more likely to recognize and correct for automated errors when they occur. In line with the findings of LOA research, guiding principles for reliability research suggest giving operators access to raw data, avoiding situations where operators are out of the loop, and fully brief operators on the reliability as well as the LOA for an autonomous system.

So we have a bit of a catch-22 concerning automation. As work becomes more complex and demands excessive on our information processing system, it is imperative workload be managed and, if possible, decreased. One way to accomplish this is through automating certain tasks, thereby lessening workload. This works well as long as the automation has nearly perfect reliability; if it does not, workload may increase. But as socio-technical systems teach us, no technology will ever have perfect reliability. So we must decide at what level does system reliability become high enough, acknowledging that the risk of errors will never be eliminated.

### 5.3. Unresolved issues in automation

Although research on LOA and automated aid reliability has covered many important issues surrounding the interaction of humans and autonomous systems and agents, there is room for more investigation. An area that has been largely overlooked in current streams of research is difference in the experience levels of operators and how that impacts performance. Whether they are UAV pilots or quality control supervisors, current research has largely ignored the fact that experience may play a large role in the interactions operators have with automation. Some research has focused on novice operators (Bruemmer, Boring, Few, Marble, & Walton, 2004), but empirical investigations comparing novices to experienced operators is needed. For example, a novice operator will likely respond poorly to an automation failure when compared to an experienced employee who knows the background processes behind the automation. Research on problem detection emphasizes the importance of expertise in identifying and interpreting cues (Klein, Pliske, Crandall, & Woods, 1999). Thus, we expect expert operators to have better performances with unreliable automation than novice operators. However, the relationship between expertise and performance should be weaker with reliable automation, particularly if the automation can compensate for any deficiencies in a novice operator.

*Proposition 7*: Operator expertise interacts with automation reliability to affect performance, such that expertise is more valuable to performance in conditions of imperfect automation.

Keeping operators "in the loop" with the task they are completing is another important determinant of performance. Research on interface design could greatly inform this issue by investigating displays that aggregate data and present automation aids, but also provide intuitive access to raw data should operators need it. An existing problem with operators who do have access to raw data is the additional workload associated with accessing it. If the information was easily available and intuitively connected to the related automation within an interface, these problems may be resolved. In addition to the numerous design variables discussed previously, automation interfaces should also focus on intuitive and easy use. The extant HRI literature has documented a wide variety of systems that seek to make information easier to understand, and interfaces easy to use. Some additional consideration should be given to the work of Vicente (1999) and others, who codify naturalistic displays for everyday use. We expect that performance comparisons would favor naturalistic displays over traditional displays of automated systems.

*Proposition 8*: Naturalistic data displays improve operator performance with automated systems over traditional displays by reducing the workload associated with accessing raw data.

Lastly, as technology allows, adaptive automation schemes should be investigated as a potential buffer to the effects of different operators or tasks. A critical issue is the degree that a system may adjust its own autonomy, or self-adjustment. Adjustable autonomy in systems could assist operators by altering their own actions based on output performance or operator responses to automation aids, as described in a positive feedback loop. For example, in a semi-autonomous quality control system, performance data could be fed back into the system to subsequently alter the LOA. If a given operator is experienced and performs better with more control of the system, he or she could then be granted more control. On the other hand, a novice operator might benefit from either higher levels of automation when output efficiency is important or from low levels of automation for training purposes. Similarly, complex tasks may demand differing levels of automation to compensate for response difficulties or personal danger. Using performance-related information from previous trials, autonomous systems or agents might be able to predict failures and correct for workers before the human operator is even aware of a problem.

*Proposition 9*: Adjustable automated systems that are partially guided by past or current operator performance should improve performance beyond other scripts for automated behavior.

## 6. Conclusion

Within the past 15 years, the HRI literature has grown significantly, including research on the problem of operator workload. Although such research has addressed the issue of HRI workload substantively, there remain several issues within the HRI literature. Overall, there are many variables for robot systems – display design, automated functions and intelligence frameworks, interface design, etc. We have covered several promising avenues for research for the topics reviews here. However, there are other gaps in research which apply to concepts beyond a single display feature or automation level.

### 6.1. Future research

Much of the extant research on systems attempt to optimize system performance, but it is unclear from our review what the empirical benefits are from system compared to another. For example, there are numerous derivations of artificial intelligence frameworks for autonomous behavior, but it is unclear what advantage one framework may have upon another. Many evaluations have been performed upon a single system, or several iterations of a given system, but validated systems have rarely been compared against one another using task performance criteria. Research should thus include more empirical comparisons between multiple systems with differing combinations of features, the results of which could inform the incremental validity of one system over another system. This line of research could also identify specific features which provide a practical advantage in HRI tasks and help integrate existing systems that serve similar functions and provide the same performance benefits.

Another concern revealed from the current review is a need for more consistency in variable definitions and measurement. For example, latency/time delay and camera perspective manipulations utilized a wide range of terminology and operations, such that identifying guiding principles for these variables was difficult. In addition to independent variables, task criteria also vary widely between studies. Within the area of FOV research, studying SA appears to be a fruitful direction, but these criteria are neglected in other types of visual demand manipulations.

For some variables, the same criterion label described different measurements. For example, error rate is reflected in numerous ways including: points acquired, targets identified, and collisions avoided. Although these data inform us about the task-specific relationships they examine independently, it is difficult to sensibly integrate them underneath a common criterion due to the task-dependency issue. This discovery brings to light the fact that more general investigations are needed which can be flexibly applied to more tasks (Miller & Parasuraman, 2003) and common measurement methods so the findings can be better utilized by a wider audience. The majority of coded studies also shared methodological constraints due to their samples, which were notably small, predominantly male, and often recruited participants in advanced education. Thus, HRI studies would also benefit from larger and more diverse samples.

The definition, operation, and measurement of study variables warrant greater attention in order to create a more unified research agenda. Once studies attend to these issues, an empirical review may be conducted in the form of a meta-analysis. Such an endeavor could quantify the relationships in this review, which used a qualitative approach. Thus, researchers and practitioners would have more precise data to inform decisions related to HRI socio-technical systems.

### 6.2. Summary

The purpose of our work was to systematically review the empirical research on workload in HRI, to draw guiding principles for managing workload, and create propositions to guide future HRI research. When appropriate, we tempered these findings by considering them within the larger perspective provided by socio-technical systems. A variety of factors in the socio-technical system may negatively impact workload, but these issues may also be addressed through careful consideration of the task demands and available system resources. In cases of high workload, optimal visual displays, multimodal feedback, and reliable automation can improve operator performance. However, we caution that the task and criterion must be considered, and that some conclusions are drawn from a relatively small sample of studies.

It is also important to consider work reported here (and the topic of HRI) within the larger socio-technical perspective. Robots are entities within a larger system of humans and organizations. Workers within those systems do not perform optimally as they are influenced by daily motivation, power, and other needs. Regardless of how well designed, systems do not perform optimally. This is in part due to the fact that failures of equipment occur. Thus, it is imperative that operators and artificial agents work together as team members monitoring one another's actions and performance. Furthermore, organizational resources are needed to provide clear task mission and the necessary equipment to perform the task. Without effective leadership and material resources, operators and autonomous agents will struggle to be effective.

Furthermore, not all events can be foreseen. Since events change systems and their states, neither can the future states be deterministically specified, nor can their interactions be foretold. This implies that the development of automated systems cannot be relied upon to always correctly cue the operator or to take other appropriate action. Perhaps the inability to understand and specify the system is due to its opacity (Revans, 1982). If so, those working in certain sections of the organization and with specific technologies (e.g., HRI, nuclear power, aviation) will always be under the veil of uncertainty and some unreliability. As such, we must carefully consider the tenants of socio-technical systems and construct our technologies and organizational systems with those in mind.

## Acknowledgements

This work was supported by government contract number DAAD19-01-C-0065, task order 83. The views, opinions, and/or findings contained in this report are those of the authors and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other documentation.

## References

*Note: *indicates a study included in review, see Tables 1–7.*


*Adams, J. J. (2009). Multiple robot/single human interaction: Effects on perceived workload. *Behavior Information and Technology, 28*(2), 183–198.

*Adelstein, B. D., Thomas, G. L., & Ellis, S. R. (2003). Head tracking latency in virtual environments: Psychophysics and a model. In *Proceedings of human factors and ergonomics society 47th annual meeting* (pp. 2083–2087).

*Allison, R. S., Zacher, J. E., Wang, D., & Shu, J. (2004). Effects of network delay on a collaborative motor task with telehaptic and televisual feedback. In *Proceedings of 2004 ACM SIGGRAPH international conference on virtual reality continuum and its applications in industry* (pp. 375–381).

Beer, S. (1966). *Decision and control: The meaning of operational research and management cybernetics*. Chichester: John Wiley.

*Bruemmer, D. J., Boring, R. L., Few, D. A., Marble, J. L., & Walton, M. C. (2004). "I call shotgun!": An evaluation of mixed-initiative control for novice users of a search and rescue robot. *IEEE Transactions on Systems, Man, and Cybernetics, 34*, 2847–2852.

Burke, J. L., Prewett, M. S., Gray, A., Yang, L., Stilson, F. R. B., Redden, E., et al. (2006). Comparing the effects of visual-auditory and visual-tactile feedback on user performance. A meta-analysis. *Proceedings of the Eighth International Conference on Multimodal Interfaces, 108*, 117.

*Calhoun, G., Draper, M., Nelson, J., Lefebvre, A., & Ruff, H. (2006). *Simulation assessment of synthetic vision concepts for UAV operations*. (Technical Report). Air Force Research Laboratory, Human Effectiveness Directorate Warfighter Interface Division at Wright-Patterson AFB, OH.

*Cassidy, A. M. (2009). Mental models, trust, and reliance: Exploring the effect of human perceptions on automation use (Master's Thesis, Naval Post-Graduate School).

*Chadwick, R. A. (2006). Operating multiple semi-autonomous robots: Monitoring, responding, detecting. In *Proceedings of the human factors and ergonomics society's 50th annual meeting.* San Francisco, CA. *HFES '06* (pp. 329–333).

*Chadwick, R. A. (2005). The impacts of multiple robots and display views: An urban search and rescue simulation. In *Proceedings of the human factors and ergonomics society's 49th annual meeting.* Orlando, FL. *HFES '05* (pp. 387–391).

*Chen, J. Y. C. (2009). Concurrent performance of military tasks and robotics tasks: Effects of automation unreliability and individual differences. In *Proceedings of the fourth human robot interaction conference 2009*, La Jolla, CA. *HRI '09* (pp. 181–188).

*Chen, J. Y. C., & Joyner, C. (2009). Concurrent performance of gunner's and robotics operator's tasks in a multi-tasking environment. *Military Psychology, 21*, 98–113.

Chen, J. Y. C., & Thropp, J. E. (2007). Review of low frame rate effects on human performance. *IEEE Transactions on Systems, Man and Cybernetics-Part A: Systems and Humans, 37*(6), 1063–1076.

*Chen, J. Y. C., Durlach, P. J., Sloan, J. A., & Bowers, L. D. (2008). Human–robot interaction in the context of simulated route reconnaissance missions. *Military Psychology, 20*, 135–149.

Chen, J. Y. C., Haas, E., & Barnes, M. (2007). Human performance issues and user interface design for teleoperated robots. *IEEE Transactions on Systems, Man, and Cybernetics, 37*(6), 1231–1245.

Coovert, M. D., & Elliott, L. R. (2009). UAV Operator Specifications from O*NET. In *Proceedings of the 15th international symposium on aviation psychology*. Dayton, OH (pp. 552–557).

Coovert, M. D., Ramakrishna, K., & Salas, E. (1989). Preferences for power in expert systems by novice users. *AI & Society, 3*, 59–61.

Coovert, M. D., Walvoord, A., Elliott, L., & Redden, E. (2008). A tool for the accumulation and evaluation of multimodal research. *IEEE Transactions on Systems, Man, and Cybernetics, 24*(5), 1884–1906.

*Cosenzo, K. A., Parasuraman, R., Novak, A., & Barnes, M. (2006). *Implementation of automation for control of robotic systems.* (Technical Report). Army Research Laboratory (ARL) at Aberdeen Proving Grounds, MD.

*Crandall, J. W., & Cummings, M. L. (2007). Developing performance metrics for the supervisory control of multiple robots. In *Proceedings of the second ACM/IEEE international conference on human–robot interaction.* Arlington, VA. *HRI '07* (pp. 33–40).

*Darken, R. P., & Cervik, H. (1999). Map usage in virtual environments: Orientation issues. *Proceedings of IEEE Conference on Virtual Reality, 133*, 140.

*Darken, R. P., Kempster, K., & Peterson, B. (2003). Effects of streaming video quality of service on spatial comprehension in a reconnaissance task. In *Proceedings of IAITSEC 2001*, Orlando, FL.

*Dixon, S. R., & Wickens, C. D. (2006). Automation reliability in unmanned aerial vehicle control: A reliance-compliance model of automation dependence in high workload. *Human Factors, 48*, 474–486.

*Dixon, S. R., & Wickens, C. D. (2003). Control of multiple-UAVs: A workload analysis. In *Proceedings of the 12th international symposium on aviation psychology.* Dayton, OH (pp. 1–5).

*Draper, M., Calhoun, G., & Nelson, J (2006). *Evaluation of synthetic vision overlay concepts for UAV sensor operations: Landmark cues and picture-in-picture*. (Technical Report). Human Effectiveness Directorate, Warfighter Interface Division at Wright-Patterson AFB, OH.

*Draper, M., Calhoun, G., Ruff, H., Williamson, D., & Barry, T. (2003). Manual versus speech input for unmanned aerial vehicle control stations. In *Proceedings of the human factors and ergonomics society's 47th annual meeting.* Denver, CO. *HFES '03* (pp. 387–391).

*Draper, J. V., Handel, S., Hood, C. C., & Kring, C. T. (1991). Three experiments with stereoscopic television: when it works and why. In *IEEE transactions on systems, man and cybernetics* (pp. 1047–1052).

*Drascic, D., & Grodski, J. (1993). Defense teleoperation and stereoscopic video. *Proceedings of SPIE: Stereoscopic Displays and Applications IV, 1915*, 1–12.

*Drury, J. L., Keyes, B., & Yanco, H. A. (2007). LASSOing HRI: Analyzing situation awareness in map-centric and video-centric interfaces. In *Proceedings of the second ACM/IEEE international conference on human–robot interaction.* Arlington, VA. *HRI '07* (pp. 279–286).

Elliott, L. R., Prewett, M. S., Coovert, M. D., Walvoord, A. G., Saboe, K. N., & Johnson, R. J. (2009). *A meta-analysis of vibrotactile and visual information displays for improving task performance.* Unpublished manuscript, Army Research Laboratory at Fort Benning, GA.

*Ellis, S. R., Mania, K., Adelstein, B. D., & Hill, M. (2004). Generalizeability of latency detection in a variety of virtual environments. In *Proceedings of the human factors and ergonomics society's 48th annual meeting*. New Orleans, LA, *HFES '04* (p. 2632).

*Endsley, M. R., & Kaber, D. B. (1999). Level of automation effects on performance, situation awareness and workload in a dynamic control task. *Ergonomics, 42*, 462–492.

*Fisher, A., McDermott, P. L., & Fagan, S. (2009). Bandwidth allocation in a military teleoperation task. In *Proceedings of the human factors and ergonomics society's 53rd annual meeting.* La Jolla, CA. *HFES '09* (pp. 287–288).

*Folds, D. J., & Gerth, J. (1994). Auditory monitoring of up to eight simultaneous sources. In *Proceedings of the human factors and ergonomics society's 38th annual meeting.* Santa Monica, CA, *HFES '94* (pp. 505–509).

*Galster, S. M., Knott, B. A., & Brown, R. D. (2006). Managing multiple UAVs: Are we asking the right questions? In *Proceedings of the human factors and ergonomics society's 50th annual meeting.* San Francisco, CA. *HFES '06* (pp. 545–549).

*Goodrich, M. A., McLain, T. W., Anderson, J. D., Sun, J., & Crandall, J. W. (2007). Managing autonomy in robot teams: Observations from four experiments. In *Proceedings of the second human robot interaction conference.* Arlington, VA. *HRI '07* (pp. 25–32).

* Hardin, B., & Goodrich, M. (2009). On using mixed-initiative control: A perspective for managing large-scale robot teams. In *Proceedings of the fourth ACM/IEEE international conference on human–robot interaction.* La Jolla, CA. *HRI '09* (pp. 165–172).

* Heath-Pastore, T. (1994). *Improved operator awareness of teleoperated land vehicle attitude*. (Technical Report). Naval Command, Control and Ocean Surveillance Center, CA.

* Hendy, K., Lao, J., & Milgram, P. (1997). Combining time and intensity effects in assessing operator information-processing load. *Human Factors, 30*, 30–47.

* Hill, S. G., & Bodt, B. (2007). A field experiment of autonomous mobility: Operator workload for one or two robots. In *Proceedings of the second ACM/IEEE international conference on human–robot interaction.* Arlington, VA. *HRI '07* (pp. 169–176).

* Hughes, S., & Lewis, M. (2005). Task-driven camera operations for robotic exploration. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans, 35*(4), 513–522.

* Humphrey, C. M., Henck, C., Sewell, G., Williamson, B. W., & Adams, J. A. (2007). Assessing the scalability of a multi-robot interface. In *Proceedings of the second ACM/IEEE international conference on human–robot interaction.* Arlington, VA. *HRI '07* (pp. 239–246).

* Kaber, D. B., & Endsley, M. R. (2003). The effects of level of automation and adaptive automation on human performance, situation awareness and workload in a dynamic control task. *Theoretical Issues in Ergonomics Science, preview*, 1–40.

* Kaber, D. B., Onal, E., & Endsley, M. (2000). Design of automation for telerobots and the effect on performance, operator situation awareness and subjective workload. *Human Factors and Ergonomics in Manufacturing, 10*, 409–430.

Klein, G., Pliske, R. M., Crandall, B., & Woods, D. (1999). Features of problem detection. In *Proceedings of the human factors and ergonomics society's 43rd annual meeting.* Houston, TX. *HFES '99* (pp. 133–137).

* Krotkov, E., Simmons, R., Cozman, F., & Koenig, S. (1996). Safeguarded teleoperation for lunar rovers: From human factors to field trials. In *Proceedings of the 26th international conference on environmental systems*. Santa Monica, CA.

* Lane, J. C., Carignan, C. R., Sullivan, B. R., Akin, D. L., Hunt, T., & Cohen, R. (2002). Effects of time delay on telerobotic control of neutral buoyancy vehicles. In *Proceedings of IEEE international conference on robotics and automation*. Washington, DC (pp. 2874–2879).

* Levinthal, B. R., & Wickens, C. D. (2006). Management of multiple UAVs with imperfect automation. In *Proceedings of the human factors and ergonomics society's 50th annual meeting.* San Francisco, CA. *HFES '06* (pp. 1941–1944).

* Lewis, M., Wang, J., Hughes, S., & Liu, X. (2003). Experiments with attitude: Attitude displays for teleoperation. *IEEE Transactions on Systems Man and Cybernetics, 33*(2), 1345–1349.

* Lif, P., Jander, H., & Borgwall, J. (2007). Tactical evaluation of unmanned ground vehicle during a MOUT exercise. In *Proceedings of the 7th international conference of engineering psychology and cognitive ergonomics*. Beijing, China. *HCI '07* (pp. 731–740).

* Lion, D. M. (1993). *Three dimensional manual tracking using a head-tracked stereoscopic display*. (Technical Report). Human Interface Technology Lab, WA.

Locke, E. A., & Latham, G. P. (1990). *A theory of goal setting and task performance*. Englewood Cliffs, NJ: Prentice-Hall.

* Luck, J. P., McDermott, P. L., Allender, L., & Russell, D. C. (2006, March). An investigation of real world control of robotic assets under communication latency. In *Proceedings of the first human robot interaction conference 2006*, Salt Lake City, UT. *HRI '06* (pp. 202–209).

* Massimino, M. J., & Sheridan, T. B. (1994). Teleoperator performance with varying force and visual feedback. *Human Factors, 36*(1), 145–157.

* Meyer, J., Feinshreiber, L., & Parmet, Y. (2003). Levels of automation in a simulated failure detection task. *IEEE Transactions on Systems, Man and Cybernetics, 33*(3), 2101–2106.

Miller, C.A., Parasuraman, R. (2003). Beyond levels of automation: An architecture for more flexible human-automation collaboration. In *Proceedings of the 47th human factors and ergonomics society* (pp. 82–85).

* Mosier, K. L., Sethi, N., McCauley, S., Khoo, L., & Orasanu, J. M. (2007). What you don't know can hurt you: Factors impacting diagnosis in the automated cockpit. *Human Factors, 49*(2), 300–310.

* Murray, S. A. (1995). Human–machine interaction with multiple autonomous sensors. In *Proceedings of the 6th IFAC/IFIP/IFORS/IEA symposium on analysis, design and evaluation of man–machine systems*.

* Muthard, E. K. & Wickens, C. D. (2003). Factors that mediate flight plan monitoring and errors in plan revision: Planning under automated and high workload conditions. In *Proceedings of the 12th international symposium on aviation psychology*, Dayton, OH.

* Nielson, C. W., & Goodrich, M. A. (2006). Comparing the usefulness of video and map information in navigation tasks. In *Proceedings of the first human robot interaction conference*. Salt Lake City, UT. *HRI '06* (pp. 95–101).

* Olmos, O., Wickens, C. D., & Chudy, A. (2000). Tactical displays for combat awareness: An examination of dimensionality and frame of reference concepts and the application of cognitive engineering. *The International Journal of Aviation Psychology, 10*(3), 247–271.

* Parasuraman, R., Galster, S., & Miller, C. (2003). Human control of multiple robots in the RoboFlag simulation environment. *IEEE Transactions on Systems, Man and Cybernetics, 33*(4), 3233–3237.

* Parasuraman, R., Gaslter, S. M., Squire, P., Furukawa, H., & Miller, C. (2005). Flexible delegation-type interface enhances system performance in human supervision of multiple robots: Empirical studies with RoboFlag. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans, 35*(4), 481–493.

* Park, S. H., & Woldstad, J. C. (2000). Multiple two-dimensional displays as an alternative to three-dimensional displays in telerobotic tasks. *Human Factors, 42*(4), 592–603.

* Pazuchanics, S. L. (2006). The effects of camera perspective and field of view on performance in teleoperated navigation. In *Proceedings of the human factors and ergonomics society's 50th annual meeting*. San Francisco, CA. *HFES '06* (pp. 1528–1532).

Prewett, M. S., Gray, A., Burke, J. L., Yang, L., Stilson, F. R. B., Elliott, L. R., et al. (2006). The benefits of multimodal information: A meta-analysis comparing visual and visual-tactile feedback. *Proceedings of the Eighth International Conference on Multimodal Interfaces*, 333–338.

* Reddy, M. (1997). *The effects of low frame rate on a measure of user performance in virtual environments*. (Technical Report). Dept of Computer Science, University of Edinburgh.

Revans, R. (1982). *The origins and growth of action learning.* Bromley: Chartwell Bratt.

* Richard, P., Birebent, G., Coiffet, P., Burdea, G., Gomez, D., & Lagrana, N. (1996). Effect of frame rate and force feedback on virtual object manipulation. *Presence, 5*(1), 95–108.

* Rovira, E., McGarry, K., & Parasuraman, R. (2007). Effects of imperfect automation on decision making in a simulated command and control task. *Human Factors, 49*, 76–87.

* Ruff, H. A., Narayan, S., & Draper, M. H. (2002). Human interaction with levels of automation and decision-aid fidelity in the supervisory control of multiple simulated unmanned air vehicles. *Presence. Teleoperators and Virtual Environments, 11*(4), 335–351.

* Ruff, H. A., Calhoun, G., Draper, M., Fontejon, J., & Abbott, J. (2004). Exploring automation issues in supervisory control of multiple UAVs. In *Proceedings of the 5th conference of human performance, situation awareness, and automation technology* (pp. 218–222).

* Schermerhorn, P., & Schultz, M. (2009). Dynamic robot autonomy: Investigating the effects of robot decision-making in a human-robot team task. In *Proceedings of the 2009 international conference on multimodal interfaces*. Cambridge, MA. *ICMI-MLMI* '09 (pp. 63–70).

* Schipani, S. P. (2003). *An evaluation of operator workload during partially autonomous vehicle operations.* (Technical Report). Army Research Laboratory at Aberdeen Proving Grounds, MD.

* Scribner, D. R., & Gombash, J. W. (1998). *The effect of stereoscopic and wide field of view conditions on teleoperator performance*. (Technical Report). Army Research Laboratory at Aberdeen Proving Grounds, MD.

* Sellner, B. P., Hiatt, L. M., Simmons, R., & Singh, S. (2006). Attaining situational awareness for sliding autonomy. In *Proceedings of the 1st ACM SIGCHI/SIGART conference on human–robot interaction.* Salt Lake City, UT. *HRI '06*.

* Shreik-Nainar, M. A., Kaber, D. B., & Chow, M.-Y. (2003). Control gain adaptation in virtual reality mediated human–telerobot interaction. *Human Factors and Ergonomics in Manufacturing, 15*(3), 259–274.

* Smyth, C. C. (2002). *Modeling indirect vision driving with fixed flat panel displays: Task performance and mental workload*. (Technical Report). Army Research Laboratory at Aberdeen Proving Grounds, MD.

* Smyth, C. C., Gombash, J. W., & Burcham, P. M. (2001). *Indirect vision driving with fixed flat panel displays for near-unity, wide, and extended fields of camera view.* (Technical Report). Army Research Laboratory at Aberdeen Proving Grounds, MD.

* Squire, P., Trafton, G., & Parasuraman, R. (2006). Human control of multiple unmanned vehicles: effects of interface type on execution and task switching times. In *Proceedings of the 1st ACM SIGCHI/SIGART conference on human–robot interaction*. Salt Lake City, UT. *HRI '06* (pp. 26–32).

* Thomas, L. C., & Wickens, C. D. (2000). *Effects of display frames of reference on spatial judgments and change direction*. (Technical Report). Army Research at Aberdeen Proving Grounds, MD.

* Trouvain, B., Schlick, C., & Mervert, M. (2005). Comparison of a map- vs. camera-based user interface in a multi-robot navigation task. In *Proceedings of IEEE interantional conference on systems, man and cybernetics.* Washington DC, USA. *SMC '03* (pp. 3224–3231).

* Trouvain, B., & Wolf, C. (2003). Design and evaluation of a multi-robot interface. In *The role of humans in intelligent and automated systems. Symposium presented at the 22nd RTO HFM*. Warsaw, Poland, (pp. 1–12).

Vicente, K. J. (1999). *Cognitive work analysis: Toward safe, productive, and healthy computer-based work*. Mahwah, NJ: LEA.

* Wang, H., Lewis, M., Velagapudi, P., Scerri, P., & Sycara, K. (2009a). How search and its subscale tasks in N robots. *Proceedings of the fourth ACM/IEEE international conference on human–robot interaction.* La Jolla, CA. *HRI '09* (pp. 141–147).

* Wang, H., Chien, S.Y., Lewis, M. Velagapudi, P., Scerri, P., & Sycara, K. (2009b). Human teams for large scale multirobot control. *Proceedings of the IEEE international conference on systems, man, & cybernetics.* San Antonio, TX, (pp. 1269–1274).

* Wang, J., Wang, H., & Lewis, M. (2008). Assessing cooperation in human control of heterogeneous robots. In *Proceedings of the third ACM/IEEE international conference on human–robot interaction.* Amsterdam, Netherlands. *HRI '08* (pp. 9–15).

* Wang, J., & Lewis, M. (2007). Human control for cooperating robot teams. In *Proceedings of the second human–robot interaction conference*. Arlington, VA. *HRI '07* (pp. 9–16).

* Wang, L., Jamieson, G. A., & Hollands, J. G. (2009). Trust and reliance on an automated combat identification system. *Human Factors, 51*, 281–291.

*Wang, W., & Milgram, P. (2003). Effects of viewpoint displacement on navigational performance in virtual environments. In *Proceedings of human factors and ergonomics society 47th annual meeting*. Denver, CO. *HFES '03* (pp. 139–143).

*Watson, B., Walker, N., Ribarsky, W., & Spaulding, V. (1998). Effects of variation in system responsiveness on user performance in virtual environments. *Human Factors, 40*(3), 403–414.

*Watson, B., Walker, N., Woytiuk, P., & Ribarsky, W. (2003). Maintaining usability during 3D placement despite delay. *Proceedings of IEEE Conference on Virtual Reality, 133*, 140.

Wickens, C. (2002). Multiple resources and performance prediction. *Theoretical Issues in Ergonomics Science, 3*(2), 159–177.

Wickens, C. (2008). Multiple resources and mental workload. *Human Factors, 50*(3), 449–454.

*Wickens, C. D., Dixon, S. R., & Chang, D. (2003). *Using interference models to predict performance in a multiple-task UAV environment- 2 UAVs.* (Technical Report). University of Illinois at Urbana-Champaign.

*Wickens, C. D., Dixon, S. R., Goh, J., & Hammer, B. (2005). *Pilot dependence on imperfect diagnostic automation in simulated UAV flights: An attentional visual scanning analysis.* (Technical Report). University of Illinois at Urbana-Champaign, Institute of Aviation.

Wier, D. (2004). Sequences of failure in complex socio-technical systems: Some implications of decision and control. *Kybernetes, 33*, 522–537.

*Witmer, B. G., & Kline, P. B. (1998). Judging perceived and traversed distance in virtual environments. *Presence, 7*(2), 144–167.

*Yeh, M., & Wickens, C. D. (2001). Display signaling in augmented reality: Effects of cue reliability and image realism on attention allocation and trust calibration. *Human Factors, 43*(3), 355–365.

*Yi, H., Song, B., Ji, D., & Yu, T. (2006). Experimental research on situation awareness of the operators for unmanned aerial vehicle. *IEEE Transactions on Systems, Man and Cybernetics, 36*(6), 1225–1228.