

Comparative Analysis of Human and Teleoperated Robotic Performance in Space Utility Operations

Bence P. Szego¹ and Dr. David Akin²

University of Maryland Space Systems Laboratory, College Park, Maryland, 20742

Human participation in space exploration introduces significant cost, risk, and engineering complexity, motivating increased reliance on robotic systems for many operational tasks. However, the degree to which robotic manipulation can effectively substitute for direct human interaction remains an open human factors question. This study evaluates performance differences using a Fitts' Law button-tapping task conducted under four distinct experimental conditions including direct human operation in a shirtsleeve environment, human manipulation using a simulated hardshell spacesuit arm, robotic manipulation with direct visual feedback of the worksite, and robotic manipulation relying on camera-based teleoperation. Following each set of trials, participants completed subjective workload assessments including the NASA Task Load Index (TLX) and the Bedford workload scale. Task performance was further quantified using Fitts' Law, taking throughput and movement time as functions of an index of difficulty, providing a consistent measure of manipulation efficiency across conditions. Results indicate that shirtsleeve and simulated suit-arm conditions produce significantly higher efficiency and lower cognitive workload than robotic operation, whereas robotic conditions reduce perceived physical effort compared to the suited condition but introduce far greater cognitive workload. These findings suggest that direct human interaction and robotic manipulation fulfill complementary roles rather than serving as direct replacements for one another, and that future missions may benefit by integrating both approaches to balance efficiency, operator workload, and operational flexibility.

I. Nomenclature

MT	= Fitts variable, mean task completion time (s)
D	= Fitts variable, distance from starting point to center of target zone (m)
W	= Fitts variable, diameter of target zone (m)
a	= Intercept of linear regression of Fitts data, corresponding to reaction time
b	= Slope of linear regression of Fitts data, corresponding to task difficulty

II. Introduction

Human involvement in space exploration is inherently dangerous, and presents risks including injury and death which make the enterprise difficult to justify. A significant body of engineering work relating to the exploration of space, then, is consumed in designing and implementing safety systems and failsafes to ensure the health and survival of the crew. Even with all the work and cost, accidents leading to injury and death still occur. In order to combat these risks, as well as reduce the costs associated with conducting science in space, many space agencies the world over have moved to primarily robotic means for data collection, analysis, and communication recently. As a result, the vast majority of current exploration missions are conducted entirely without human crews. These robotic missions are capable of performing increasingly complex scientific tasks, operating autonomously in extreme environments, and remaining active for years or even decades without the need for life-support. Their success has

¹ Undergraduate Researcher, Aerospace Engineering, 4298 Campus Drive, College Park, MD 20742, AIAA University Student Member 1925774

² Director, Space Systems Lab, Aerospace Engineering, 382 Technology Drive, College Park, MD 20742, AIAA University Educator Member

highlighted the economic and practical advantages of uncrewed exploration, particularly when compared to the significant cost, mass, and risk associated with human spaceflight. Consequently, the prevalence of robotic missions has led to debate over the necessity of sending humans into space at all. As robotic systems continue to advance in capability and reliability, the justification for exposing astronauts to substantial danger is increasingly questioned, prompting broader consideration of whether future space exploration should prioritize human presence or rely primarily on autonomous or remotely operated systems.

This study will use a combination of four trials to establish the necessity of human presence in space in order to maintain an acceptable degree of productivity and reduce the difficulty of performing tasks. As most space exploration involves inherently novel or unstructured tasks, little of the work can be entirely automated, instead relying on human operators guiding remote systems from afar. The additional constraints imposed by physical separation, limited sensory feedback, and interface mediation may significantly degrade performance even in basic tasks. To quantify these effects, this study will investigate task efficiency using Fitts' law. Comparing performance across trials where subjects are directly involved in tasks and trials where subjects operate a robotic system remotely enables a systematic, human factors driven assessment of how decreasing levels of direct human involvement impact task execution, and establishes a comparative framework for evaluating when direct human presence provides a meaningful operational advantage over teleoperated robotic systems. The results are intended to inform trade studies in mission architecture and human-robot interface development by identifying performance thresholds beyond which the productivity benefits of crewed operations may or may not justify their associated risks and costs. Fitts' law performance will be quantified using throughput and movement time as functions of index of difficulty, providing a well-established, task-independent measure of human and teleoperated manipulation efficiency.

In addition to the objective performance metric provided by Fitts' law, this study will quantify perceived workload and effort associated with each trial condition using established human-factors assessment tools. Subjective workload will be evaluated using the NASA Task Load Index (NASA-TLX), capturing mental demand, physical demand, temporal demand, performance, effort, and frustration. Workload and the attention requirement associated with completing the task will further be assessed using the Bedford scale. Together, these instruments enable comparison of not only task completion efficiency, but also the cognitive and physical burden imposed on the operator across the tested conditions, offering a more complete assessment of the human cost of reduced presence and mediated control in space operations.

III. Background

A. Space Utility Vehicle Concept

The Space Systems Lab at the University of Maryland, College Park, has been at the forefront of research into the space utility vehicle (SUV) concept, which is imagined as an alternative to EVA for orbital or microgravity operations. The general concept proposes a small maneuverable platform that allows an operator to perform inspection, assembly, maintenance, or scientific work outside a spacecraft without requiring a traditional extravehicular activity. By relocating the operator into a dedicated work vehicle equipped with propulsion, robotic manipulators, and integrated control interfaces, the SUV seeks to improve astronaut safety and operational efficiency while reducing the physical constraints imposed by conventional spacesuits. In principle, such systems allow operators to maintain close proximity to the worksite while benefiting from enhanced mobility, improved dexterity through robotic arms, and a more controlled operating environment [1]. As a result, the SUV concept has been proposed as a potential solution for a wide range of orbital servicing and construction tasks, particularly as interest grows in on-orbit assembly, satellite servicing, and long-duration infrastructure in Earth orbit.

Despite broad agreement on the potential advantages of the SUV concept, there remains significant disagreement regarding the appropriate degree of human involvement in these systems. Some proposed architectures emphasize direct human interaction with the worksite, prioritizing visibility and direct, hands-on manipulation in order to maximize dexterity and situational awareness. Other designs instead favor robotic or teleoperated systems in which the human operator remains physically separated from the work environment, relying on cameras, displays, and robotic control interfaces to perform tasks. This approach can reduce risk to crew members and limit the need for life-support systems in exposed environments, but may also introduce performance penalties associated with reduced sensory feedback, interface mediation, and limited depth perception. Determining how these competing design philosophies affect operator performance is therefore an important human factors question. Evaluating the relative efficiency, workload, and controllability of direct human manipulation versus robotic operation provides insight into how SUVs and related space systems should balance human presence with robotic capability in order to achieve the most effective operational design.

One concrete instantiation of the SUV concept is the SCOUT single-person spacecraft (SPS) design developed at the University of Maryland. SCOUT integrates both human and robotic manipulation capabilities,

allowing the operator to interact with the worksite directly through human hands while also controlling robotic arms for extended reach or precision tasks. Crucially, the design emphasizes direct observation, giving the operator immediate visual feedback of the task environment, which complements the proposed mechanical advantages provided by both the suit-constrained human arms and the robotic manipulators [2]. Understanding the comparative advantages and limitations of these different interaction systems is essential for informing future SUV design. Specifically, quantifying when human dexterity or direct vision provides a measurable benefit, versus when robotic arms suffice, can guide decisions about which capabilities are necessary for a given mission. This knowledge could also clarify whether the inclusion of both human and robotic manipulation systems is operationally justified, or whether one modality alone could achieve mission objectives efficiently and safely, ultimately supporting more effective and risk-conscious SUV architectures.

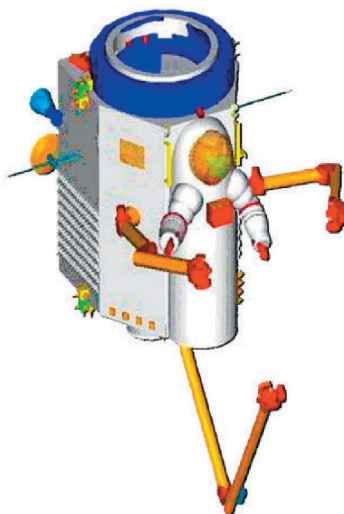


Fig. 1: SCOUT Single-Person Spacecraft Render [2]

B. Review of Existing Literature

A past 2023 study conducted by a team at the University of Maryland Space Systems Laboratory investigated similar questions regarding operator performance in direct-contact and robotic task environments using a Fitts' Law tapping experiment. That work compared five operational conditions: shirtsleeve interaction, conventional spacesuit arms, SPS-style suit arms, robotic manipulation with direct vision of the worksite, and robotic manipulation through teleoperation using camera displays. Performance was evaluated using movement time and Fitts' Law regression, and subjective handling qualities were assessed using the Cooper-Harper rating scale. The study found that all hands-on conditions produced significantly faster task completion times than the robotic control environments, while direct visual control of the robot outperformed teleoperation through video displays. However, the analysis reported relatively weak linear correlations in several of the fitted Fitts' Law relationships, resulting in low R^2 values that limited the strength of the performance modeling. Additionally, the use of the Cooper-Harper scale focused primarily on handling quality rather than operator workload, leaving the cognitive and physical demands associated with each operating condition largely unexplored [3].

The present study builds on this previous work while addressing several of its limitations. The task board, which was not discussed in previous work, has been designed to offer significant variation in the distances between buttons, which was cited as a shortcoming in the previous work, and may offer an improvement to the fitting of the Fitts' law regression. Further, a different simulated spacesuit arm configuration is used to more clearly represent constrained human manipulation. In addition, workload is evaluated once again using the NASA-TLX to identify any gaps in the previous data, as well as newly using the Bedford Workload Scale rather than relying on Cooper-Harper ratings, allowing the experiment to quantify the cognitive and physical effort required to complete the task under each condition. While prior work primarily focused on ranking the relative performance of each operating mode, the goal of this research is to expand the analysis toward defining the operational roles best suited to each configuration. By combining objective Fitts' Law performance metrics with workload measurements, this study seeks to identify how direct human interaction, constrained human operation, and robotic manipulation may be most effectively integrated in future SUV architectures rather than treated simply as competing alternatives.

In essence, this study seeks to extend the work by this team and determine

IV. Materials and Methods

A. Chosen Measurement Methods

In order to effectively analyze the throughput and perceived strain of robotic- and human-input task completion, first, a standard task had to be designed which could be completed by both humans and robots. Further, this task had to provide data which would provide both throughput and strain information.

To measure throughput, and provide a strict quantitative scale through which task difficulty could be measured, analysis using Fitts' Law was performed. Fitts' law specifically measures the throughput involved with moving from a fixed point to an area of arbitrary size, and is modelled with Eq. (1) below.

$$MT = a + b * \log_2\left(\frac{2D}{W}\right) \quad (1)$$

Here, MT refers to time to complete the movement, D refers to distance from the starting point to the center of the target, and W refers to the diameter of the target. In addition to these known variables, which can be recovered from testing, a and b are fitted as a linear regression to the data of a task, which define the input delay and acceleration, respectively. The quantity of $\frac{1}{b}$ can then be interpreted as an index of performance, or "information processing capacity" of a system, with greater quantities, or equally, lesser quantities of b itself, representing higher performance [4].

The final requirement, of returning strain information, was simpler to design experimentation around, considering the wealth of survey systems which exist to provide quantitative data in human factors testing, especially relating to perceived effort, workload, and performance. For this study, the NASA-TLX and the Bedford Scale were chosen due to their widespread usage in industry and applicability, as well as their ability to, when combined, measure both individual components of overall effort, as well as workload independently.

The NASA-TLX is a subjective workload assessment tool developed at NASA Ames Research Center to evaluate the demands placed on a human operator while performing a task. It measures perceived workload across six dimensions: mental demand, physical demand, temporal demand, performance, effort, and frustration, which together provide an overall indication of task difficulty [5]. Because it captures multiple aspects of workload, NASA-TLX is widely used in human-factors studies to evaluate human interaction with complex systems. In this study, NASA-TLX complements the objective performance metrics derived from Fitts' law by capturing the operator's perceived effort while completing each trial condition. This is particularly relevant when comparing direct human manipulation and robotic teleoperation, where physical effort may decrease but cognitive demand may increase due to mediated control and reduced sensory feedback.

The Bedford scale is a subjective workload assessment method commonly used in aviation and spaceflight human-factors research to evaluate the cognitive demand placed on an operator during a task. The scale measures an operator's spare mental capacity using a single numerical rating from 1 to 10, determined through a hierarchical decision process that guides the user to select the description that best matches their workload level. Lower ratings correspond to minimal workload and ample spare capacity, while higher ratings indicate increasing mental demand and reduced ability to manage additional tasks. NASA uses the Bedford scale as a standard method for verifying acceptable cognitive workload in crew interfaces, with typical tasks targeting ratings of 3 or less [6]. In this study, the Bedford scale complements the NASA-TLX by providing a simple overall measure of whether the cognitive demands of each experimental condition remain within acceptable limits for sustained operation.

Together, Fitts' law, the NASA-TLX, and the Bedford Workload Scale provide a comprehensive framework for evaluating task difficulty from both objective and subjective perspectives. Fitts' law quantifies task efficiency through measurable metrics such as movement time and throughput, capturing how effectively an operator can perform a motor task under different constraints. In contrast, NASA-TLX and the Bedford scale measure the human cost of that performance by assessing perceived physical and cognitive workload. By combining these tools, it becomes possible to determine not only which operational modes allow tasks to be completed most efficiently, but also which impose acceptable levels of effort and mental demand on the operator. This integrated approach enables clearer identification of the environments in which direct human manipulation, constrained human operation, or robotic control are most appropriate, even when a single measure of efficiency might dictate a different method, helping guide design decisions and mission architectures that balance productivity, safety, and operator workload in novel ways.

B. Chosen Task

In order for Fitts' law to provide accurate data on subject behavior and performance, a specific task analogous to moving to a target area of some arbitrary size must be chosen. For the purposes of this paper, the task was refined slightly to include tapping circular buttons. By using multiple buttons, the distance between start and end position could be easily defined and varied, and by connecting the buttons to a microcontroller, which reads the

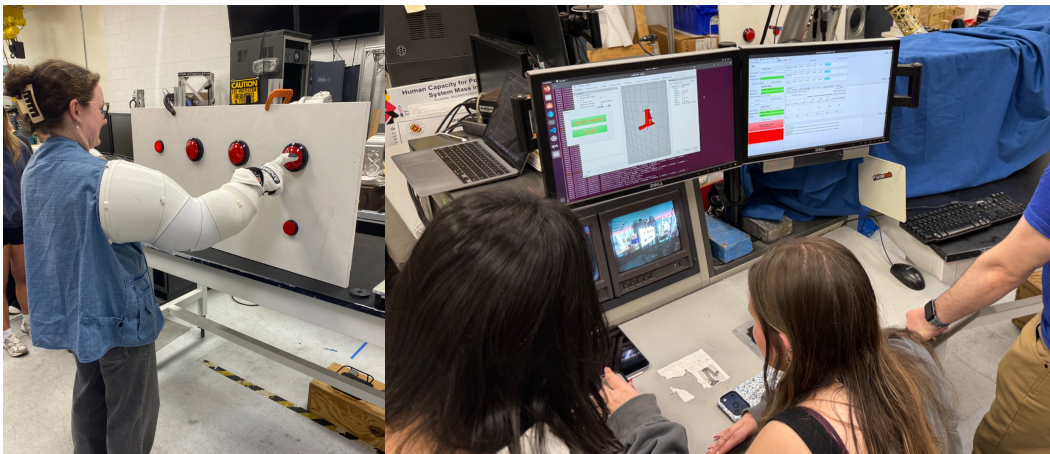
time the buttons are pressed, the time that it takes the subject to move between the buttons was also recovered. These features, along with the known, varying size of each button, allowed all three measured quantities in Fitts' law to be recovered. The button-pushing task also had the ancillary benefit of being easy to perform using a robotic arm, which was one of the primary requirements of the task selection. The act of pushing a button was quite simple, requiring rudimentary translational control of a fixed end effector, rather than a complex task involving rotation or manipulation of an object, which would be much more complex with a robotic system. The button pushing task was also simple to learn for subjects, reducing the concern of task performance being hampered by a lack of understanding or training on the details of a task.

C. Procedure

The experiment consisted of four sequential trials of a Fitts' law button-selection task designed to evaluate performance under varying operational constraints. In each trial, subjects were instructed to press illuminated buttons on a target panel as they activated one at a time in sequence. The task required participants to move between targets of varying spacing and size, allowing movement time and accuracy to be recorded for subsequent Fitts' law analysis. Upon pressing the illuminated button, the next target in the sequence would activate, continuing the task for the duration of the trial.

The trials were conducted in four different operational conditions intended to simulate increasing levels of constraint on the operator. The first condition was a shirtsleeve environment, in which subjects used their bare hands to interact directly with the button panel, serving as the baseline control condition. In the second condition, participants repeated the task using a simulated spacesuit arm apparatus designed to mimic proposed hard-shelled suits, with bearing joints allowing for mobility. The suit effectively restricted dexterity and increased mechanical impedance relative to natural human movement. In the third condition, subjects controlled a robotic arm to press the illuminated buttons while maintaining direct visual observation of the worksite, allowing them to gauge the positioning of the robot with their own eyes. In the fourth condition, the robotic manipulation task was repeated with the operator relying solely on camera-based video feeds for visual feedback, simulating teleoperated control without direct line-of-sight to the task environment.

After completing each trial condition, participants were asked to complete a brief assessment consisting of the NASA-TLX and the Bedford scale. These surveys were used to capture the participant's perceived workload associated with each operational mode before proceeding to the next trial.



Figs. 2, 3: Subjects completing testing in suit arm, using robotic arm with teleoperation

D. Materials for Task

The task field which the button pushing task was performed on includes eight buttons, four each of two varying sizes, measuring with a diameter of 20 cm or 12 cm. This allowed for variety in the quantity W in the Fitts' law relation. Further, the buttons were positioned every 25.4 centimeters in a grid pattern of two rows and four columns. The positioning of these buttons around the board allowed for a large variety of distances D to be encountered during a trial, depending on the start and end button of a specific pushing sequence. Within the top row, from right to left, the button sizes were small, large, large, large, while in the bottom row, again from right to left, the button sizes were large, small, small, small. This task board was kept constant, and used for every trial, no matter the condition of the subject, whether hands-on or robotic interaction. In addition to the task board, for the suited condition, subjects

were fitted with a simulated space suit arm, printed in plastic. The arm is made up of a series of bearing joints, which allow the hard-shell structure to rotate and accommodate the movement of the subject. In addition to the hard shell arm, a glove was used as well to model the decreased sensory feedback and dexterity that wearing a pressurized suit would cause. The glove and arm were attached to a vest to allow the subjects to comfortably wear the system while completing the button pushing task. The task board used for all the trials as well as the suit arm are shown below.

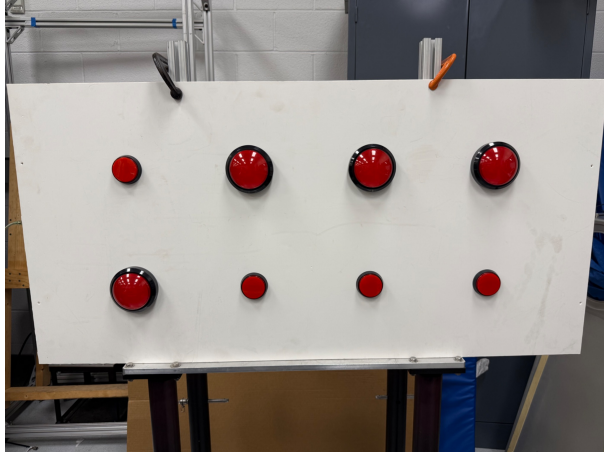


Fig. 4: Button task board.



Fig. 5: Simulated space suit arm used in testing.

For robotic trials, a six degree of freedom robotic arm, originally designed by the University of Maryland Space Systems Lab as the camera operation arm for the Ranger neutral buoyancy vehicle, was used. This arm was controlled using a three degree of freedom joystick, which controlled translation in x-y-z, as well as a secondary joystick providing roll, pitch, and yaw control [7]. In addition to this, for trials without direct vision, two cameras were used, one mounted to the end effector of the robotic arm which monitored the contact point of the arm, as well as a camera mounted to the side of the system offering a view of the workspace. This allowed participants to both see the exact position of the end effector with reference to the button, as well as a view of the entire board, ensuring the next lit button could always be located. The Ranger arm and control station are pictured below.

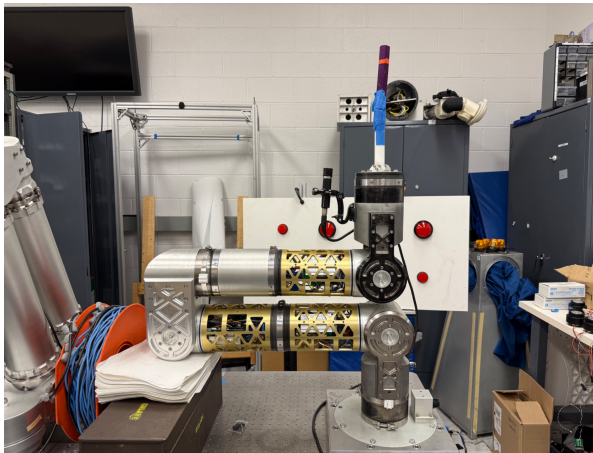


Fig. 6: Robotic arm used in testing.



Fig. 7: Control interface with video monitors.

V. Results

A. Average Fitts' Law Results, Tabulated

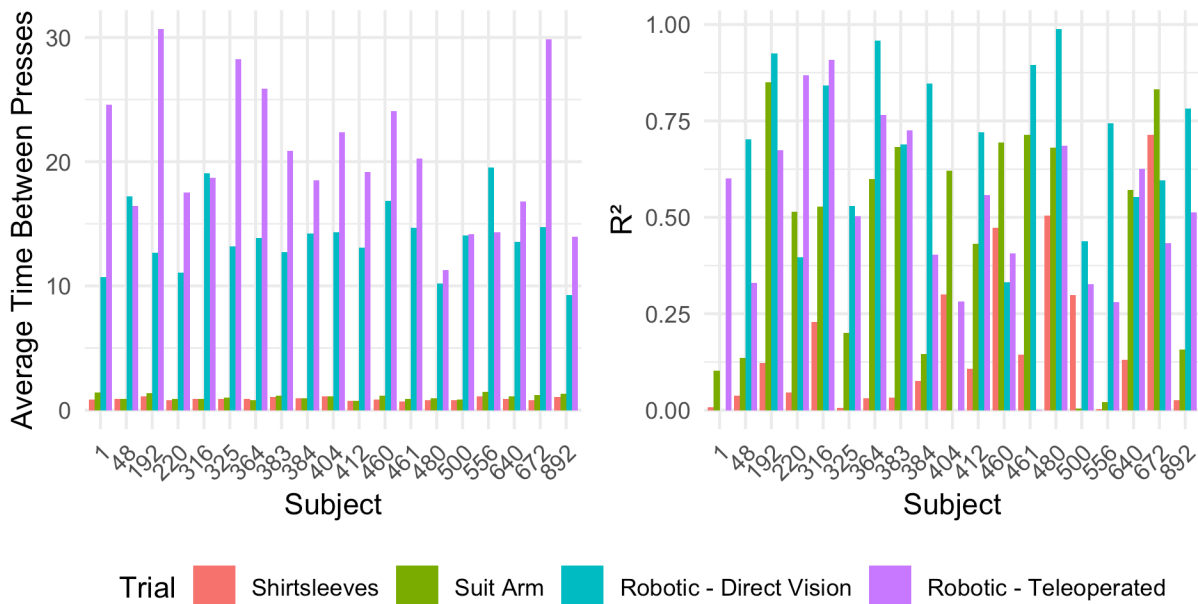
The measured task completion times varied substantially across the four operating conditions. In the shirtsleeve condition, the average time between presses was 0.921, with average regression parameters $a = 0.760$ and $b = 0.068$, producing a relatively weak fit on average, with $R^2 = 0.173$. Performance in the suited condition was slightly slower, with an average press time of 1.067, and average regression parameters $a = 0.708$ and $b = 0.152$, with a moderate improvement in average fit with $R^2 = 0.447$. The direct-vision robotic condition resulted in much longer

press times, averaging 13.95, and yielded average parameters $a = 1.897$ and $b = 5.424$, with the strongest relationship on average among the four cases with $R^2 = 0.628$. The teleoperated robotic condition produced the longest average press time at 20.41 seconds, with average parameters $a = 1.230$ and $b = 7.753$ and a moderately strong fit on average with $R^2 = 0.521$. Overall, robotic operation resulted in substantially longer task times compared with human-only conditions, though the regression models saw a larger amount of variance in those cases. Individual regression data for each subject can be seen in the graphs below.

Trial Condition	Average time between press	a	b	R^2	$\frac{1}{b}$	% of control $\frac{1}{b}$
Shirtsleeve	0.921	0.760	0.068	0.173	14.706	100.00%
Suited	1.067	0.708	0.152	0.447	6.588	44.80%
Direct Vision Robotic	13.953	1.897	5.424	0.628	0.184	1.25%
Teleoperated Robotic	20.405	1.230	7.753	0.521	0.129	0.88%

Table 1: Average Fitts' regression results across trials.

B. Fitts' Law Regression Results, Graphs



Figs. 8, 9: Average Time between Presses, Regression R²

Regression Lines by Trial

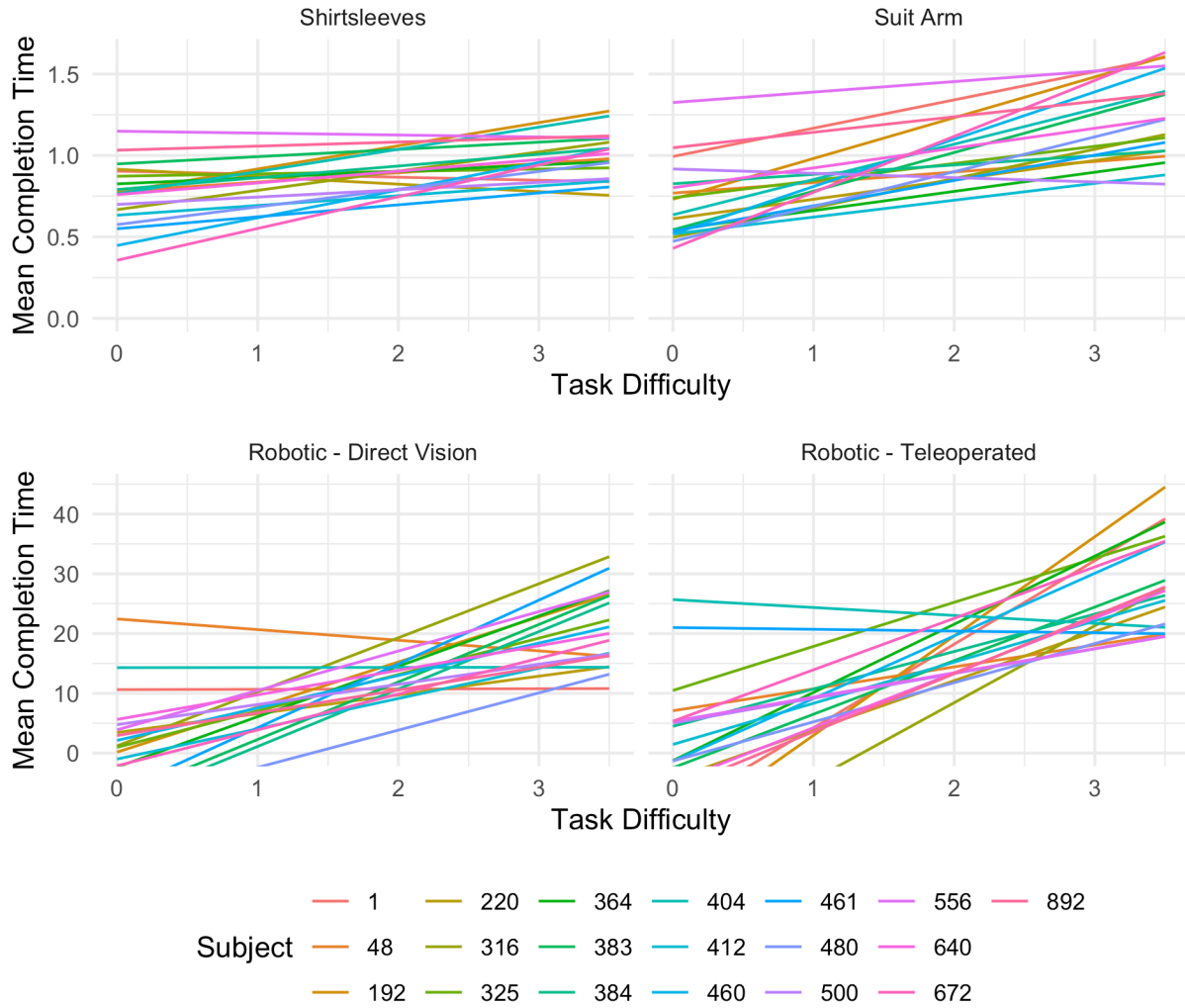


Fig. 10: Regression Lines by Trial, Subject

C. NASA-TLX and Bedford Scale Results

Trial Condition	Mental Demand (out of 7, 1 is lowest)	Physical Demand (out of 7, 1 is lowest)	Temporal Demand (out of 7, 1 is lowest)	Performance (out of 7, 1 is success)	Effort (out of 7, 1 is lowest)	Frustration (out of 7, 1 is lowest)	Bedford Scale Rating (out of 10)
Shirtsleeve	1.68	1.68	2.47	1.21	1.68	1.11	1.84
Suited	2.50	3.33	2.17	1.94	2.94	2.28	3.11
Direct Vision Robotic	4.22	2.50	2.78	2.06	3.94	3.00	4.33
Teleoperated Robotic	4.11	2.11	2.61	2.61	4.00	2.89	4.50

Table 2: Average survey results across trials.

VI. Discussion

A. Task Difficulty via Fitts' Law

In the context of Fitts' Law, the coefficient $\frac{1}{b}$ can be interpreted as an indicator of throughput or information-processing capacity, describing how efficiently a user performs aimed movements as task difficulty increases. Larger values of $\frac{1}{b}$ indicate higher throughput and therefore lower comparative difficulty, while smaller values indicate that performance degrades more strongly as the index of difficulty increases. Given that the task itself was unchanged in each trial, a lower value of $\frac{1}{b}$ than the control value from the shirtsleeves condition would indicate that the operating condition imposed additional difficulty onto the operator to complete the task. For the shirtsleeve condition, $\frac{1}{b} = 14.71$, indicating the highest throughput and lowest relative difficulty among the tested conditions. The suited condition yields $\frac{1}{b} = 6.59$, reflecting a reduction of performance by 65.2%, likely associated with mobility constraints. In contrast, the direct-vision robotic and teleoperated robotic conditions produce substantially lower values of $\frac{1}{b} = 0.18$ and $\frac{1}{b} = 0.13$ respectively, indicating much lower throughput and therefore greater operational difficulty.

As expected, task performance degraded progressively as operational constraints increased, with direct human operation in shirtsleeves providing a control baseline for both efficiency and perceived workload. Each subsequent condition introduced additional difficulty: the simulated spacesuit arm reduced throughput and increased movement time due to mechanical impedance and limited dexterity, while robotic manipulation further degraded performance as a result of interface mediation, limited system speed, and reduced sensory feedback. These results indicate that although wearing a suit introduces a measurable performance penalty, robotic operation imposes a substantially larger reduction in efficiency due to the increased difficulty of performing the button-pressing task.

It is notable that the difficulty penalty associated with robotic operation under direct vision and teleoperation are actually quite similar, their reduction from control processing capacity differing by only 0.37%. While this is not negligible, it is still far less than the difference in performance even between any of the other conditions. This could indicate that the source of visual information, be that cameras or direct eyesight, does not present as significant of a decrease in performance as the presence of tactile feedback from a subject's hands. Effectively, the most significant performance penalty is imposed once the subject must use the robot to complete the task, with additional constraints of worsened visual feedback not adding too much additional difficulty.

B. R^2 and Likely Causes of Variance

The coefficient of determination, R^2 , provides a measure of how well the linear model predicted by Fitts' Law explains the relationship between movement time and index of difficulty in the collected data. The average R^2 values observed here range from 0.173 in the shirtsleeve condition to 0.628 in the direct-vision robotic condition, indicating varying strengths of linear correlation across the tested scenarios. The relatively low R^2 values in some cases may be partially attributable to the experimental structure, in which multiple times could be gathered against a single level of Fitts difficulty, introducing multiple data points at the same index of difficulty which had to be accounted for in the fit. This ensures that there will be points which aren't able to be fitted to the regression line, lowering R^2 . Despite this scatter, the overall trends remain meaningful because the relationship described by Fitts' Law is well established in human motor control literature. Furthermore, many individual subsets of the data, particularly those with more variance of index of difficulty, produced fits with R^2 values well above 0.7, with some reaching over 0.9, while maintaining regression coefficients in line with the other trials. This suggests that while variability across repeated indices of difficulty reduces the fit on aggregate, the underlying relationship between task difficulty and movement time is still clearly present in the experimental results.

C. NASA-TLX and Bedford Scale Differences

The NASA TLX results show a clear shift in the type of workload experienced as the task environment changed. The shirtsleeve condition produced the lowest ratings across nearly all categories, reflecting the relative ease of direct human manipulation. When the simulated suit was introduced, physical demand increased substantially, with an average score of 3.33 compared with 1.68 in shirtsleeves, along with increases in perceived effort and insecurity, indicating the mechanical and dexterity limitations imposed by the suit. In contrast, both robotic conditions showed significantly lower physical demand than the suited case, with an average score of 2.50 for direct vision and 2.11 for teleoperation, but much higher mental demand, with ratings above 4 in both cases. This suggests that while the robot reduced the physical burden of the task, it shifted the workload toward cognitive processing and control of the

interface. Notably, temporal demand remained relatively constant across all conditions, averaging between 2.2 and 2.8, even though the robotic tasks took substantially longer to complete. This indicates that subjects did not feel significantly more hurried during the robotic trials, suggesting that while their objective performance slowed, their perceived pace of work remained similar.

The Bedford workload ratings show a similar trend in perceived task difficulty, particularly for the robotic conditions. The shirtsleeve case produced the lowest Bedford score, with an average score of 1.84, indicating ample spare attentional capacity. The suited condition increased the average workload rating to 3.11, reflecting a noticeable reduction in spare capacity as the subject managed the additional mechanical constraints. Both robotic conditions produced substantially higher Bedford scores, with an average of 4.33 for direct-vision operation and 4.50 for teleoperation. These values are very similar to one another and both lie above the tier described as “Insufficient spare capacity for easy attention to additional tasks.” [6] This suggests that operating the robotic system required sustained attention and cognitive engagement regardless of whether the robot was controlled directly or through teleoperation. Together, these results indicate that robotic manipulation significantly increases attentional workload even while reducing the physical demands associated with the task.

D. Implications for Mission Design and Application of Conditions

The results have several implications for mission design, particularly when selecting between human extravehicular activity and robotic manipulation. Tasks that are cognitively demanding, such as scientific exploration, inspection, or in-situ research, may be poorly suited to robotic operation. The NASA-TLX data indicates that robotic control significantly increases mental demand, and the Bedford ratings suggest that operators have little spare cognitive capacity when performing even simple tasks using a robotic system. In mission scenarios where astronauts must carefully interpret visual information, make judgments quickly, or coordinate multiple tasks simultaneously, the added cognitive burden of controlling a robotic system could reduce effectiveness or increase the likelihood of error. Additionally, the Fitts’ Law results show that robotic manipulation substantially increases movement time compared with direct human operation. For tasks requiring speed, efficiency, or rapid repeated interactions, such as equipment manipulation or assembly, the lower task efficiency and slower times observed in the robotic conditions suggest that relying on robotic control would likely reduce operational productivity.

However, the results also indicate several mission scenarios where robotic manipulation could be advantageous. The NASA TLX results show that robotic operation significantly reduces physical demand when compared with working in a simulated spacesuit, suggesting that robotics could be particularly effective for monotonous, repetitive, or physically taxing activities where cognitive demands are relatively low. Examples might include routine inspection, transport of objects, or other simple manipulation tasks that do not require rapid execution or complex decision-making. Importantly, the similar performance and workload results between the direct-vision robotic and teleoperated conditions suggest that physical human presence near the robot may not be necessary to achieve comparable performance. This opens the possibility of assigning such tasks to teleoperators located safely inside a spacecraft or habitat, or potentially even at a remote control station, thereby reducing astronaut exposure to the risks associated with extravehicular activity while still benefiting from human supervision and adaptability.

As for applying these results to the SUV concept as a whole, a few things are clear. First and foremost, there is a clear and demonstrated benefit for the inclusion of people in space exploration, and within the design of microgravity spacecraft, including the SUV. The efficiency gains associated with human presence, specifically in direct human contact with tasks, significantly offset the additional cost and risks associated with human presence. Our findings indicate that significant productivity is lost when tasks are performed solely via robotic or teleoperated systems, with the greatest reductions occurring when robotic manipulators replace direct human contact. Interestingly, the difference in efficiency between direct robotic manipulation and teleoperation is relatively small, suggesting that for single-person spacecraft equipped only with robotic arms, task performance would approach that of fully remotely piloted systems, rather than matching the productivity achievable with direct human involvement. Hybrid designs like SCOUT, which integrate both human and robotic manipulators under direct human supervision, offer a practical solution to this tradeoff: they allow operators to maintain the high efficiency of direct contact when tasks demand dexterity and situational awareness, while providing the option to reduce physical demand through robotic operation when appropriate. These results provide concrete guidance for SUV design, indicating when human presence is operationally essential and when robotic augmentation can supplement or extend human capability without substantial efficiency loss.

VII. Conclusion

Overall, the results demonstrate clear tradeoffs between direct human manipulation and robotic operation in terms of speed, perceived difficulty, and cognitive demand. As expected, human operation in shirtsleeves provided the highest throughput and lowest workload overall, while the simulated space suit introduced increased physical

demand and reduced dexterity. Robotic manipulation greatly increased task time and mental workload and reduced efficiency significantly, but also reduced the physical effort required to complete the task. These findings have important implications for SUV design, as well as mission design in future human spaceflight programs. In particular, they may inform the development of operational concepts for missions associated with the Lunar Gateway and other future space stations, where human-robot interaction is expected to play a major role. In such environments, robotic manipulation may be valuable for routine or physically demanding tasks, while more time-intensive or cognitively complex work could remain better suited to direct human interaction.

Several areas of additional work could further strengthen these findings. First, future experiments should include a wider range of Fitts difficulty conditions in order to produce more robust regression fits and improve the correlation values observed in the present data. This could be accomplished by increasing the variability in button sizes, or by adding additional buttons to increase the variability in spacing. Collecting additional trials across more indices of difficulty would help establish stronger relationships between index of difficulty and movement time, reducing the variability that contributed to the relatively modest R^2 values in some cases. Second, the experimental framework could be expanded to include different types of representative spaceflight tasks beyond simple button pressing. For example, similar trials could be conducted using simulated orbital repair or assembly tasks, which would better capture the types of manipulation astronauts may perform during maintenance or construction activities in orbit. Conducting such experiments in a neutral buoyancy environment would be particularly valuable, as it would more closely replicate the dynamics of microgravity and allow for direct evaluation of robotic and human performance within an SUV operational concept. Finally, additional trials could be conducted with experienced operators of the robotic equipment, to better understand how the task efficiency of a practiced subject, more analogous to an astronaut, changes under different conditions. Together, these extensions would help refine the understanding of when robotic systems are most beneficial and how they can best complement human operators in future space missions.

Acknowledgments

It would be remiss of me not to express my deep and profound gratitude to the team at the University of Maryland Space Systems Laboratory, who have made this research possible from the first day onwards with their support, advice, and resources. Thank you specifically to Dr. David Akin for his guidance and support throughout this project, to Charlie Hanner, Daniil Gribok, and Romeo Perlstein, who provided their expertise and were each integral to the completion of this research. I would also like to thank all those who were so generous in donating their time to help me investigate and act as participants in this research. Hopefully you have learned something interesting as well, or at the very least gotten a cool photo with the robotic arm! And last, though certainly far from least, thank you finally to Emily Martinez, Levente Szego, Leo Szego, and Catherine Stock, without whose encouragement this project would not have gotten across the finish line.

References

- [1] D. Akin, "Investigation of the Effectiveness of Space Utility Vehicles for Extravehicular Operations," *ASCEND 2021*, American Institute of Aeronautics and Astronautics, Nov. 2021. <https://doi.org/10.2514/6.2021-4098>
- [2] D. Akin, and M. Bowden, "SCOUT: EVA Capabilities of the Space Construction and Orbital Utility Transport," *SAE Technical Paper 2004-01-2295*, 2004, <https://doi.org/10.4271/2004-01-2295>.
- [3] L. Allee-Press, N. Bindra, L. Cha, and W. Dacey, *ORBIT: Orbital Repairs By Innovative Technology*, Digital Repository at the University of Maryland, College Park, MD, 2023. <https://doi.org/10.13016/QMKZ-9FEC>
- [4] Stuart K. Card, William K. English, and Betty J. Burr, "Evaluation of Mouse, Rate-Controlled Isometric Joystick, Step Keys, and Text Keys for Text Selection on a CRT," *Ergonomics*, Vol. 21, No. 8, 1978, pp. 601-613.
- [5] National Aeronautics and Space Administration, "NASA Task Load Index (TLX)," Human Systems Integration Division, NASA Ames Research Center.
- [6] National Aeronautics and Space Administration, "Cognitive Workload Assessment Methods: Bedford Workload Scale," NASA-STD-3001 Technical Brief OCHMO-TB-032, NASA, Washington, DC.
- [7] University of Maryland Space Systems Laboratory, "Ranger Neutral Buoyancy Vehicle," Space Systems Laboratory Robotics Program, University of Maryland, College Park, MD.