



# Human Productivity in a Workspace Shared with a Safe Robotic Manipulator

Catharine L. R. McGhan\* and Ella M. Atkins<sup>†</sup>  
University of Michigan, Ann Arbor, Michigan 48109

DOI: 10.2514/1.54993

Robotics technology is approaching a state where highly capable autonomous platforms can directly interact with human companions. This paper explores the ability of a human and manipulator to efficiently complete independent tasks in a shared physical workspace without explicit communication of their planned actions. With focus on quantifying the impact of robot presence on human task performance and workload, a series of human-subject experiments were conducted in which a seated human was asked to perform cognitive and mobility tasks in this shared workspace. The manipulator could unintentionally interrupt these activities, blocking the field of view or physical path to a worksite while completing its own tasks. To avert impending conflict, the robot received and reacted to ideal sensor data identifying the human's tasks upon initiation. Tests over a progressively complex set of tasks for the human-manipulator team were conducted. Results indicate minimal impact on a human companion's productivity when a manipulator is introduced to the workspace, provided it responds quickly to avoid blocking constraints. Future work involves maturation of the robot's perception and prediction capabilities followed by additional shared workspace experiments.

## Nomenclature

$A, B, \dots$	=	types of scenarios
$a_{i-1}$	=	distance from $\hat{Z}_{i-1}$ to $\hat{Z}_i$ measured along $\hat{X}_{i-1}$ , in.
$d_i$	=	distance from $\hat{X}_{i-1}$ to $\hat{X}_i$ measured along $\hat{Z}_i$ , in.
$N$	=	number of moves per second for joint angle waypoint calculations
$R_i$	=	Task Load Index adjusted rating for a single load source $i$
$R_W$	=	Task Load Index weighted rating, or overall workload
$r_i$	=	Task Load Index rating
$T_s$	=	start time, when at initial pose
$T_f$	=	arrival time, when at final pose
$W_i$	=	Task Load Index weight
$\alpha_{i-1}$	=	angle from $\hat{Z}_{i-1}$ to $\hat{Z}_i$ measured about $\hat{X}_{i-1}$ , deg
$\theta_i$	=	(variable) angle from $\hat{X}_{i-1}$ to $\hat{X}_i$ measured about $\hat{Z}_i$ , deg
$\theta_{si}$	=	starting angle of initial pose for joint angle $i$
$\theta_{fi}$	=	resulting angle of final pose for joint angle $i$

## I. Introduction

HUMAN-ROBOT interaction can potentially make hazardous work traditionally done by humans safer and more productive. We define task performance as a function of safety (or risk), efficiency (task completion over time), and user preference (task priority) and productivity as performance when impacted by workload (adverse conditions). In a space environment, risk to an astronaut is cumulative with exposure time at a high risk level, and even 'routine' astronaut extravehicular activities (EVA) have appreciable risk. A highly capable robotic system could accomplish the simpler, well-modeled tasks usually performed by human astronauts on EVA; potentially, this could reduce EVA duration, which reduces risk and may also increase overall productivity. Overall productivity can also increase by adding intelligence to those robotic systems such that close supervision is no longer necessary and tasks can be completed independently. If working locally (e.g., servicing a spacecraft or constructing a lunar habitat), humans and robots may need to operate in a physically proximal environment. The robotic systems' presence must neither introduce unacceptable levels of risk nor interfere in a way that reduces productivity.

For effective collaboration, both human-centric and robot-centric views must be balanced, and, for efficiency, either performance metrics must be common across the team or each agent must act in a manner that maximizes its own reward and does not negatively impact other agent(s). Given a common operating environment, physical actions must be coordinated through sensing and decision making, ideally by both human and robotic agents. In this work, we present a controlled human-robot collaboration environment to study the impact of physically proximal robotic manipulation on human productivity in a shared workspace. We believe that the key to improved performance is prediction of intent by human and robot agents. Intent information could be shared by direct communication, but this can detract from goal-oriented task execution and reduce efficiency. Thus, in this work, we constrain ourselves to implicit communication only, ultimately through sensing and model-based prediction of the collaborator's intent. A test conductor 'senses' the human's activities and directly relays them to the robot, eliminating the need to directly sense and convey to the robot the human's motions over time.

This paper presents results from a series of human-subject experiments aimed at quantifying the impact of a physically proximal robot manipulator on human task performance and workload. To this end, we have developed a laboratory workspace in which a seated human

Received 18 May 2011; revision received 26 June 2012; accepted for publication 27 June 2012; published online 30 January 2014. Copyright © 2012 by Catharine L. R. McGhan and Ella M. Atkins. Published by the American Institute of Aeronautics and Astronautics, Inc., with permission. Copies of this paper may be made for personal or internal use, on condition that the copier pay the \$10.00 per-copy fee to the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923; include the code 2327-3097/14 and \$10.00 in correspondence with the CCC.

\*Ph.D. Candidate, Aerospace Engineering Department. Student Member AIAA.

<sup>†</sup>Associate Professor, Aerospace Engineering Department. Associate Fellow AIAA.

completes cognitive and physical tasks in a workspace also occupied by an active robotic manipulator. Our principal hypothesis is stated as follows.

Productivity of a collaborating human–robot team operating in a shared workspace can be maximized when the human has no need to supervise the robot. Supervision is no longer necessary when a robot can autonomously operate safely and efficiently with acceptable or no impact on its companion’s productivity.

For our work, safety implies that the robotic manipulator will not physically contact the human or that any potential physical contact will not introduce risk of serious harm to the human. The human must also feel safe around the robot. Efficiency implies optimal or near-optimal decision making by all agents. For the robot, this implies the existence of task, environment, and human activity models (e.g., through intent) that allow it to effectively complete its tasks without slowing or compromising the human’s activities. The previous hypothesis statement makes a critical assumption, stated as follows.

We presume human–robot team overall productivity will be higher than the productivity achieved should the human or robot act alone.

This paper seeks to support or disprove our principle hypothesis as well as to analyze the validity of the previous assumption through a series of human–robot experiments. We describe a set of human–robot collaboration experiments designed to enable evaluation of the impact of a robot’s presence and motions on human performance, workload, and focus of attention using a series of objective and subjective metrics. During test operations, a seated human test subject was asked to complete simple cognitive and motor tasks, some with the robot idle and others with it moving to accomplish independent goals. To focus our attention on human–robot productivity rather than robot capabilities and limitations, we provide ‘ideal’ intent data to the robot: a human test conductor presses keys indicating which next task the human test subject appears to be pursuing when that task is initiated. This enables the robot to predict near-term conflicts and react appropriately to complete its own motion-based task(s), either by avoiding physical contact with the human or line-of-sight occlusion of the human’s gaze [1].

Next, we discuss related work and its influence on the set of simplifying assumptions used to appropriately scope our experiments. We also describe the manipulator and computer systems deployed in our tests and then specify the experiments used to assess our hypothesis. Next, the test methodology is explained, followed by our results and conclusions drawn from processing the suite of subjective and objective datasets.

## II. Background

The study of human–robot interaction requires the integration of concepts from computer science (artificial intelligence, multilayer architectures, symbolic decision making, neural networks, machine learning, etc.), physics-based control systems (feedback control, navigation, mechanisms, planning and sensing), human factors or cognitive engineering, and psychology. Most deployed space robotic systems have, to date, been human-centric, with space robotic human-support systems either requiring direct human supervision or having been designed to halt when close to impinging on a human’s work envelope to prevent injury. There are many examples of these systems. AERCam [2,3] is a free-flying six-degree-of-freedom (DOF) spherical camera platform flown on shuttle mission STS-87. AERCam was intended to improve situational awareness for shuttle and extravehicular activity (EVA) missions. The Personal Satellite Assistant (PSA) [4,5], a similar zero-g 6-DOF free-flyer, was developed and tested at NASA Ames. While AERCam was teleoperated, the PSA had sufficient autonomy to station-keep based on fiducial markings in its environment, but it did not sense or react to human presence, unless the human carried fiducial(s) to track. The Mars Exploration Rover (MER) [6] pair, Spirit and Opportunity, are remote teleoperated systems designed for planetary surface exploration without humans in their physical environment. Over their highly successful deployments, MER rovers have seen substantial upgrades to their autonomy software for data compression, navigation, and planning/scheduling. These technologies are transferable to collaborative missions including augmentations in the robotic system’s ability to perceive and react to other agents, human or robotic.

Several large-scale space manipulator systems have defined the state of the art for space-based manipulation: Ranger, Canadarm, Canadarm2, and Dextre [7–9]. The latter two are currently on the International Space Station (ISS) and have recently been joined by the smaller Robonaut 2 platform. The Ranger and Dextre systems have two highly dexterous ‘arms’ designed to complete scripted EVA activities while astronauts remain indoors. The Canadarm is a single multipurpose arm, as is Canadarm2; in addition to manipulation without an astronaut, both Canadarms have been used to interact with an astronaut on EVA by use of their end effector as a work platform upon which astronauts can stand and be maneuvered about, decreasing their physical movement effort. Ranger and the Canadarms are teleoperated systems, while Dextre is a supervised system capable of executing scripted automation sequences. Launched in 2008, Dextre completed its long testing cycle in December 2010 and successfully finished its first official repair job in February 2011 [10,11]; as of June 2012, it has completed two rounds of joint operations with NASA’s Robotic Refueling Mission, demonstrating the feasibility of on-orbit robotic satellite servicing and repair [12,13]. There have also been advances in systems meant to physically collaborate with humans. Notably, NASA’s Robonaut systems [14–16], highly dexterous human-analogue designs, have been extensively tested on Earth for eventual automated space operations. Robonaut was originally meant to replace astronauts but now is intended to complement humans on EVA. The first Robonaut (R1) was initially teleoperated, with incremental implementation of capabilities, allowing it to automatically execute task sequences (such as grabbing objects) while operating near humans with minimal workspace impingement [14]. The Robonaut 2 (R2) system was developed in collaboration with General Motors, and one of two prototypes now resides on the ISS; its mission dictates it undergo a year-long testing cycle inside an isolated chamber to study and validate its operational characteristics in a zero-g interior environment before use [15]. R2 is equipped with force-sensing capabilities that will stop the robot’s motion if it contacts a human (or other object) or shut it down completely if struck with sufficient force; the arm itself is well-padded in case such a hit occurs [17,18]. This capability is an important first step in enabling safe physical human–robot interaction [19,20]. This combined with research into intent prediction and smart planning may contribute to a further increase in the autonomy of such highly capable systems, allowing for direct proximal human–robot collaboration. Our research is complementary to the R2 tests; we focus on human–robot interaction in a ‘safe lab’ but do not attempt to mimic the zero-gravity-constrained ISS environment.

Examples of recent research into tools supporting safe and efficient autonomy include safe real-time trajectories for physically proximal operations [21], safety/injury metrics [19,20], industrial collaboration safety standards [22–24], the efficient/productive distribution of teams of agents and the breakdown and assignment of tasks (both homogeneous and heterogeneous) [25], scheduling problems (centralized and distributed) [26], real-time plan repair [27], and implicit or explicit communication between agents [28,29]. In [21], experiments were conducted where a robot was directed to hand an object to a human. The human’s position was sensed using an automated laser-based positioning system and real-time trajectories were created by running an A\* search on a potential field ‘safety bubble’ determined from a cost function. In our work, we have a higher degree of close-proximity interaction, but we control the environment to allow the use of preplanned trajectories and human entry (rather than sensing) of human position information and task intent. In [19,20], the need for good safety metrics is emphasized, as it is clearly demonstrated that the current widespread use of car-crash safety metrics for these purposes is flawed. An example of this is a lack of warning in a crushing case, where the metrics for acceleration would be low or negative (deceleration), normally indicating a safer situation, while conversely the manipulator could be pinning a person between itself and an immovable object. Our work focuses on the robot’s impact on human productivity rather than increasing system safety by using metrics for alteration of the robot’s trajectory, circumventing an immediate need to directly address

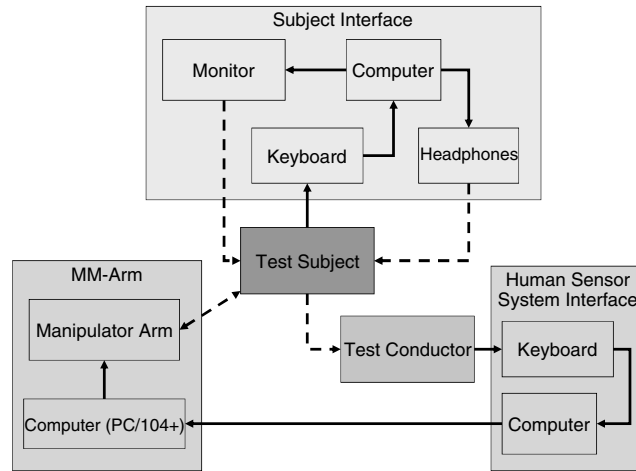


Fig. 1 Hardware subsystems for human-robot experiments.

the ‘safe trajectory planning’ challenge. Industry is incrementally revising their industrial robot safety standards to allow limited human movement within the reachable workspace during active operation [22–24]. The factory floor environment is more engineered and thus predictable than a planetary surface and likely even an EVA environment, so application of such standards to space robotics would be a challenging endeavor itself outside the scope of this paper, driving our testing constraints.

Human-robot operations require the communication of intent between agents, as a shared understanding and agreement upon common goals is necessary for effective collaboration. Robotic system understanding of human intent is critical in a shared environment. When the human’s activities are prioritized over robot activities, as assumed in our work, the robot must first ensure it does not interfere with the human before attempting to efficiently achieve its own goals. Although supervision by the human could eliminate the possibility of misinterpretation of intent, there is a nontrivial workload associated with communicating clear, concise, and relevant intent and instructions to a robot, and the situational awareness issues associated with such oversight further decrease the human’s productivity level [30,31].

Researchers have concluded that inclusion of a ‘smart’ manipulator results in a decrease in overall human workload and increase in productivity [32,33]. Our paper investigates similar workload and productivity issues when more intrusive and persistent incursions into the shared workspace occur. To gauge human productivity in our experiments, we use simple objective and subjective metrics of human performance and workload, as we focus on the robot’s impact rather than human-robot task activity tradeoffs. Examples of additional system- and agent-level performance measures to gauge productivity level are given in [34]. A survey of alternative noninvasive visual human sensing methods, from which intent could be derived, is provided in [35]. Other works also directly discuss prediction of intent from explicit and implicit nonverbal communication without action-mapping [36,37]; these show the feasibility of using ‘ideal’ intent information without regard for source, as is done in our experiments. This allows us to eliminate additional workload from implicit communication because the test subjects know they can ‘trust’ the robot to determine intent without overt verbal or gestural cues.

### III. Experimental Apparatus

Our experiments placed a robotic manipulator and seated human in a common physical workspace populated with fixtures that allow the robot and human to accomplish simple but realistic task-level goals. Figures 1 and 2 provide an overview of the test setup, detailing the three interfaces used in human-subject testing: the robotic manipulator arm, the test-subject interface, and the human sensor system (test conductor) interface.

Note that the dashed lines in Fig. 1 show data that have an indirect impact on the receiving block, while solid lines indicate explicit communication between blocks via keystrokes on a keyboard or a direct data connection between components. Also note that, although the task activations (‘button activations’ and ‘message display information’) in Fig. 2 were created offline, they are not known to the test subject or robotic agents before their activation times.

The computers used by the test subject and test conductor were Dell single-core systems with adequate processing power running a Linux-based operating system. The robotic manipulator used in the experiments is called the MichiganMan(ipulator) arm, or MM-arm, a low-cost platform designed in-house and constructed primarily from commercial off-the-shelf (COTS) hardware [1,38]. The MM-arm is controlled by a PC/104+ computer stack, which receives ‘ideal’ sensor data from the ‘human sensor system’ computer. These data are generated in real time by a

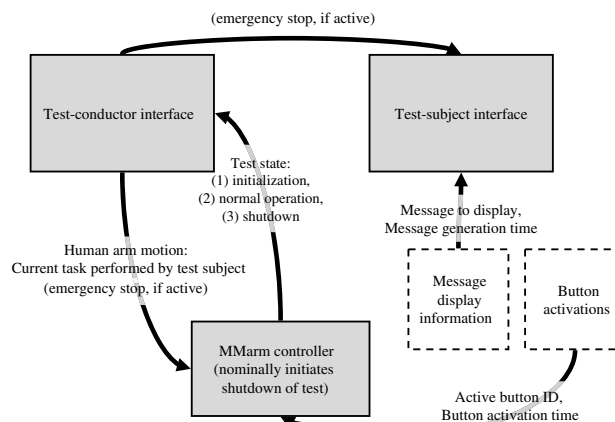


Fig. 2 Software infrastructure.



**Fig. 3 Test-conductor interface keyboard bindings.**

human test conductor who indicates changes in the human test subject's task status via keystrokes, circumventing many challenges in sensing and inference associated with human task recognition that would otherwise complicate the testing. Meanwhile, a test subject interacts with a separate computer interface, nominally completing cognitive (math) tasks and sporadically completing physical tasks prompted by computer-generated messages (e.g., drink soda, eat a chip, press button). Test subjects listened to low-impact music on noise-cancelling headphones to tune out background noise, in particular the manipulator motor noise that introduces additional, possibly distracting, auditory information whenever the manipulator moves between target poses.

Communication between computers occurs through a wireless router. During testing, the data shown in Fig. 2 were communicated between modules in real time via basic TCP/IP socket protocols. The data processed posttest included this information as well as information logging keystrokes and internal processes. All three C++ software modules in Fig. 2 (one per Fig. 1 hardware module) time-stamped these data and stored them locally on each computer system in text files; the computers had synchronized system clocks, and data files were transferred to a main server after each test. The MM-arm control software included a communications interface, a task scheduler, and a motion controller. The test conductor interface included a communications interface and a keystroke logger, while the test-subject interface included a more advanced visually interactive keylogger and a timed reader–displayer to display task activation messages. Both user interfaces were command-line text interfaces, capturing user keystrokes and reacting as described next. Shell scripts were set up on all three systems to help automate this process: reading a command from the console to determine the test set and test to run, synching the times, running the software modules using the corresponding test script, and copying the collected data to a centralized location once complete.

#### A. Test Conductor Interface

The purpose of the test conductor interface (simulated human sensor system) was to receive keystrokes from the test conductor corresponding to portions of the test subject's motion over the course of their task completion progress and send this information to the MM-arm control software for use. The keys used on the test conductor's keyboard were marked with colored paper with the possible test subject actions. Keys indicate human arm and hand motion segments enacted to complete the particular task in progress. For example, the 'eat chip' task requires human arm movement outward to the chip bowl, grabbing a chip at the bowl, and return arm movement to bring the chip to the mouth for consumption. Keys for a subtask sequence were chosen in a close grouping together on the keyboard and given the same color-coding. Figure 3 shows the key layout used.

Note that no differentiation was made between motions for a particular task by the MM-arm control software, and mappings from task to conflict were predefined. The test conductor interface software was a simple text-based program run at the console. During testing, the screen listed the key bindings and last selected motion. One key was also set up as an 'emergency stop' switch for the MM-arm robot, as a last resort safety precaution. This switch stops all three programs gracefully, pausing the MM-arm before shutting it down and exiting the two interfaces.

#### B. Test-Subject Interface

The purpose of the test-subject interface was two-fold: 1) to show math problems and task activations to the screen, record keystrokes, and update the screen with partial math solutions or clear messages over the course of each test, and 2) to receive survey data from the test subject after each test. The keys used on the test subject's keyboard were marked with colored paper: the number line, backspace, delete, enter key, and spacebar. Math problems were solved right to left in a tabular format using the number line (not number pad) and enter key to submit answers; this promoted use of only the right hand during testing. The ~ symbol was used to represent the user's current cursor position onscreen, and the backspace and delete keys removed the number value to the right of this cursor, to allow for the correction of mistypes. At the bottom of the screen below the math problems, predetermined messages instructing completion of an interruptive physical task were displayed on an interval specified by a test script. Test subjects were instructed to press the spacebar once when a task activation message is first registered and again after completing the corresponding physical task. Task activation messages were presented sequentially; in the event that a human-assigned task was not complete before a new task activation, the old task was dropped and the new task replaced the old. The test-subject interface software was a simple text-based program run at the console, as shown in Fig. 4. After each human–robot test series, this interface was replaced with an instance of an open-source spreadsheet editor and a survey to be completed.

#### C. MichiganMan(ipulator) Arm

Figure 5 gives an overview of the physical workspace in which human-subject testing was conducted. The manipulator was located in the center of the space (shown at full extension), while the human subject sat in the chair (lower center) with task locations to the left, right, and forward of



**Fig. 4 Sample math problem display (with blue waypoint target in foreground).**

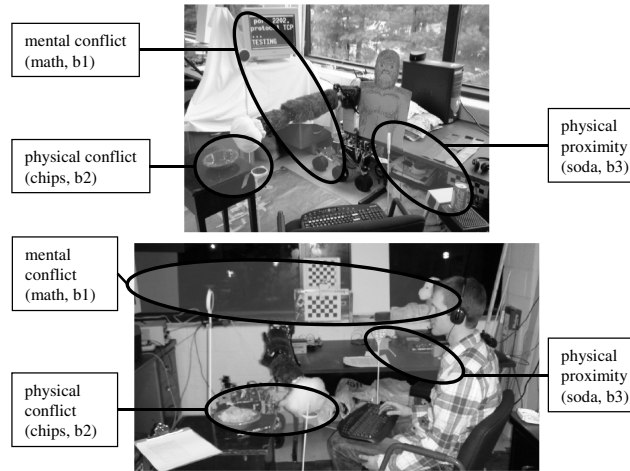


Fig. 5 Workspace setup with MM-arm. Buttons b1, b2, and b3 are indicated to the test subject by blue reflectors.

his/her station. A monitor for display was situated directly in front of the chair and inclined lightly above head level. The manipulator and human were situated near each other so they shared the majority of their physical arm and end effector (hand) workspaces. This overlap in collaborative space did not extend into the space containing the human's head, torso, or legs when the human had a seated neutral body position, a design feature we conveyed to help test subjects focus on tasks completed in the overlapping workspace. The position of tasks and of objects associated with task completion were chosen such that manipulator arm and human arm movements would come into conflict but also in such a way that manipulator would only physically conflict within a small envelope of trajectories the human was expected to take when completing physical tasks.

The MichiganMan(ipulator) arm (MM-arm) was designed to move in a workspace comparable to that reachable by the arm of a seated human. Emphasis was placed on ensuring that the MM-arm would be safe for the human-subject testing described previously. The MM-arm is a low-power, lightweight 4-DOF roll-pitch-roll-pitch manipulator developed from low-cost components by University of Michigan students. This fixed-base manipulator has size, speed, and range of motion similar to a human arm and its D-H (Denavit-Hartenberg) parameters, which follow Craig's convention [39], approximately correspond to both a human arm and NASA's Robonaut 1 (R1) system [1,40]. These are given in Table 1. For our implementation, we specify an offset from the final joint axes to a fixed tooltip frame at  $d_5 = -11.25$  in., which extends from the wrist to the center of the soft padded 'hand'. The forward kinematics of the manipulator can then be represented by a single transformation matrix. The MM-arm has an analytical solution for its inverse kinematics, due to the joint alignment.

MM-arm joints are plastic pan-tilt shoulder and elbow joint mechanisms, while the linkages between the joints are composed of multiple carbon fiber tubes. Cushioning materials of a neutral color are added over the quasi-rigid structures to mitigate any impact force a collision with the structure might impart, increasing the safety of the system. The manipulator wrist is covered with a ball of white cotton batting material to make it visually attentive and to soften the end. Speed and torque of the manipulator actuators are constrained to levels that would not be capable of injuring the seated test subject, particularly given the padding affixed to the manipulator. Commercial off-the-shelf (COTS) digital servos were used as the joint actuators, because they are robust, reliable, and relatively low-torque and low-power compared to typical industrial manipulator motors and motor-driver systems. The particular servos used have internal PD controllers and can be treated as black-box mechanisms with only joint angle inputs. Battery power is fed into a line driver board, preventing runaway and overvoltage power situations. This low-power setup mitigates the fact that there is currently no sensory closed loop; the MM-arm cannot tell whether it has reached its destination or whether contact has or may shortly occur. Rekimoto [41] gives an example of a different research platform with such sensing capabilities.

MM-arm's closed-world model includes all goal location and human conflict information for goal-based decision making. 'Closed world' for the MM-arm domain is a valid assumption in our experimental setup due to the further assumptions that all targets of interest remain at known, fixed positions, that the MM-arm has a fixed base, and that the human is seated in a known location with features (e.g., arm length, gaze height) similar to those of a 'generic' human model. We explicitly specify a conflict mapping before testing. For each human task, we define a set of robot poses that could or would potentially cause a physical or mental (gaze) conflict with the human. Because of the closed-world assumption, we can compute offline the set of trajectory poses or three-dimensional envelope through which the robot will command movements while completing each task. We can compute a similar envelope for the human tasks using a generic human kinematic model and then compare these to determine the set of conflicting task-pose pairs. Although this procedure was sufficient for tests described in this paper, real-world applications where either agent may not have a fixed-base location would likely require online identification of local conflict sets based on target, robot base, and human body movements near and through the shared workspace.

During each test, the MM-arm control software receives the human's current task, checks the conflict mapping, and selects its current goal. A PC/104+ computer stack, two serial servo controller boards, and a line driver board are used to command and control the manipulator servos at the 9600 maximum baud rate imposed by the servo controller boards. For this work, the robot's task scheduler is a simple first-in first-out queue with the task sequence for each test predefined in a text file (test script). To reduce risk and annoyance in a shared workspace, the robot was instructed to defer to human activities to 1) avoid physical contact, and 2) minimize human gaze obstruction when the human is viewing the computer screen. This simple strategy effectively gives the human tasks priority over robot tasks. In the event of a conflict, the queue blocks as long as the conflict exists. The first task with no conflict is initiated; if no such task exists, the robot moves to a preset 'neutral' (e.g., stowed) pose that never conflicts with the human given our workspace configuration or maintains its station if already at that pose. This simple task queue is sufficient for our experiments and would be replaced by a planner-scheduler in a deployed platform.

Table 1 MM-arm D-H parameters

$i$	$\alpha_{i-1}$ , deg	$a_{i-1}$ , in.	$d_i$ , in.	$\theta_i$ deg
1	0	0	3.814	$\theta_1$
2	-90	0.345	0	$\theta_2$
3	-90	-0.345	-16.5	$\theta_3$
4	90	0	0	$\theta_4$

**Table 2** MM-arm poses

Pose name	Joint angle pose deg				End-effector location from MM-arm base, in.		
	$\theta_1$	$\theta_2$	$\theta_3$	$\theta_4$	$x$	$y$	$z$
Stowed	90	-90	0	0	0	-27.405	3.469
Unstowed 1	0	-90	0	90	-16.155	0	14.719
Unstowed 2	0	-32	0	90	0.8493	0	23.5856
b1	-55	-53	0	47	-8.154	11.6451	24.6568
b2	-35	10	21	6	4.5867	-3.7261	30.9510
b3	0	45	0	68	22.124	0	11.3295

The MM-arm was designed to emulate human arm motions in part because this has been a convention for human-interactive robots (e.g., for Robonaut) and because humans are good at predicting how other humans move [42]. Although explicit communication is disallowed for the testing, the human may still want to implicitly predict what the MM-arm is doing, even if this prediction is merely to establish a higher level of comfort or confidence in the robot. MM-arm movement is determined by a simple ‘direct path’ joint-space algorithm that plans a smooth joint-space trajectory between forward-kinematic poses. Each robot-assigned task had only one unique ‘goal pose’ associated with it. These poses were found by using a test program to move the MM-arm such that the tooltip location corresponded to the necessary goal location (e.g., button locations), and the pose was selected according to the visual obstruction requirements for that goal pose. Once the final pose was determined, the joint angle set was recorded and used in subsequent testing.

Table 2 gives representative poses. Once a task with corresponding goal pose is selected, an arrival time  $T_f$  to reach this goal pose from the initial pose is determined (generally the initial time of goal selection  $T_s$  plus a preset time-of-motion, e.g., 2 s). To execute this maneuver, joint angle commands are incremented as shown in Eqs. (1) and (2):

$$\Delta\theta = (\theta_{fi} - \theta_{si})/N \quad (1)$$

$$\theta_i(t + \Delta t) = \theta_i(t) + \Delta\theta_i, \quad \theta_i(T_s) = \theta_{si}, \quad \theta_i(T_f) = \theta_{fi} \quad (2)$$

$N$  for our experiments was 30. A larger  $N$  helps reduce the speed of motion by commanding more interim waypoints to be met. This constrains manipulator joints to move at speeds significantly slower than the maximum rates despite the use of COTS servo modules. Such slow speeds promote safety and minimize distraction for the test subject. Within the range of acceptable safe speeds, our choice of  $N$  and time-of-motion were made carefully. A perception of too-slow motion would cause irritation and lower productivity, due to increased chance and duration of conflict, while motion perceived too quick would cause undue stress, apprehension, or fear. Either would lead to eventual distrust of the robot’s motions. Feedback from test subjects on manipulator speed is described next. Note that because the arm movements are done in joint space, the speed of tooltip motion is not consistent between poses, but for these tests safety and distraction factors dominate concerns over consistency in tooltip speed. Manipulator and servo constraints restrict the choice of valid angles for each joint. In software, we further limit the elbow joint to emulate a ‘human-like’ range of motion. With smooth motions between arm poses, the MM-arm never attempts to move outside the joint limits.

#### IV. Experiment Assumptions and Constraints

To realistically scope experiments and focus results on the examination of efficiency and workload, we make the following assumptions.

- 1) No faults or failures occur in either actuators or sensors (including test conductor data entry). Tests in which any such issues occurred were terminated and not used in our analysis.
- 2) Seated human torso and manipulator base locations are known and constant, and all waypoint targets for human and robot tasks are stationary and predefined. This enables definition of robot trajectories and conflict sets before test script creation and testing.
- 3) No appreciable learning curve, no task-switching subject overload, and no variability from choice of handedness existed. We investigate this learning curve assumption as part of our experimental results.
- 4) Test subjects stay on task unless the robot distracts them. The learning curve and task focus assumptions allow us to assume user preference and task priorities are constant.
- 5) Tasking with the same level of difficulty require approximately the same execution time to simplify data processing, and each task has common best-case and worst-case execution times. This last assumption allows the use of test scripts developed offline for all test subjects.

#### V. Test Matrix

In this section, we define the task sets the human–manipulator team complete and identify appropriate combinations of conflict scenarios to test our hypothesis. The test subject is asked to complete three types of tasks: cognitive tasks, multistep physical tasks, and movement-only physical tasks. The robot can only accomplish (physical) movement-only tasks in the shared workspace. Although the necessary tools and objects for task completion are shared in some cases (e.g., button pressing), for our experiments human and robot tasks are presumed independent, i.e., neither human nor robot can offload tasks from the other. To identify changes in human performance, we perform direct comparisons between complementary pairs of tests: one where the human performs all tasks and one where tasks the robot can accomplish are offloaded from human to robot. (In these experiments, responsibility for completing movement-only physical tasks is given to only the human or only the robot for any particular test. The human and robot are not given the same type of task in the same test; real-time collaboration in task assignment is beyond the scope of this work.) We treat the cognitive task as ‘most important’ and the physical tasks as desired but either unplanned or unexpected. To identify changes in test-subject workload, we chose low-impact tasks with minimal learning curve so the metrics would be more sensitive to changes from the robot’s activity than differences in task sequences. To accommodate this, simple, everyday tasks were chosen that have analogues to tasks completed during on-orbit operations. These tasks were independent, easily completed, and not expected to create cumulative fatigue given the short test duration, making robot activity the sole major impact on productivity.

To identify changes in (human) focus of attention, avoidable collisions or close-proximity motion of the robot near the human are represented by classifying human–robot paired tasks as conflicting versus nonconflicting. This yields four task classes: 1) nonconflicting, robot stationary, 2) nonconflicting, robot moving, 3) visual (gaze) conflict (mental conflict), and 4) physical conflict. The first represents tests in which the robot

• Nominal robot action
• Robot action without consideration for human test subject actions (showcases all arm movements to the human, a demonstration case)
• Ideal
• Nominal human cognition only (math problem-solving with the manipulator at a static workspace position)
• Human-only
• Human problem-solving with interruption (math tasks with interspersed messages to eat chips, drink soda, press buttons, or some combination thereof)
• Robot-active, no-conflict
• Collaborative operation baseline with manipulator sporadically avoiding the human (similar to human-only, but with button tasks offloaded to the robot – best case with robot)
• Robot-active with conflict
• Collaborative operation inducing frequent conflicts (mental conflicts, physical conflicts, or a mix of both)
• Robot-active with conflict, overtasking
• Collaborative operation “stress tests” (more tasks than can be accomplished, or close to it – worst-case with robot)

Fig. 6 Test scenarios.

was not moving to complete any tasks, while the latter three encompass active robot tests. The use of test scripts of tasks and activation times can then ensure the human–manipulator team will encounter the appropriate suite of conflicts.

We first defined specifics of each task, ensuring at least one task of each type was specified. We used mathematical addition as a cognitive task, consumption for the multistep physical tasks, and ‘button pressing’ for the movement-only physical tasks. These tasks were chosen because they have similar analogues to common tasks in a space environment, such as an EVA for spacecraft repair. The cognitive task requires moderate concentration and a clear line-of-sight, as might be required when following familiar steps to inspect or diagnose a problem with an electronics module. The physical tasks involving consumables are multimotion pick-and-place tasks, one for stowing or retrieving consumables such as repair or habitation module supplies. The button-pressing involved ‘quick’ movement to handle an overriding concern, such as grabbing a toolbox that might be able to float away.

For the math problems task, we chose randomized simple addition problems, in a three-line  $XXX + XX = ?$  format to avoid cognitive delay that would otherwise result from switching math operations. Input was acquired right to left (ones, tens, hundreds, thousands place) from the keyboard number line to nominally occupy both hands and to avoid an ‘expert’ number pad typist biasing statistics. The two consumption tasks were eating chips from a nearby chip bowl and drinking soda from a nearby soda can. ‘Button pushing’ required touching reflectors at static locations b1, b2, and b3 and was simulated as an external interruptive event. Next, we chose test-subject task locations and robot poses to satisfy the task classes: nonconflicting stationary as the robot’s default (robot at rest and stowed so it posed no conflict, nc), nonconflicting human–robot task combinations with nearby moving robot (e.g., drinking soda with robot goal to press b3, nc), and human–robot tasks with conflict either in the subject’s visual gaze given math problems with robot goal to press b1 (mental conflict, mc) or in physical workspace overlap eating chips with robot goal to press b2 (physical conflict, pc). To best compare the impact of the robot on human performance, we constructed six test scenarios as shown in Fig. 6. The chosen physical placements allowed multiple robot-active no-conflict comparisons. Robot movement could occur both near and far from the human, or the robot could not move at all. The ‘eat chips during b2 activation’ and ‘solve math during b1 activation’ tasks were the only activities with conflicts studied in our comparisons. We expected changes in human productivity according to the types of robot task (proximity of the robot) and the speed of the robot, and we attempted to choose task activations exploring these factors in testing.

Next, we built a test matrix over a set of test scenarios and conflict combinations. Table 3 shows the mix of non/conflicting tasks for the first test session. The robot-only scenario (Z) is introduced before the first test session to familiarize the test subject with the robot’s baseline motions. All test sessions begin with a math-only test (ideal scenario, A) to provide a best-case performance benchmark for each subject and to help check for the existence of a learning curve. Tests then become progressively more complex, introducing consumption and button-pressing tasks into the human’s schedule and moving the robot through a series of progressively more conflicting task sequences. Task timings allow tasks to be completed before reactivation, except in overtasking scenarios.

The second test session consists of paired complementary tests to the first session given in the same time sequence as the first session. This avoids the possibility of unaccountable differences due to fatigue that could result from a differing order of completion. Complementary test scripts are created from a test script by modifying it so that button tasks assigned to the human are reassigned to the robot (leaving the human more time for math problem completion) when the original script is human-only. When the original script is robot-active, button tasks assigned to the robot are reassigned to the human, but only when the human does not already have a task scheduled. Physical tasks that would overlap in time (within an assumed minimum time of completion) are discarded because the human subject is not asked to remember a task queue. We discard overlapping button tasks when creating human-solo complementary tests because this should lighten the workload in the human-solo tests, and we want detrimental effects from the robot’s presence to be clear. The third test session also included tests across the Table 3 scenarios. For this session, the MM-arm moves more quickly when aborting motion toward b1 in an attempt to minimize screen occlusion (conflict) when the test

Table 3 Test session 1: distribution of task category combinations

Scenario type	Human tasks			Robot	Conflicts		Overtasking
	Math	Food/drink	Buttons	Buttons	b1	b2	
Z				X			
A	X						
B	X	X					
C	X		X				
D	X	X	X				
E	X			X			
F	X	X		X			
G	X	X		X		X	
H	X	X		X	X		
I	X	X		X	X	X	
J	X	X		X	X	X	X

subject resumes solving cognitive tasks earlier than expected. Comparisons with previous tests allow determination of manipulator speed impact on human performance.

The Appendix contains details on the number of interactions and conflicts that occurred in each test in every test set.

## VI. Test Metrics

In our human-subject experiments, we track changes in human performance and workload to support or disprove our hypothesis following the standard design of experiments practices established by the human factors and engineering communities [43,44]. We computed our objective performance measures from test subject keystroke data. First, we captured task completion time as well as the time elapsed between subject acknowledgement and completion of a task. We also evaluated cognitive task set completeness, correctness, and incorrectness rates, computed from the number of cognitive tasks (completed, correctly completed, or incorrectly completed) divided by the amount of time not spent on the physical tasks. These statistics provide measures of a test subject's ability to focus on the cognitive (math) tasks. The use of rates allows comparisons of tests across different test subjects regardless of test duration, number of task activations, or the comparative difficulty of the physical tasks.

We also investigated subjective performance measures. To establish metrics for subjective survey data acquired during our tests, we adopted the NASA Task Load Index (TLX), a procedure for obtaining subjective workload assessments developed during a 3-year, 40-experiment research effort by the Human Performance Group at NASA Ames Research Center [45]. The Task Load Index has been subjected to validation studies, some of which included supervisory control and laboratory tasks.

The TLX has six load sources: mental, physical, temporal, performance, effort, and frustration. These are measured by a combination of weights and ratings. The weights and ratings are selected by the test subject after each test. Each rating  $r_i$  is selected as a tick mark upon a number line between 0 and 100, with gradations of five points. Weights  $W_i$  for each load source are found by comparing pairs of load sources in unrepeatable combinations and tallying the number of times that load source is chosen. This gives a maximum of 15 points across all six weights, given by

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} = \frac{6!}{2!(4!)} = \frac{6 * 5}{2} = \frac{3 * 5}{1} = 15$$

with the maximum possible tally for any weight being 5. Weights account for "differences in workload definition between [subjects] within a task and differences in the sources of workload between tasks," while ratings "reflect the magnitude of that factor in a given task" [45]. Derived data such as the adjusted ratings  $R_i$  and the weighted rating  $R_W$  defined next can be compared between subjects with less variability than other methods:

$$R_i = W_i * r_i \quad (3)$$

$$R_W = \frac{1}{15} \sum_{i=1}^6 R_i \quad (4)$$

During data processing, we use the adjusted ratings for performance, effort, and frustration as well as overall workload  $R_W$ . Performance provides an evaluation of a test subject's view of their own performance. Effort and frustration capture workload changes directly related to the robot's involvement. Given its definition in the TLX instructions, effort should be the most sensitive load source to any learning curve. Overall workload is divorced from test and test-subject specifics and is used for cross-subject comparisons, similarly for the objective rates.

To put the TLX load scale ratings in perspective, adjusted ratings  $R_i$  have a range of 0 through 500, and the weighted rating  $R_W$  has a range of 0 through 100. We define 'noise' as a change in the weighted rating of a load source by up to  $\sim 10$ –20 because a change in rating or weight by one gradation has a proportional effect on the adjusted weight. We describe how specific weights and rankings were determined in Sec. VIII. Higher ratings indicate the test subject felt higher workload or pressure from that particular source, except for performance rating where lower ratings indicate a feeling of poor performance during that particular test. Higher weights indicate the test subject felt those load sources were more important contributors for that test than those with lower weights.

To identify changes in human performance, we perform direct comparisons between complementary pairs of tests, and the robot's impact, positive or negative, is inferred from differences in the task completeness and correctness rate and the human's perception of their own performance between paired tests. To identify changes in test-subject workload, we infer trends through 1) task-specific workload, comparing task completion times and task incorrectness rates between task types, and 2) test-specific workload, from the TLX survey data [45]. The TLX load sources demonstrating workload come from cognitive task completion (performance rating) and robot presence (frustration rating) as well as overall workload  $R_W$ . To identify changes in (human) focus of attention, we infer the robot's impact on level of distraction from differences in performance and workload between the human-solo and robot-active tests.

The ideal outcome of our experiments would be increased performance, decreased workload, and increased focus of attention on human tasks when the robot is active. This or neutral results in all cases would support our hypothesis and make a strong case for improvement of the robot's collaborative processes. Major sources of distraction could result from the visible motion of the robot or hidden stresses from the lack of explicit human-robot communication such as fear of collision. Fear of collision would suggest lack of trust in the robot, where the human would be more comfortable or efficient with supervisory control or a nonoverlapping workspace. Unavoidably, some distraction is expected due to hard-wired biological reflexes when confronted with motion in the field of view, but a lack of consistency in task completion times may indicate that the human was diverting more attention to the robot than attributable to uncontrollable impulse.

## VII. Test Methodology

This section describes our test setup and execution procedures. Human-subject testing was postapproved by the University of Michigan's Institutional Review Board (IRB) in Behavioral Sciences.

### A. Test-Subject Selection

All test subjects were Aerospace Engineering student volunteers found by posting flyers and via email solicitation. Test incentives were free food before, during, and after test sessions. A nearly even mix of graduate and undergraduate students participated, with one female subject. No subject had appreciable background in robotics or human-subject experiments.



## B. Test-Conductor Responsibilities

The testing process for each subject was supervised by a test conductor. Test conductors were required to take online certification tests created by the Behavioral Sciences IRB. Before testing, the test conductor educated test subjects on the testing process and fully informed them of all procedures before being asked to sign a consent form. The test conductors were also tasked with maintaining a safe, comfortable, and unpressured test environment as well as test-subject confidentiality; preparing and running the test equipment; and acting as the ‘ideal observer’ for the human sensor system interface.

## C. Testing Procedure

Three test sets of nine tests were developed. Nine test subjects were run through all three test sets in the main round of testing; an additional three subjects completed only the first test set in the initial round. Test activities and timings were the same across subject and managed by the automated test scripts. To avoid cumulative fatigue, each subject completed all three test sets within a two-week period, with a maximum of two test sessions per week and one test session per day within  $\sim 1.5$ –2 h, with each test running 3–4 min. Before testing, subjects were given a quick ‘demo’ of the robotic arm and had the opportunity to hold and physically move the manipulator when in a depowered state. This allowed test subjects to become more knowledgeable of the weight, inertia, and padding of the robotic arm. This introduction was designed to allow the subjects to make or better validate their own safety assessment when working near the robot. Subjects were also shown a ‘demo’ of robotic motion at this time to familiarize them with the robot’s common poses and movements during task completion. Test subjects were told before each test whether the robot would be active; this was done to avoid unnecessary ‘surprise’ or stress when the robot started its tasks. Without this information, the subjects might have felt increased levels of agitation or stress if/when the robot did not perform as the subject might naively expect, which would have required in-test discovery of this information, potentially resulting in a higher learning curve. Subjects were also told what types of tasks to expect but nothing regarding task timing or frequency, so they could not anticipate or plan task timings or sequences. Surveys were completed immediately posttest.

During testing, the test subject completed as many cognitive tasks as possible but was sporadically directed to complete the higher-priority physical tasks. Messages requesting the user complete a physical task were shown in large print along the bottom of the monitor where cognitive tasks were presented. Once requested, cognitive work could not continue until the physical task was completed and task initiation (acknowledgement of message) and task completion was logged by the test subject. Test subjects were instructed to complete physical movements at their preferred pace to create a more realistic and relaxed environment. They were told the robot would defer to their movements, and taking longer to complete nonconflicting tasks would allow the arm more time to complete its own tasks but instructed to focus on their own tasks rather than the manipulator’s motions or task progress. They were also told not to complete ‘button pressing’ tasks for the robot.

## D. Data Collection

We collected two main types of data to test our hypothesis: test-subject data (in three subsets) and ‘sensor’ data. Objective quantitative data derived from test-subject keystrokes (type I), which captured all GUI interface interactions, were postprocessed to retrieve the time duration and success or failure of each completed activity as related to human performance for the computation of rates. Subjective qualitative data derived from NASATLX surveys (type II) capture the test subject’s opinions of how efficient they thought they were, the relative stress they felt from the given workload, and their comfort level with the robot. An open-ended questionnaire (type III) allowed more direct answers on the test subject’s comfort level with the testing and robot interaction. Quantitative sensor data obtained from keystrokes by the test conductor were available to the robot in real time and tracked the test subject’s intended task goal (type IV).

## VIII. Results

This section describes objective and subjective results compiled for our nine test subjects over the metrics and test scenarios outlined previously. As described previously, human productivity was determined from performance and workload comparisons between paired complementary tests and between human-only versus robot-active tests. Objective rate data were used for all 12 test subjects to assess any learning curve effects and evaluate human performance. However, data from the three test subjects who only completed test set 1 were removed from the learning curve and paired complementary curve comparisons. Subjective TLX data were collected in digital form for nine test subjects to provide further evidence for existence or absence of a learning curve as well as workload. We used eight of the nine test subjects’ data. We excluded subject 9’s TLX data, as his ratings were anomalous. For subject 9, the data saturated the low end of the recordable TLX rating scale and showed minimal, if any, variation in the ratings. Overall TLX workload scores for subject 9 were outliers at the low end compared to all other subject TLX scores. Because we examined differences in ratings, these data were generally unhelpful for subject-specific analysis. Note that exclusion of subject 9 data did not give the data a hypothesis-friendly bias; inclusion of the subject 9 data would in fact have helped bias support toward our primary hypothesis.

We also discarded incomplete TLX survey data on a test-by-test basis. Comparisons between the two types of data identified possible correlations between workload and objective performance. For subjective metrics, we define ‘no significant change’ as data occurring within one standard deviation. We plot this as simple lines for single-test comparisons and box plots for aggregated tests, with the box centered on the average and spanning standard deviation. For objective metrics, we define ‘no significant change’ as confidence within noise of the median value(s). We plot this using bowtie plots, with the lower quartile, median, and upper quartile values as horizontal lines; vertical whiskers extend to the extrema while +s at the outer edges symbolize outliers, and the ‘notches’ (sloping inward lines of the bowtie) show the span of negligible noise. No overlap between ‘notches’ implies that the data have different medians with 95% confidence (also known as the difference at the 5% significance level, similar to the T-test for means). We discuss statistical outliers in the context of our hypothesis. Next, we first examine learning curve effects. Then, we perform paired complementary test and task comparisons. Finally, we study task completion time and subject-reaction comparison results.

### A. Learning Curve

The TLX data gave mixed results in subject perception regarding learning curve. For the first test set, the order in which tests were run was: test Z, solo tests A–D in that order, and the rest included robot assistance (see Table 3 for ‘scenario type’ identifiers). Test A was the math-only baseline, B was consumption-only, C was button-pressing only, and D included all human task types. Expected results according to the mix of task type and relative difficulty level are given in Table 4; test letter indicates what metric value listed on that row should be plugged in for that test. Figure 7 gives an example of a case exhibiting no learning curve effect.

From the test set 1 data, aggregating across all subjects, the standard deviation is  $\pm 20R_i$  or  $R_w$ , and the noise level is  $\sim 0.04$  – 0.05 problems per second (about 2–3 problems per minute). The specifics of single-subject comparisons to these values for no significant change are listed in Table 5; on a row, numbers indicate subjects for which that type or lack of learning curve held true, — indicate subjects for which it was unindicative of a learning curve, and \* represent subjects excluded from comparison due to anomaly or lack of data. We include a symbol  $\hat{\phantom{x}}$  denoting a not-statistically-significant learning curve (a trend seen with noise level of 0) for completeness. The TLX data shows no perceived (subjective)

**Table 4** Expected relationship of test result data in test set 1

Metric being compared	Learning curve			
	No	Yes (distinct)	Only math	Only consumables
TLX rating (subjective)	$\min(B, C) < D$ or $D < \max(B, C)$ ( $\min(B, C) < D$ )	$D < \min(B, C)$		
Correctness rate (objective)	$D < \max(B, C)$ or $\max(B, C) < A$	$A < \min(B, C)$ and $\max(B, C) < D$	$A < \min(B, C)$ and ( $\min(B, C) < D$ ) or $D < \max(B, C)$	$\max(B, C) < \min(D, A)$

**Table 5** Exhibited statistically-significant learning curves: test 1 comparison

Learning curve	Subject number											
None — overall workload	1	2	3	4	5	6	7	8	*	*	*	*
None — correctness rate	1	2	3	4	5	6	7	8	9	10	11	12
Distinct — effort rating	1	2	—	—	—	—	—	—	*	*	*	*
Distinct — correctness rate	—	—	^	^	—	—	^	—	—	—	—	—
Any — correctness rate	^	—	^	^	^	—	^	^	^	^	—	^
Only consumables — correctness rate	^	—	—	—	^	—	—	^	^	^	—	^
Only math — correctness rate	—	—	—	—	—	—	—	—	—	—	—	—

\* denotes subjects that were excluded from the comparison, while a dash denotes a negative conclusion. ^ denotes a not-statistically-significant learning curve (noise level of 0); a number denotes significance.

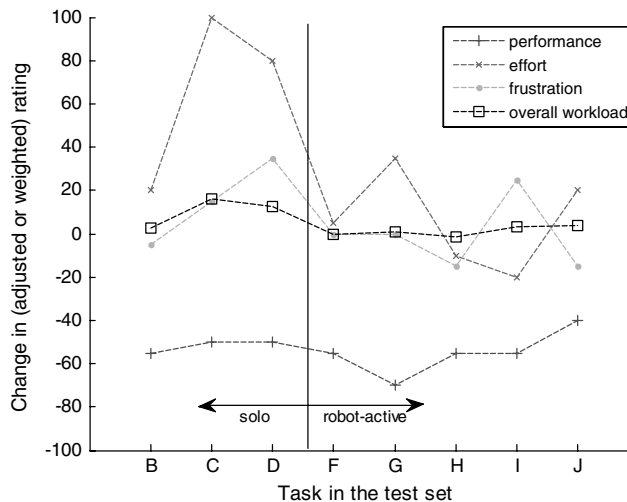
learning curve for overall workload for any subject and a first-day learning curve for two test subjects in effort rating. The correctness rate data shows that no subject exhibited a statistically-significant learning curve. Three subjects exhibit no learning curve trend at all, and three subjects exhibited a distinct but not-statistically-significant learning curve on the first day. So, 1/4 of the subjects showed a clear trend, while three-quarters of the subjects showed a trend of some sort. There is a clear discrepancy between the TLX data and the objective data.

We also check for learning effects across test sets, which is indicated by a marked decline in the baseline cases across all three days that the test sets were performed. However, the correctness rate data show no significant decline between the baselines, and neither do the TLX results shown in Fig. 8. Both the objective and TLX data thus indicate no learning curve across different days.

Figure 9 shows the aggregated correctness rate data for each test type on the first-day across all subjects. Differences are within noise for tests B-D, indicating no general first-day learning curve for test subjects. We also verified there is no significant change across the multiple test days (not shown). Thus, no general learning curve from both the objective and subjective datasets is seen across multiple days, nor does the objective rate data show a trend to indicate the presence of learning curve on the first day. Because the TLX data straightforwardly showed no learning curve for individuals when comparing overall workload (the main metric), there is also no general learning curve from the TLX results. There are individual subjects who appeared to believe they experienced a first-day learning curve based on TLX effort rating (which could bias the results of the robot-active cases to seem more efficient); a slight learning curve effect might also be discerned in the rate data. However, rate differences were not appreciable, were inconsistent between individuals, and did not impact our conclusions, so we do not remove initial tests from consideration.

**B. Paired Complementary Test Comparisons—Robot as Subordinate**

To determine objective performance, we compare the correctness rates of paired complementary tests both singly and as aggregates across all subjects. This provides a comparison of human performance when the human must accomplish all tasks versus when the robot helps by completing button-pushing tasks for the human. Note that an increase in correctness rate represents performance improvement. A significant change is noted when correctness rates for single-subject solo versus robot-active cases are outside noise. An increase in correctness rate indicates



**Fig. 7** Selected TLX load source ratings relative to baseline: subject 5, test set 1.

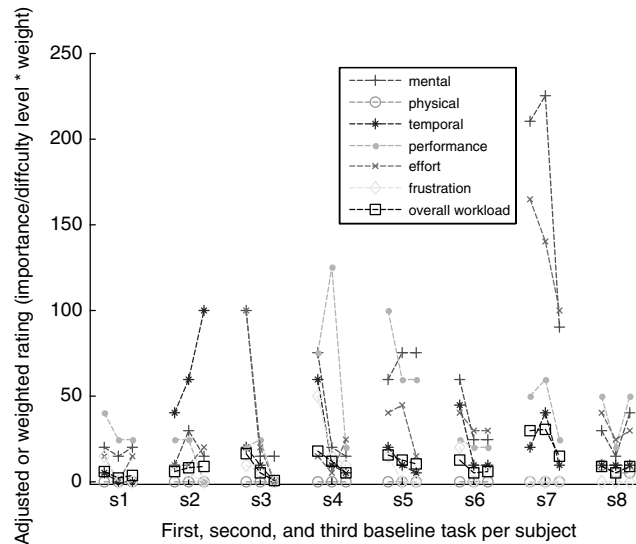


Fig. 8 TLX load source ratings for baseline cases over all test subjects.

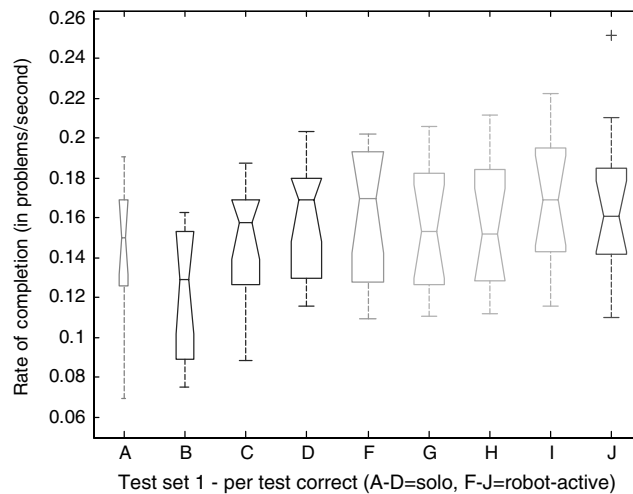


Fig. 9 Correctness rate for math problems: across all subjects, test set 1.

that the change in rate was outside error and higher in the robot case than the solo case; a decrease in correctness rate indicates that the change in rate was outside the error range and lower in the robot case than the solo case. Based on aggregate rates, some of which are shown in Fig. 9, the solo versus robot-active tests are within statistical noise in every case, with an average noise range of about  $\pm 0.02$  math problems per second. This implies no significant difference in the subjects' objective workload between when there is and is not visible robot motion in the human's field of view.

Table 6 shows a distribution of the number of test subjects who showed significant change in objective performance between their solo and robot-active paired tests. In this direct comparison, the nc cases (*E*, *F*) and one mc case (*H1*) show clear improvement in correctness rate when the robot takes over tasks, while clear performance degradation only occurred for pc cases (*G*, *I*). There are mixed results for the second mc case (*H2*). These results suggest that mental conflicts disrupt the subjects less than physical conflicts, while no conflict cases do not disrupt. However, this could be an artifact of the particular task because the mental math problems did not require the subjects to watch the screen at all times but rather encouraged many quick bursts of attention over the course of each problem. The overtasking case (*J*) also displays clear improvement relative to conflict cases without overtasking. This illustrates the existence of a tradeoff between overtasking and inclusion of the robot because the human was overtasked but not distracted by the robot in scenario *J*. These objective data provide evidence that a human does not have increased workload

Table 6 Trends for correctness rates: objective data<sup>a</sup>

Scenario type (robot case)	Number of significant increase	Number within noise	Number of significant decrease
E	4	5	0
F	3	6	0
G	0	5	4
H(1)	2	7	0
H(2)	1	7	1
I	0	6	3
J	4	5	0

<sup>a</sup>E-F = robot-active no-conflict, G-I = robot-active with conflict, J = overtasking.

**Table 7 Trends for overall workload: TLX data<sup>a</sup>**

Scenario type (robot case)	Number of significant increase	Number of within standard deviation	Number of significant decrease
E	0	6	3
F	0	8	1
G	0	8	1
H(1)	0	8	1
H(2)	0	7	2
I	0	7	2
J	0	3	6

<sup>a</sup>E-F = robot-active no-conflict, G-I = robot-active with conflict, J = overtasking.

when there is no conflict or only mental conflict with an active robot but that a human may have increased workload when there are physical conflicts with the robot. The worst-case degradation is from a pc test case, which resulted in a decrease of 0.07 correct per second. With average correctness rate of 0.15–0.2 correct per second, this is a drop of up to half the average rate, or 12–20 problems over a 3–4 min test. This result suggests we want to avoid close physical operations when possible and identify and mitigate all conflicts as soon as possible when close physical proximity of robot and human is required.

Table 7 shows a distribution of the number of test subjects who showed significant change in subjective overall workload between solo and robot-active paired tests. Note that an increase in workload implies the robot had a negative impact on performance. The TLX workload data have standard deviation of  $\sim 10$  for perceived overall workload, which is within the 10–20 of the TLX ratings scale noise discussed previously. Overall workload remains the same or decreases with the robot active, as does the effort subscale rating; none are higher than standard deviation. There are no clear indications from the TLX data (increases in workload) that correspond to the degradation (decreases in rate data) seen in the objective rates data. This dichotomy suggests that human subjects are more forgiving and ‘feel’ less stressed by the robot than the objective measure of their work output heralds.

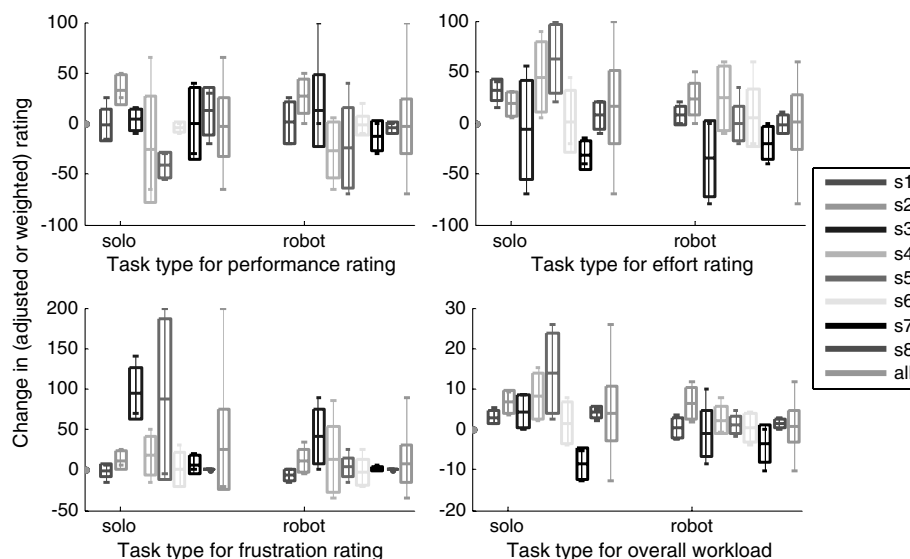
### C. Task Category Comparisons

#### 1. Solo Versus Robot-Active Cases

In this section, we compare two test groupings: those where the human works alone (human-solo) and those where the robot is moving to press buttons (robot-active). A significant increase in test-subject workload in the robot-active cases provides evidence that the presence of the robot was distracting the human, while no increase or a decrease would suggest the robot does not distract. Figure 10 shows adjusted ratings per-person across all tests within a test grouping, excluding overtasking cases. For each subject, for each load source, and for all cases, the change in TLX ratings was either within standard deviation or showed significant decrease. This indicates that the inclusion of the robot’s involvement did not change subjective performance or workload significantly. Those cases in which a clear decrease occurred were unsurprising, given that in the robot-active cases less physical tasks were assigned to the human and the mental task of solving math problems is of lower impact than other tasks.

It should be noted that, in almost all cases shown in Fig. 10, the maximum change in adjusted rating is not higher than 150, and the change in overall workload is not higher than 25. This implies that, worst case, there is a 30% change in adjusted workload and 25% in overall workload between nonovertasking cases. We also compute the objective incorrectness rate, providing an indication of math mistakes expected to be more frequent when a test subject is distracted by the robot. These data (not shown) are within statistical noise when comparing solo and robot-active cases per test set, per subject, except for subject 7 where the incorrectness rate was significantly lower than noise for the solo tests on test day 1. This is the only time this happened, so it was likely not a true improvement and this subject exhibited learning curve effects on test day 1. Our data therefore indicate that significant increase in test subject’s objective workload when the robot was active.

Next, we look for more subtle changes in correctness rates. The aggregate correctness rates in Fig. 11 show that, within each test set, the solo versus robot-active tests were within statistical noise, with an average error range of about  $\pm 0.02$  math problems per second. Thus, the rates show no appreciable difference from each other. Across all tests, per person, the correctness rates are all consistent in average. Differences are within noise when comparing the solo and robot-active cases, except for subjects 3 and 7 who exhibit improved (higher than noise) correctness rates for



**Fig. 10 Selected TLX load source ratings relative to baseline by task type, subject.**

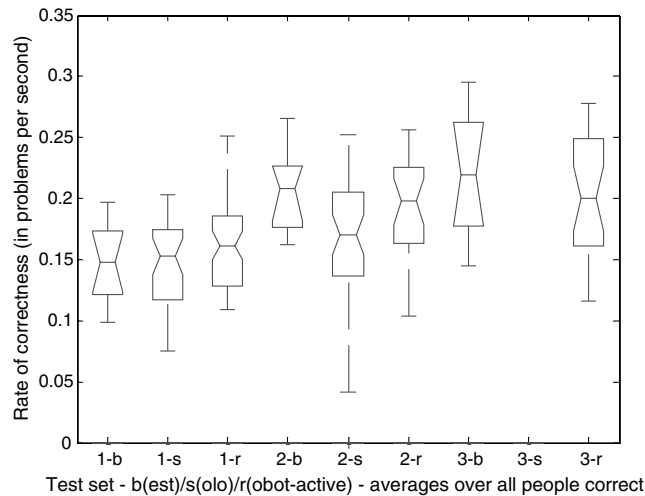


Fig. 11 Comparing correctness rates between test groupings across all subjects.

the robot-active tests than the solo tests on only the first test day, also probably due to learning curve, for reasons similar to the aforementioned, and thus negligible. These data also indicate that there are no major differences in subjective or objective workload with robot activity.

## 2. Robot-Active Conflict Cases

Although there is negligible distraction effect seen between the solo and robot-active cases, there is some variation within the datasets themselves. As discussed previously and shown in Table 3, there are several types of robot-active cases. For completeness, we also looked for trends in workload for the different conflict subcases: no conflict (nc), physical conflict (pc) only, mental conflict (mc) only, and both physical and mental conflicts (pc and mc). Note that the intersection of any of these subcases is empty.

We next compared individual TLX ratings over the robot-active conflict cases. Frustration rating was high for all subjects in the mc cases. Verbal and written survey results indicated the subjects wanted the robot to move away faster in these cases. The subjective performance rating for pc and mc was much higher than for other subcases; dealing with more conflicts while still completing all tasks may have given a greater sense of accomplishment than less challenging tests. Objectively, we see that, per person, the range of correctness rates across tests is 0.02 correct per second within a test set (and 0.05 correct per second across all tests) with the error about half that. This is a noticeable but not dramatic change because this only translates to a maximum difference of 4–6 problems over a 3–4 min test. Also, trends indicated no particular conflict type was better or worse, with overall performance consistent across conflict types in a per-person comparison. Type of conflict therefore does not appear to play a significant role in task completion or correctness, although subjects can be more frustrated by some conflicts than others.

In Fig. 12, we show aggregate results across all test subjects and test types. It is apparent that, regardless of the robot's involvement, differences within rating appear to fall within standard deviation. An interesting caveat is that the solo test-type cases have a larger standard deviation for effort and frustration and double the standard deviation in overall workload than do the robot-active cases. This echoes the type III data; many subjects responded that they preferred the robot to complete the 'annoying' button-pushing tasks. Such comments may be linked to the physical distance to the task location (the test subject had to lean forward to reach button b1), and nearby tasks reportedly caused less to no stress, according to type III data. Future work could explore this issue in more detail by varying button locations between tests, allowing mixed button pushing (collaborative task assignments) between human and robot and using directed surveys querying subjects about this specific preference. An alternative explanation is that the test subject realized that the robot could not complete two of the four nearby tasks (eating/drinking) and thus considered all tasks within this near physical space to be candidates for more efficient completion by the human test subject. In summary, both the subjective and objective datasets indicate that test-subject workload increases with differing types of robot involvement, which supports our initial hypothesis.

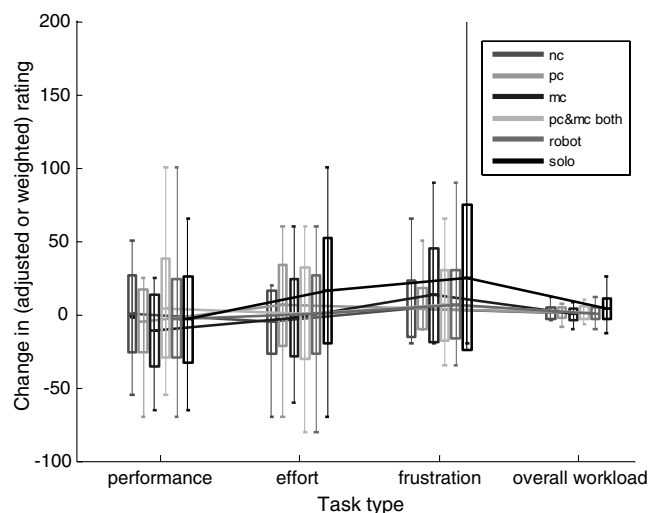


Fig. 12 Selected TLX load source ratings relative to baseline across all subjects and tests by task type (no overtasking cases).

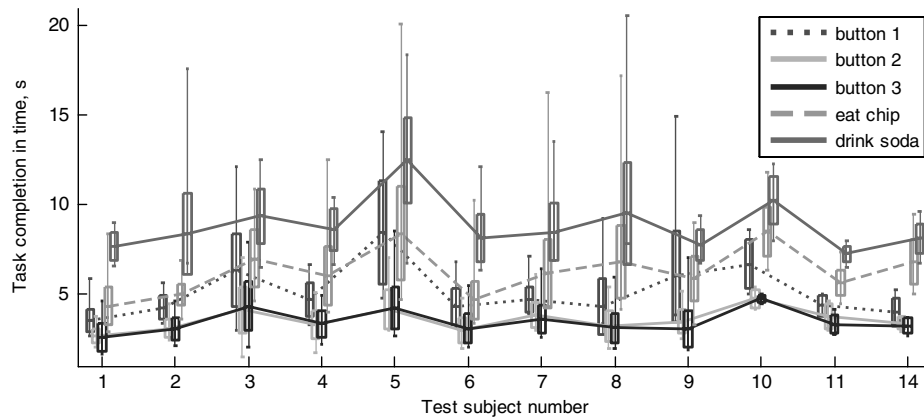


Fig. 13 Task completion times across all tests per subject.

#### D. Task Completion Times

If the offset of standard deviation for average completion time of a type of task does not overlap sufficiently between different tasks, then the workloads of the tasks can be considered different. Figure 13 shows aggregate task completion times.

From the figure, button b2 and b3 tasks have similar workload to each other but less than b1, which was most physically distant from the test subjects. Chip eating tasks by different subjects had a consistent standard deviation of  $\sim 1\text{--}2$  s, but the average time of completion differs substantially between subjects. Soda drinking tasks have less variation per-subject and took 3–5 s more time to complete than the chip tasks; however, the standard deviations are so large for food/drink consumption tasks that this is not a statistically significant result. Individually, some subjects were internally consistent (low standard deviation) for the consumption tasks; however, those who were inconsistent followed no pattern to link with a distraction. From observation, the chip and soda tasks have more timing variation in portions of the movement done at the task locations (grasping the chip/can and the actual eating/drinking) rather than in physical extension and retraction arm movements. Although this might initially appear to be motivation to change the eating/drinking tasks in future work, the larger variation in delays observed at worksites would be expected in real-world tasks where the human or robot would typically grasp, complete a physical or inspection activity, or release an object at each site.

#### E. Subject Reaction

Test subjects provided notable feedback regarding robot speed and conflicts. Many test subjects indicated the speed of the manipulator was too slow, particularly in conflict cases. However, many of these same test subjects said that the manipulator was moving too fast when the motions were sped up slightly in test set 3 (from a 3 s traversal to a 1.5 s traversal), though the faster motions occurred infrequently (the speedup was only for movement away from button 1 if the task was interrupted, and could only occur up to two times per relevant test). These data suggest the existence of a crossover point at which the manipulator's speed is perceived differently by the test subjects. Qualitatively, this crossover point likely exists between these two speeds for most subjects. Unfortunately, we cannot draw a strong conclusion at this time, as we did not include an explicit survey question to formally collect this information from every subject in this round of testing; this could be future work. However, despite this verbal reaction regarding manipulator speed, it is notable that, when asked directly, subjects said that they felt safe working near the manipulator at all times. A dichotomy of behavior was observed during the mc cases with b1; some subjects would attempt to look over the manipulator to continue their work, while others either waited patiently for the arm to move away or slowed their movements during other tasks to give the manipulator time to finish. This behavior suggests subjects were unclear on the proper protocol for this conflict, suggesting additional guidance in how to react to conflict (wait versus attempt to circumvent problem) in future testing. In this case, the uncontrolled factor was the tradeoff between the relative importance of completing the designated cognitive tasks in the timeliest manner versus conserving energy by patiently waiting.

In one interesting case not included in our statistics, a physical conflict occurred but the sensor signal was not sent by the test operator. The test subject quickly noticed the robot continued the conflicting task, paused to wait for the robot to finish, then completed their own task while casually commenting verbally upon the lack of response. This single data point suggests the human test subject could recognize and handle differences in the robot's behavior between cases in which it reacts to the human and cases when it does not (but remains a safe companion). This is important once the robot is responsible for sensing the human and predicting intent, as the goal would be for the robot to correctly predict human response in most cases, but it will be difficult to prove the robot's predictions will be accurate in all cases.

## IX. Conclusions

A hypothesis has been proposed that a collaborating human–robot team operating in a shared workspace can enjoy maximum productivity when the human need not supervise the robot. To support or refute this hypothesis through human–subject tests, a human–robot experimental apparatus was constructed and used that placed a seated human and fixed-base robotic manipulator in a shared physical workspace. Test-subject performance was characterized given a series of experiments in which a human completes cognitive and physical tasks with and without a manipulator executing its own conflicting or nonconflicting tasks. The goal of these experiments was to determine the extent to which a robot manipulator impacts human task performance and workload when operating in a shared physical workspace. The results, both objective and subjective, support the hypothesis: productivity of the human in the shared workspace remains comparable to productivity of the human working alone so long as the robot does not interfere directly with the human's physical motions or perceptual focus of attention.

Key results from the previous data processing section are summarized as follows.

1) No first-day learning curve is evident from overall workload. There are mixed results regarding first-day learning curve in objective performance (correctness rate); no statistically significant learning curve was evident from the object data, but there were some noticeable trends. No multiday learning curve is evident from either objective or subjective datasets.

2) The paired complementary test objective rate data indicate that test subjects are unaffected by robot presence except when physical conflicts occur. The subjective Task Load Index data, however, show no clear increase in workload under the tested conditions.

- 3) In solo versus robot-active test aggregate comparisons, data indicate that robot involvement did not significantly change subjective or objective performance or workload.
  - 4) Differing types of robot involvement show no general trend that test-subject workload increases in either subjective or objective datasets.
  - 5) All tasks had task completion times with standard deviation on the order of their task completion times.
- From our hypothesis, experiments, and data processing efforts, we draw the following conclusions.
- 1) Our hypothesis and assumptions were supported by our results, with the exception of physical conflicts where the robot was not able to perform efficiently.
  - 2) Fast real-time response by the robot is essential to avoid human productivity decrease in the case of physical conflict.
  - 3) Cases were observed where the robot needs to begin moving away from a possible conflict site before the human physically moves toward completing this conflicting action due to upper (safety-constrained) bounds on the robot's movement speed.
  - 4) Test subjects consistently felt safe with their robotic companion, with some subjects requesting higher manipulator speeds to minimize conflicts/delays.
  - 5) Subjects could easily interpret whether the robot was continuing to pursue its goals versus executing a conflict-avoidance action. The robot was not yet equipped to predict human intent beyond keyboard inputs provided by the test director.

The physical conflict scenarios suggest that intent prediction, potentially achieved with look-ahead or model-predictive planning to avoid the need for explicit communication of intent, could increase efficiency and minimize frustration when physical conflicts would otherwise arise. Intent prediction might redirect the robot to other objectives or might simply require the robot to expedite or delay its current task sequence. Scheduling algorithms with probabilistic prediction will be key to meeting safety constraints while also maximizing collaborative performance. Test-subject confidence in such a fully automated system will be important to evaluate in future tests.

Future work could also include longer-duration testing to determine the impact of human fatigue on ongoing human performance and overall system performance. If the robotic system is able to offload dull and repetitive tasks from the human without significantly impacting human productivity levels, we anticipate the human-robot team will be more productive than the human alone. Longer-duration tests can enable a study of the tradeoffs between varying human: robot task allocation and task intensity versus work session duration on cumulative fatigue. There is also a question of when or whether a tradeoff occurs in level of distraction due to robot proximity for the different types of tasks. This was tested somewhat but not exhaustively in this work; testing for these purposes would involve changing the physical placement of the task in the workspace (eating near versus eating far) as well as the proximity of robot motion. Direct comparisons of near-conflicts versus far-conflicts for physical and mental tasks were also not tested in this work (only near-physical and far-mental tasks were performed); testing with fewer or no mixtures of task types would be interesting to pursue. It will also be interesting to conduct a longer series of tests over several weeks or months, seeking individuals with a background in robotics, or offering less-knowledgeable participants a more-thorough understanding of the robotic test platform before testing. Tests could then examine the effects of extended prior knowledge and/or newly learned knowledge versus multitest experience on human productivity. Productivity comparisons broken down by gender and age might also be interesting, if a study with enough individuals for a statistically significant comparison can be conducted.

### Appendix: Task Timelines for Test Sets

Tables A1–A3 detail the number of interactions and conflicts and when they occurred in each test given.

**Table A1 Task timelines for human and robotic agents during tests in test set 1**

Test number	Scenario type	Tasks	Task timeline											
			(b1, b2, b3 = buttons 1, 2, 3; ch = chips, dr = drink; ma(th) all other times; nc, pc, mc = no conflict, physical conflict, mental conflict; = no-op)											
1	Z	Human	—	—	—	—	—	—	—	—	—	—	—	
1	Z	Robot	b1	b2	b3	b1	b3	b2	b1					
1	Z	Conflict												
2	A	Human												
2	A	Robot	—											
2	A	Conflict												
3	B	Human	ch	dr	ch	dr	ch	dr	ch	dr				
3	B	Robot	—											
3	B	Conflict												
4	C	Human	b1	b2	b3	b2	b3	b1	b3	b1	b2			
4	C	Robot	—											
4	C	Conflict												
5	D	Human	ch	b3	b1	dr	b2	b1	ch	b2	b3	dr		
5	D	Robot	—											
5	D	Conflict												
6	F	Human			ch		dr		dr		dr	ch		
6	F	Robot	b2	b3	b3	b3	b2	b3	b2	b2	b3	b3		
6	F	Conflict			nc		nc		nc		nc	nc		
7	G	Human	ch		ch		ch		ch		ch	ch		
7	G	Robot	b3	b2	b3	b3	b2	b2	b2	b2	b3	b3		
7	G	Conflict	nc		nc		pc		pc		nc	nc		
8	H	Human	ch		ch	dr			ch			ch	ch	
8	H	Robot	b3	b2	b1		b3	b2		b1	b1	b2	b3	
8	H	Conflict	nc		nc	nc			nc	mc	mc	nc	nc	
9	I	Human	ch		ch		ch		ch		ch	ch		
9	I	Robot	b2	b1	b2	b1	b2	b1	b2	b1	b2	b1	b2	
9	I	Conflict	pc	mc	pc	mc	pc	mc	pc	mc	pc	mc	pc	
10	J	Human	ch		dr		ch		dr		ch	dr		
10	J	Robot												
10	J	Conflict	pc	mc	nc	mc	pc	mc	nc	mc	pc	mc	nc	

**Table A2 Task timelines for human and robotic agents during tests in test set 2**

Test number	Scenario type	Tasks	Task timeline																
			(b1, b2, ba = buttons 1, 2, 3, all; ch = chips, dr = drink; ma(th) all other times; nc, pc, mc = no conflict, physical conflict, mental conflict; = no-op)																
1	A	Human																	
1	A	Robot	—																
1	A	Conflict																	
2	E	Human																	
2	E	Robot	b2	b3	b2	b3	b2	b3	b2	b3	b2	b3	b2	b3	b2	b3	b2	b3	b2
2	E	Conflict																	
3	H	Human																	
3	H	Robot	b1	b2	b3	b2	b3	b1	b3	b1	b2								
3	H	Conflict	mc				mc												
4	H	Human	ch				dr				ch				dr				
4	H	Robot		b3	b1	b2	b1	b2	b3										
4	H	Conflict		mc			mc												
5	D	Human	b2	b3	ch	b3	dr	b3	dr	b2	dr	b3	ch						
5	D	Robot	—																
5	D	Conflict																	
6	D	Human	ch	b2	ch	b3	ch	b2	ch	b2	ch	b3	ch						
6	D	Robot	—																
6	D	Conflict																	
7	D	Human	ch	b2	ch	dr	b3	b2	ch	b1	b1	ch	b2	b3	ch	b3	b3		
7	D	Robot																	
7	D	Conflict																	
8	D	Human	ch	b1	ch	b1	ch	b1	ch	b1	ch	b1	ch						
8	D	Robot																	
8	D	Conflict																	
9	D	Human	ch	ba	dr	ba	ch	ba	dr	ba	ch	ba	dr						
9	D	Robot	—																
9	D	Conflict																	

**Table A3 Task timelines for human and robotic agents during tests in test set 3**

Test number	Scenario type	Tasks	Task timeline																
			(b1, b2, ba = buttons 1, 2, 3, all; ch = chips, dr = drink; ma(th) all other times; nc, pc, mc = no conflict, physical conflict, mental conflict; = no-op)																
1	A	Human																	
1	A	Robot	—																
1	A	Conflict																	
2	G	Human	ch				ch				ch				ch				
2	G	Robot	b3	b2	b3	b3	b2	b2	b2	b2	b3	b3	b3						
2	G	Conflict	nc	nc			pc			pc			nc			nc			
3	I	Human	ch	ch			ch			ch			ch						
3	I	Robot	b2	b1	b2	b1	b2	b1	b2	b1	b2	b1	b2						
3	I	Conflict	pc	mc	pc	mc	pc	mc	pc	mc	pc	mc	pc	mc	pc				
4	J	Human	ch	dr			ch			dr			ch			dr			
4	J	Robot	[b1, b2, b3 reactivate every 5 s]																
4	J	Conflict	pc	mc	nc	mc	pc	mc	nc	mc	pc	mc	nc	mc					
5	G/I	Human																	
5	G/I	Robot	b2	b3	b2	b3	b2	b3	b2	b3	b2	b3	b2	b3	b2	b3	b2	b3	b2
5	G/I	Conflict																	
6	C/D	Human	b2	b3	b2	b2	b3												
6	C/D	Robot	—																
6	C/D	Conflict																	
7	C/D	Human	b1	b1	b1	b1	b1												
7	C/D	Robot	—																
7	C/D	Conflict																	
8	H/I	Human																	
8	H/I	Robot	b3	b1	b2	b1	b2	b3											
8	H/I	Conflict		mc			mc												
9	C/D	Human	b3	b1	b2	b1	b2	b3											
9	C/D	Robot	—																
9	C/D	Conflict																	
6	C/D	Human	b2	b3	b2	b2	b3												
6	C/D	Robot	—																
6	C/D	Conflict																	
7	C/D	Human	b1	b1	b1	b1	b1												
7	C/D	Robot	—																
7	C/D	Conflict																	
8	H/I	Human																	
8	H/I	Robot	b3	b1	b2	b1	b2	b3											
8	H/I	Conflict		mc			mc												



**Table A3** (Continued.)

Test number	Scenario type	Tasks	Task timeline					
			(b1, b2, ba = buttons 1, 2, 3, all; ch = chips, dr = drink; ma(th) all other times; nc, pc, mc = no conflict, physical conflict, mental conflict; = no-op)					
9	C/D	Human	b3	b1	b2	b1	b2	b3
9	C/D	Robot	—					
9	C/D	Conflict						

<sup>a</sup>Note that, for test set 3, in tests 5 through 9 no consumption messages were displayed but the human was told to eat or drink whenever they wished (thus, conflicts varied per person). These tests were not included in our data processing for this paper due to the high variability between subjects of when conflict cases could, and did, occur. It was collected towards later determination of variability and frequency of consumption task occurrence per subject in future work.

### Acknowledgments

The authors would like to thank the graduate and undergraduate students from the Department of Aerospace Engineering at the University of Michigan who volunteered to be test subjects for the MichiganMan(ipulator) arm (MM-arm) experiments. The authors also thank Jeremy Green, Kevin Matzen, and Ryan Wolcott, the student group involved in the construction of the MM-arm, as well as Gabriel Arroyo, a summer student who conducted several of the MM-arm tests.

### References

- [1] McGhan, C. L. R., and Atkins, E. M., "Physically-Proximal Human-Robot Collaboration: Enhancing Safety and Efficiency Through Intent Prediction," *AIAA Infotech@Aerospace Conference*, AIAA Paper 2009-1950, April 2009.
- [2] Williams, T., and Tanygin, S., "On-Orbit Engineering Tests of the AERCAM Sprint Robotic Camera Vehicle," *8th AAS/AIAA Space Flight Mechanics Meeting*, Monterey, CA, AAS Paper 98-171, Feb. 1998.
- [3] Wagenknecht, J., Fredrickson, S., Manning, T., and Jones, B., "Design, Development and Testing of the Miniature Autonomous Extravehicular Robotic Camera (Mini AERCAM) Guidance, Navigation, and Control System," *26th Annual American Astronautical Society Guidance and Control Conference*, American Astronautical Society, Feb. 2003.
- [4] Dorais, G. A., and Gawdiak, Y., "The Personal Satellite Assistant: An Internal Spacecraft Autonomous Mobile Monitor," *Proceedings of the IEEE Aerospace Conference*, Vol. 1, IEEE Publ., Piscataway, NJ, 2003, pp. 333–348.
- [5] Hexmoor, H., and Vaughn, J., "Computational Adjustable Autonomy for NASA Personal Satellite Assistants," *Proceedings of the 2002 ACM Symposium on Applied Computing*, ACM, New York, NY, 2002, pp. 21–26.
- [6] Biesiadecki, J. J., Leger, P. C., and Maimone, M. W., "Tradeoffs Between Directed and Autonomous Driving on the Mars Exploration Rovers," *International Journal of Robotics Research*, Vol. 26, Jan. 2007, pp. 91–104. doi:10.1177/0278364907073777
- [7] Roderick, S., Roberts, B., Atkins, E., and Akin, D., "The Ranger Robotic Satellite Servicer and Its Autonomous Software-Based Safety System," *IEEE Intelligent Systems*, Vol. 19, No. 5, Sept.–Oct. 2004, pp. 12–19. doi:10.1109/MIS.2004.53
- [8] Haidegger, T., "Advanced Robotic Arms In Space," *55th International Astronautical Congress*, International Astronautical Federation, Vancouver, Canada, 2004, pp. 1–10; also Paper IAC-04-IAA.3.6.2.04.
- [9] Coleshilla, E., Oshinowa, L., Rembalaa, R., Binaa, B., Reyb, D., and Sindelar, S., "Dextre: Improving Maintenance Operations on the International Space Station," *Acta Astronautica*, Vol. 64, Nos. 9–10, May–June 2009, pp. 869–874. doi:10.1016/j.actaastro.2008.11.011
- [10] "Dextre's Final Exam Scheduled for December 22–23, 2010," NASA, [http://www.nasa.gov/mission\\_pages/station/structure/dextre\\_final\\_exam.html](http://www.nasa.gov/mission_pages/station/structure/dextre_final_exam.html) [retrieved 31 March 2011].
- [11] "Dextre Successfully Completes Its First Official Job," NASA, [http://www.nasa.gov/mission\\_pages/station/expeditions/expedition26/dextre\\_firstjob.html](http://www.nasa.gov/mission_pages/station/expeditions/expedition26/dextre_firstjob.html) [retrieved 31 March 2011].
- [12] Harding, P., "Dextre and RRM Complete Record Breaking Week of Robotics on ISS," <http://www.nasaspaceflight.com/2012/03/dextre-rrm-completerecord-breaking-week-robotics-iss/> [retrieved 26 June 2012].
- [13] Harding, P., "ISS: Dextre and RRM Complete Second Round of Joint Ops—CDRA Recovered," <http://www.nasaspaceflight.com/2012/06/iss-dextre-rrmcomplete-second-round-joint-ops-cdra-recovered/> [retrieved 26 June 2012].
- [14] Diftler, M. A., Culbert, C. J., Ambrose, R. O., Platt, R. J., and Bluethmann, W. J., "Evolution of the NASA/DARPA Robonaut Control System," *Proceedings of the IEEE International Conference on Robotics and Automation*, Vol. 2, IEEE Publ., Piscataway, NJ, 2003, pp. 2543–2548.
- [15] Jha, A., "Meet Robonaut 2, Astronaut Assistant," <http://www.guardian.co.uk/science/2010/nov/02/robonaut-2-international-space-station> [retrieved 31 March 2011].
- [16] Diftler, M. A., Mehling, J. S., Abdallah, M. E., Radford, N. A., Bridgwater, L. B., and Sanders, A. M. et al., "Robonaut 2—The First Humanoid Robot in Space," *2011 IEEE International Conference on Robotics and Automation*, IEEE Publ., Piscataway, NJ, May 2011.
- [17] Pappas, S., "Humanoid Robot Hitching Space Ride on Shuttle Discovery," <http://www.space.com/9384-humanoid-robot-hitching-space-ride-shuttlediscovery.html> [retrieved 31 March 2011].
- [18] Choi, C. Q., "New Robot Could Aid Astronauts in Space," <http://www.space.com/7871-robot-aid-astronauts-space.html> [retrieved 31 March 2011].
- [19] Haddadin, S., Albu-Schaffer, A., Frommberger, M., Rossmann, J., and Hirzinger, G., "The 'DLR Crash Report': Towards a Standard Crash-Testing Protocol for Robot Safety—Part 1: Results," *IEEE International Conference on Robotics and Automation*, IEEE Publ., Piscataway, NJ, 2009, pp. 272–279.
- [20] Haddadin, S., Albu-Schaffer, A., Frommberger, M., Rossmann, J., and Hirzinger, G., "The 'DLR Crash Report': Towards a Standard Crash-Testing Protocol for Robot Safety—Part 2: Discussions," *IEEE International Conference on Robotics and Automation*, IEEE Publ., Piscataway, NJ, 2009, pp. 280–287.
- [21] Sisbot, E., Clodic, A., Alami, R., and Ransan, M., "Supervision and Motion Planning for a Mobile Manipulator Interacting with Humans," *Proceedings of the 3rd ACM/IEEE International Conference on Human Robot Interaction*, ACM, New York, NY, 2008, pp. 327–334.
- [22] "American National Standard for Industrial Robots and Robot Systems—Safety Requirements," Robotic Industries Association, American National Standards Institute, Rept. R15.06-1999, 1999.
- [23] "Robots for Industrial Environment—Safety Requirements—Part 1—Robot," American National Standards Institute, International Standard Organization, Rept. 10218-1:2007, 2007.
- [24] Schuster, G., and Winrich, M., "Robotics Safety," Rockwell Automation Paper WP009A-EN-P, Dec. 2009.
- [25] Gerkey, B. P., and Mataric, M. J., "A Formal Analysis and Taxonomy of Task Allocation in Multi-Robot Systems," *International Journal of Robotics Research*, Vol. 23, No. 9, Sept. 2004, pp. 939–954. doi:10.1177/0278364904045564

- [26] Carpin, S., and Parker, L. E., "Cooperative Leader Following in a Distributed Multi-Robot System," *Proceedings of IEEE International Conference on Robotics and Automation*, IEEE Publ., Piscataway, NJ, 2002, pp. 2994–3001.
- [27] Van Der Krogt, R., and De Weerd, M., "Plan Repair as an Extension of Planning," *Proceedings of the 15th International Conference on Automated Planning and Scheduling (ICAPS-05)*, AAAI Press, Menlo Park, CA, 2005, pp. 161–170.
- [28] Greenstein, J. S., and Revesman, M. E., "Two Simulation Studies Investigating Means of Human–Computer Communication for Dynamic Task Allocation," *IEEE Transactions on Man and Cybernetics Systems*, Vol. 16, No. 5, Sept. 1986, pp. 726–730. doi:10.1109/TSMC.1986.289317
- [29] Breazeal, C., Kidd, C. D., Thomaz, A. L., Hoffman, G., and Berlin, M., "Effects of Nonverbal Communication on Efficiency and Robustness in Human–Robot Teamwork," *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'05)*, IEEE Publ., Piscataway, NJ, 2005, pp. 708–713.
- [30] Kaupp, T., Makarenko, A., and Durrant-Whyte, H., "Human–Robot Communication for Collaborative Decision Making—A Probabilistic Approach," *Robotics and Autonomous Systems*, Vol. 58, No. 5, May 2010, pp. 444–456. doi:10.1016/j.robot.2010.02.003
- [31] Sheridan, T. B., *Telerobotics, Automation, and Human Supervisory Control*, MIT Press, Cambridge, MA, 1992, pp. 1, 41–65, 301–321.
- [32] Ikeura, R., Inooka, H., and Mizutani, K., "Subjective Evaluation for Maneuverability of a Robot Cooperating with Human," *Proceedings of the 1999 IEEE International Workshop on Robot and Human Interaction*, IEEE Publ., Piscataway, NJ, 1999, pp. 201–205.
- [33] Kato, R., and Arai, T., "Assessment of Mental Stress on Human Operators Induced by the Assembly Support in a Robot-Assisted 'Cellular Manufacturing' Assembly System," *International Journal of Automation Technology*, Vol. 3, No. 5, 2009, pp. 569–579.
- [34] Steinfeld, A., Fong, T., Kaber, D., Lewis, M., Scholtz, J., Scholtz, A., and Goodrich, M., "Common Metrics for Human–Robot Interaction," *Proceedings of the 1st ACM SIGCHI/SIGART Conference on Human–Robot Interaction*, ACM, New York, NY, 2006, pp. 33–40.
- [35] Gavril, D. M., "The Visual Analysis of Human Movement: A Survey," *Computer Vision and Image Understanding*, Vol. 73, No. 1, Jan. 1999, pp. 82–98. doi:10.1006/cviu.1998.0716
- [36] Takeda, T., Hirata, Y., and Kosuge, K., "Dance Step Estimation Method Based on HMM for Dance Partner Robot," *IEEE Transactions on Industrial Electronics*, Vol. 54, No. 2, 2007, pp. 699–706. doi:10.1109/TIE.2007.891642
- [37] Anh, M., Ho, T., Yamada, Y., and Umetani, Y., "An Adaptive Visual Attentive Tracker for Human Communicational Behaviors Using HMM-Based TD Learning with New State Distinction Capability," *IEEE Transactions on Robotics and Automation*, Vol. 21, No. 3, 2005, pp. 497–504. doi:10.1109/TRO.2004.840912
- [38] McGhan, C. L. R., and Atkins, E. M., "A Low-Cost Manipulator for Space Research and Undergraduate Engineering Education," *AIAA Infotech@Aerospace 2010*, AIAA Paper 2010-3394, April 2010.
- [39] Craig, J. J., *Introduction to Robotics: Mechanics and Control*, 3rd ed., Pearson Education, Upper Saddle River, NJ, 2005, pp. 28, 35–36, 68–69.
- [40] Nickels, K., "Hand-Eye Calibration of Robonaut," *San Antonio Chapter IEEE Computer Society Past Meeting Presentations*, 16 September 2004, [http://www.ieee-cs-cts.org/past\\_meetings.htm](http://www.ieee-cs-cts.org/past_meetings.htm) [retrieved 2 Jan. 2011].
- [41] Rekimoto, J., "SmartSkin: An Infrastructure for Freehand Manipulation on Interactive Surfaces," *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM, New York, NY, 2002, pp. 113–120.
- [42] Yarrow, K., Brown, P., and Krakauer, J. W., "Inside the Brain of an Elite Athlete: The Neural Processes That Support High Achievement in Sports," *Nature Reviews Neuroscience*, Vol. 10, No. 8, Jan. 2009, pp. 585–596. doi:10.1038/nrn2672
- [43] Box, G., and Hunter, W., *Statistics for Experimenters*, Wiley-Interscience, New York, 2005, pp. 1–132, 363–488.
- [44] Montgomery, D., *Design and Analysis of Experiments*, Wiley, Hoboken, NJ, 2008, pp. 1–106.
- [45] "NASATLX: Task Load Index," NASA, <http://humansystems.arc.nasa.gov/groups/TLX/> [retrieved 2 Jan. 2011].

A. Srivastava  
Associate Editor