Worth the Wait: Understanding How the Benefits of Performative Autonomy Depend on Communication Latency

Rafael Sousa Silva[1], Michelle Lieng[1], Emil Muly[1], and Tom Williams[1]

***Abstract*— Robots deployed in space exploration contexts need to efficiently communicate with both co-located and remote teammates to perform tasks and resolve points of uncertainty. In recent work, researchers have proposed *Performative Autonomy*, an autonomy design strategy for enabling language-capable robots in these contexts to enhance interactants' Situation Awareness. However, it is not yet clear how the efficacy of this autonomy design strategy might be impacted by the extreme latency that characterizes interplanetary communication. In this work, we thus present the results of the first study exploring the impact of interaction latency on the effectiveness of *Performative Autonomy*. Our results suggest that while *Performative Autonomy* exacerbates the increased task performance times required under high latency, this autonomy design strategy can be used without increasing cognitive load, even under substantial communication latency. Moreover, our results suggest that robots performing lower levels of autonomy were viewed as better teammates, and that this autonomy design strategy helped provide resilience to degradation to such perceptions that would otherwise be caused by increasing levels of latency. Overall, these results motivate further work within the new *Performative Autonomy* paradigm for both remote and proximal human-robot interactions, in both space-oriented and traditional, terrestrial, human-robot interaction domains.**

I. INTRODUCTION

Robots hold significant potential for space exploration, not only on planetary and lunar surfaces, but also within space stations and ships. On space stations like the International Space Station (ISS) and the proposed Lunar Orbital Platform-Gateway (LOP-G), robots are expected to perform inspection tasks (spot checks, surveys, change detection, and problem localization [1]). Performing these repetitive duties during crewed periods allow human astronauts to better allocate their time and resources, and allow for continual maintenance during uncrewed periods. The Astrobee robot [2], for instance, is designed to operate aboard the ISS to monitor (1) radiation and carbon dioxide levels, (2) ISS integrity, and (3) the locations of tools that might be needed by crewmates [1]. These robots can also be teleoperated or tasked by remote operators during both crewed and uncrewed periods to perform activities in the ship.

Robots deployed on current and future space stations must be capable of natural language interaction [3] to accept natural language commands, answer questions from crewmates, report critical information (e.g., about detected anomalies [1]) and ask for assistance. Moreover, natural and effective human-robot communication is especially important in space contexts due to the acute hazards that can arise from miscommunication and human loss of Situation Awareness (SA) in such conditions [4].

In recent work, Roy et al. [5] proposed *Performative Autonomy* (PA): an effective strategy in which robots ask questions to increase SA (rather than to gather information) without cognitively overloading interactants. But while this prior research has shown the value of PA for promoting SA and task performance without increasing cognitive load, it is unclear whether these benefits, in particular the effects on task performance and the lack of effects on workload, will hold across relevant Human-Robot Interaction (HRI) space contexts. In this work we consider whether the benefits (and lack of costs) will be maintained across contexts that vary in terms of *communication latency*.

While proximal HRI (e.g., interactions between humans and robots aboard the ISS) may not be subject to communication latency, high communication latency is a defining characteristic of *remote* HRI contexts (e.g., communications between planetary robots and ground control operators on earth). Bandwidth concerns already lead to loss-of-signal and dropout aboard the ISS [1], and these concerns are likely to become more acute as space missions begin to range farther from earth. Future space operations will occur in domains known to be bandwidth-limited (e.g., permanently shadowed lunar regions [6]). Moreover, future operations are expected to experience increasing levels of communication latency as missions move farther from earth, to the ISS, LOP-G, Moon, Mars, and beyond.

For robots using autonomy design strategies like PA to be successfully deployed in these future crewed or uncrewed contexts, several key questions need to be answered:

RQ1: How does *Performative Autonomy* impact the success of interactions with remote, language-capable robots?
RQ2: How does communication latency mediate the effectiveness of language-capable robots' use of *Performative Autonomy*?

In this work, we present the results of the first study exploring the impact of interaction latency on the efficacy of PA. In Section II we will provide additional research background and introduce our hypotheses. In Section III we will present the design of a human-subject study designed to test these hypotheses. In Section IV we will present the results of this experiment. Finally, in Section V we will discuss these results before concluding in Section VI.

[1]These authors are with the MIRRORLab, Colorado School of Mines, Golden, CO 80401, USA. rsousasilva@mines.edu, mylieng@mines.edu, emuly@mines.edu, twilliams@mines.edu

II. BACKGROUND

A. Language-Capable Robots in Space

Natural language communication is critical for coordination and task performance in robotic space deployments [3]. Furthermore, this need for communication is accentuated by the dangerous character of space exploration contexts. First, the high cost of error in this context makes it critical for robots to be able to ask for help, advice, and clarification when necessary. Second, astronauts and ground control operators need to maintain high SA to prevent catastrophic error [7]. Close calls like U.S. ISS IVA 23 demonstrate how gaps in SA can have deadly consequences in space contexts [4]. High SA is critical during both crewed periods (in which human and robot activities may have intricate interdependencies) and uncrewed periods (in which it may be more challenging for human teammates to maintain high SA). As such, it is critical for robots deployed in space contexts to keep human teammates in the loop of their activities, regardless of whether those teammates are proximal or remote. Critically, the ability for a robot to keep interactants in the loop is mediated by the robot's *level of autonomy* [8].

B. Levels of Autonomy

Researchers have proposed different taxonomies for computational levels of autonomy. For instance, [9] proposed ten categories that define ways in which tasks can be divided between a human and a computer. These levels of autonomy can be manipulated through approaches such as adaptive automation [8], in response to changes in factors like teammate workload [10], [11]. For example, an adaptive automation system might increase the robot's level of autonomy to help an overloaded human teammate [12]. However, as autonomy increases, the ability for robots to keep teammates situationally aware decreases [13], [14], [15].

This challenge is partly due to the relationship between levels of autonomy and the communication strategies that are natural at each level. As robot autonomy decreases, there is increasing need to make statements and to ask for permission or ask questions in general. Inspired by literature [16], [17] on the types of queries a robot can use under different levels of autonomy, [5] proposed *Performative Autonomy*: instead of increasing autonomy to decrease mental workload at the cost of SA, robots decrease autonomy to increase SA, at the potential cost of mental workload.

C. Performative Autonomy

PA is an autonomy design strategy in which robots periodically *perform* lower levels of autonomy than they are truly capable of to strategically increase SA. Because lower levels of autonomy are associated with asking for permission or asking for supervisor decisions, PA manifests as robots asking questions they do not (believe they) truly need the answers to. As an example, when a robot checking for noise sources in a spacecraft finds a source that can be fixed without human assistance, it might choose to ask for input before proceeding even though it does not believe it actually needs such input. By doing so, the robot momentarily distracts their teammate, but forces them to engage with the robot's task and establish awareness to provide a cogent response.

Roy et al. [5] demonstrated that robots can use PA in simulated space contexts to increase interactants' SA without noticeably increasing their cognitive load. This suggests substantial opportunity for the use of PA in safety-critical contexts like space exploration. However, it is unclear how several key dimensions of interaction might be affected by this strategy in remote versus proximal interaction contexts. Specifically, because language-based interaction is critical for both contexts, we argue that the success of PA as an autonomy design strategy must be evaluated in terms of factors like overall task efficiency and perceived quality of interaction, which may be degraded by a key characteristic of remote interaction: communication latency.

D. Latency

The proximity between a human and a robot can determine the type of interaction they can engage in. While in proximal interactions, humans and robots are co-located and can communicate face-to-face, remote interactions are separated in space and/or in time [18]. Many human-robot interactions in space contexts, such as those between robots and ground operators during uncrewed periods, are necessarily remote [19], [20], [21]. However, bandwidth limitations and the immense spatial distances at play in remote human-robot interactions in space contexts lead to temporal distances due to the time needed for communication to be relayed between interactants [22]. Moreover, the effects of this delay can only be partially mitigated through robot automation or through more effective networking [23], [24].

Communication latency is an obvious barrier to natural, effective communication. Because the time required for natural language interaction geometrically scales along with distance, as communication latency begins to accumulate, the time-related benefits of unnecessary natural language interaction are likely to decrease. Specifically, we would expect that as communication latency increases, PA will likely lead to worse task performance, but will likely be perceived more positively by interactants, and will likely manifest no detriments to cognitive load, since they were not previously observed in latency-free interactions.

E. Hypotheses

Put more formally, we propose the following hypotheses:

- **(H1)** As the PA level decreases, task completion time, perceived robot dependency, and teaming quality ratings will increase. Cognitive load will remain unaffected.
- **(H2)** As latency increases, the negative consequences associated with low PA strategies will be exacerbated.

III. METHODOLOGY

A. Experimental Design

To investigate our hypotheses, we conducted a within-subjects experiment in which each participant performed a series of three tasks. Two key independent variables (*Latency*

and *PA Strategy*) varied according to a Greco-Latin Square design. That is, the three tasks performed by the participant each had a different *Latency* level (*Low Latency* [0s], *Medium Latency* [5s], or *High Latency* [10s]) and in each of these tasks, the robot teammate used a *PA Strategy* based on different levels from [5]'s scale of dialogue autonomy (*Low PA*, which corresponded to level 1 of dialogue autonomy; *Medium PA*, which corresponded to level 2 of dialogue autonomy; or *High PA*, which corresponded to level 5 of dialogue autonomy). The Greco-Latin Square design simultaneously controlled for ordering effects for both variables.

B. Experimental Context

A remote interaction context mimicked key dimensions of interactions that could occur between a robot in the ISS and a human operator on Earth. In this experimental context (based on the task designed by [25]), the human participant's task is to complete a resource management game, in which their continued progress requires certain resources to have been previously collected by their robotic partner. This design serves as a metaphor for the unidirectional task reliance that categorizes remote HRI tasks.

In this game, the human explores a 10x10 board, with the goal of clearing all locations containing "resource stations". To clear a resource station, the participant must expend resources collected by their robot partner. At any point, the participant can see an estimate of the current reserves of each resource, and can inspect the robot partner to determine what resource they are collecting (at any point, the robot faced a folder whose color matched one of the game's four resource colors). The participant thus has a noisy estimate of the current resource state, but maintaining SA of key parts of that state (i.e., what resources the robot is collecting) require the interactant to intentionally observe the robot.

Our experimental context deviated in several ways from the original task on which it was based [25]. First, we did not allow participants to re-task the robot at any time to collect a new resource. Second, in our experiment, participants viewed their remotely located robot teammate through a video window that showed a live stream of the robot (a Misty II). Third, we manipulated *PA Strategy* by having the robot periodically re-evaluate its choices every 50 seconds, and if needed, engaged in different types of verbal interactions with the participant. In the *Low PA* condition, the robot asked participants to help them arbitrate between multiple options (e.g., "I was collecting red resources. Should I keep collecting red resources, or switch to orange resources, or pink resources?"). In the *Medium PA* condition, the robot asked participants for confirmation on a decision (e.g., "I was collecting red resources. Can I now switch to collecting orange resources?"). In the *High PA* condition, the robot merely stated its decision (e.g., "I was collecting blue resources. I am now going to switch to collecting pink resources.").

These messages were simultaneously displayed as text and spoken audibly on the participant's computer. The robot's decision in all cases was made as in [25], by considering the ratio of resources available to resources needed in future visible tasks. Communication was only performed if the resource with the lowest such ratio was not the resource already being collected. In the *Low PA* condition, the current resource and the two other resources with the lowest such ratio were listed in the robot's communication, while in the *Medium PA* and *High PA* conditions the current resource and the resource with the lowest such ratio were listed. Note that in all cases, the robot had the ability to act optimally, meaning that the *Low PA* and *Medium PA* behaviors truly were *Performances* of Autonomy.

Finally, we manipulated *Latency* by simulating a delay into the video stream of the robot. In the *Low Latency* condition, no delay was simulated. In the *Medium Latency* condition, a 5-second delay was simulated. In the *High Latency* condition, a 10-second delay was simulated. These delays, in practice, translate to pauses of three times the length associated with the condition. In a game with 10 seconds of latency, for example, the robot would take (a) 10 seconds to send a message to the player, (b) 10 seconds to process input (if any) and reposition itself, and (c) 10 additional seconds before it started collecting resources of the next color. Between (a) and (b) the robot also stayed idle for the amount of time taken by the participant to respond to the robot.

C. Measures

Our hypotheses' key dependent variables were task performance time, perceived robot dependency, cognitive load, and teaming quality.

1) To measure *task performance*, we computed the time taken by participants to complete each game.

For the measurement of the other dependent variables, surveys were administered after each game was completed:

2) A survey measured *perceived robot dependency* by asking the player "How necessary do you believe your input was for the robot to successfully complete its task?", (1=not needed at all to 5=very needed).
3) A survey measured *cognitive load* through the mental demand, temporal demand, performance, effort, and frustration items of the NASA TLX questionnaire [26].
4) To measure *teaming quality*, this survey included two questions in which participants were asked: (a) "Would you consider the robot a good teammate?", (1=poor teammate to 5=good teammate); and (b) "How likely would you be to choose this robot as your teammate if you were to complete this task again?", (1=very unlikely to 5=very likely). These items had good internal consistency, with a Cronbach's alpha value of 0.913. The average of each participant's responses to these questions was used as the final measure of teaming quality.

D. Procedure

After providing informed consent, the experimenter explained the game rules to participants. After participants' questions were answered (if any), they were taken to a room with a laptop and a monitor, positioned to the left of the laptop. The experimenter explained that, on the monitor

display, they could see the Misty II robot surrounded by the four colored stations from where the resources were being collected, and that when the game started, the robot would collect resources from the station that it was facing.

Participants were told that the game interface would show up shortly on the laptop screen and that they could start playing as soon as the interface showed up. The experimenter then headed to a Audio-Visual Monitoring Room and started the game from a computer that was remotely connected to the participant's laptop. After the game was completed, the experimenter ushered participants to the waiting area, where they filled out a post-game survey. While participants were completing the survey, the experimenter set up the following game, which used a different latency level and a different PA strategy. After participants were done with the survey, they were ushered back to the room with the laptop and the monitor, where the next game would happen. This procedure was repeated until participants were done with their third game. Finally, participants were debriefed and paid.

E. Participants

73 students, faculty, and staff (49 M, 24 F) from a small engineering university were recruited. 13 datapoints had to be discarded due to robot network problems, which resulted in 20 participants assigned to each experimental condition (39 M, 21 F). Participants were each paid $15 for their participation.

F. Analysis

Our data (https://bit.ly/roman2023) were analyzed using Bayesian Repeated Measures (RM) ANOVAs with Bayes Factor Analysis, using the bayestestR [27] and BayesFactor [28] R packages. These packages were used to calculate Inclusion Bayes Factors (BF_{10}) across matched models through model averaging [29], which indicate the relative strength of evidence for models including each candidate main effect or interaction effect, in terms of ability to explain gathered data. When a main or interaction effect could not be ruled out ($BF_{10} > 0.333$, i.e. evidence *against* inclusion (BF_{01}) no greater than 3:1), post hoc Bayesian t-tests were used to examine pairwise comparisons between conditions. The next section reports Inclusion Bayes Factors (BF_{10}).

Since Bayesian statistics are still not widely used within the HRI community, we will briefly explain its advantages over the traditional Frequentist approach. Bayesian statistics allows researchers to perform the analysis of data in terms of the strength of evidence for competing hypotheses, allowing for gathering of evidence both for and against hypotheses of interest [30]. This approach allows the acquisition of useful priors from previous study results, making it easier to continue research on the same topic [31], [32]. Critically, Bayesian statistics do not rely on p-values, which have been questioned by recent literature [33], [34], [35], [36]. Because of this, Bayesian Analysis is not grounded on the central limit theorem, adding robustness to small sample size.

IV. EXPERIMENTAL RESULTS

In this section, we will describe the results obtained through our Bayesian analysis. Table I presents the means and standard deviation values for each of the analyses described below.

A. Task Performance Time

A Bayesian RM ANOVA revealed extreme evidence for effects of Latency, PA Strategy, and the interaction between these factors, on task completion time (see Table II). Post hoc results are reported in Tables III and IV. Results for task performance time are visualized in Figure 1a.

Post hoc tests revealed that higher levels of Latency led to longer task completion times. In addition, they revealed that lower PA led to higher task completion times. Finally, post hoc tests suggested that larger levels of latency did indeed lead to larger differences in task completion times between PA strategies.

B. Perceived Robot Dependency

A Bayesian RM ANOVA revealed extreme evidence for an effect of PA on perceived robot dependency and anecdotal evidence for an interaction between PA and Latency, but moderate evidence against a main effect of Latency (see Table II). Post hoc results are reported in Tables III and IV. Results for task performance time are visualized in Figure 1b.

Post hoc tests suggested that lower PA led to higher perceptions of robot dependency. In addition, these tests suggested that larger levels of latency might lead to different levels of perceived dependency across PA strategies. For the interaction effects between Latency and PA Strategy, the primary difference observed is that while under *Low Latency* or *High Latency*, *Medium PA* produced a perception of high dependency alike to that produced by *Low PA*, under *Medium Latency*, *Medium PA* produced a perception of low dependency alike to that produced by *High PA*.

C. Perceived Teaming Quality

A Bayesian RM ANOVA revealed extreme evidence for effects of Latency and PA strategy on perceived teaming quality, and anecdotal evidence against an interaction between the two factors (see Table II). Post hoc results are reported in Tables III and IV. Results for task performance time are visualized in Figure 1c.

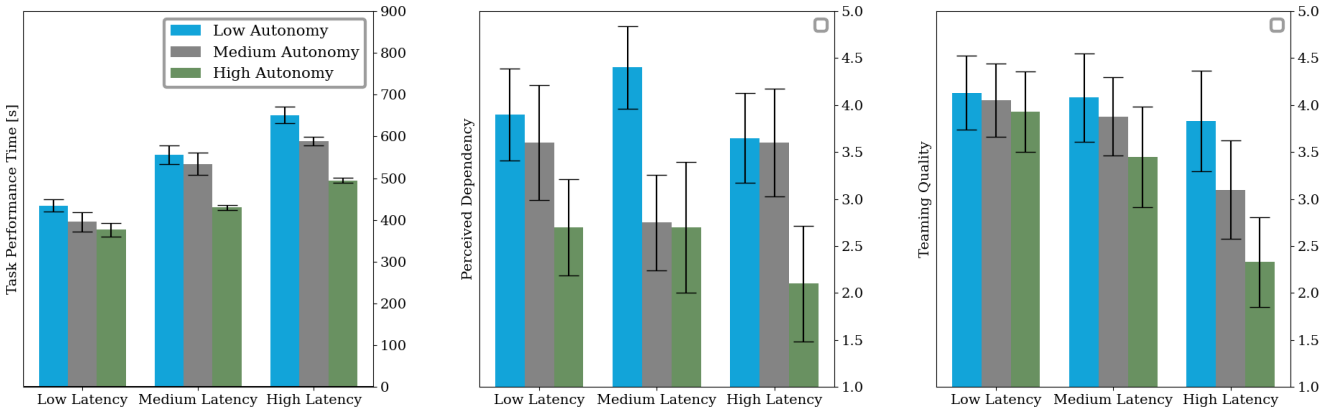
Post hoc tests revealed that high levels of latency led to worse perceptions of teaming quality. In addition, these tests revealed that low levels of PA may have led to better perceptions of teaming quality. Finally, post hoc tests suggested that larger levels of Latency led to larger differences in perceived teaming quality between PA strategy.

D. Cognitive Load

A Bayesian RM ANOVA revealed strong evidence against any effect of Latency, and moderate evidence against any effect of PA strategy or of an interaction between the two factors on participants' cognitive load (see Table II).

	Task Performance Time		Perceived Robot Dependency		Perceived Teaming Quality	
	Mean	**SD**	**Mean**	**SD**	**Mean**	**SD**
Low Latency (LL)	402.033	48.036	3.400	1.317	4.033	0.911
Medium Latency (ML)	506.700	71.925	3.283	1.485	3.800	1.098
High Latency (HL)	578.150	71.596	3.117	1.451	3.083	1.309
Low PA (LPA)	547.100	99.451	3.983	1.097	4.008	1.064
Medium PA (MPA)	506.083	94.854	3.317	1.334	3.675	1.085
High PA (HPA)	433.700	54.514	2.500	1.408	3.233	1.277
LL + LPA	434.350	33.918	3.900	1.119	4.125	0.901
LL + MPA	395.350	52.351	3.600	1.392	4.050	0.887
LL + HPA	376.400	37.897	2.700	1.174	3.925	0.977
ML + LPA	556.000	51.637	4.400	0.995	4.075	1.067
ML + MPA	534.150	60.337	2.750	1.164	3.875	0.944
ML + HPA	429.950	14.314	2.700	1.593	3.450	1.224
HL + LPA	650.950	45.589	3.650	1.089	3.825	1.228
HL + MPA	588.750	25.051	3.600	1.314	3.100	1.199
HL + HPA	494.750	13.894	2.100	1.411	2.325	1.092

TABLE I: Means and Standard Deviation values for each of the analysis groups.



(a) Effects on task performance (b) Effects on perceived dependency. (c) Effects on perceived teaming quality.

Fig. 1: Results: Effects of Latency and PA Strategy on key dependent variables. Error bars represent 95% CI.

	Latency	PA Strategy	Latency * PA Strategy
Task Performance Time	**9.86e+49**	**5.04e+28**	**6.94e+5**
Perceived Robot Dependency	0.140	**1.17e+8**	1.050
Perceived Teaming Quality	**7.21e+4**	**418.990**	0.902
Cognitive Load	0.064	0.103	0.120

TABLE II: Bayesian RM ANOVA Inclusion Bayes Factors. Results with conclusively positive evidence are bolded; results with conclusively negative evidence are grayed out.

V. DISCUSSION

A. Hypothesis One

Hypothesis **(H1)** was that task completion time, perceived robot dependency, and teaming quality would increase as PA level increased, and that cognitive load would remain unaffected. This hypothesis was supported. Performance of lower levels of autonomy increased teaming quality ratings, task completion times, and perceived robot dependency ratings, without increasing cognitive load. We will now cover our findings.

1) Perceived Teaming Quality: Robots performing lower levels of autonomy were generally perceived as better teammates. This is likely not only because robots performing lower levels of autonomy are better able to take human preferences into account, but also because they demonstrate their desire to do so, thus affirming interactants' face needs [37] and demonstrating their social agency [38]. If so, this would promote key long-term benefits arising from this autonomy design strategy, even outside of space exploration contexts. For example, social companion robots, such as Pepper [39], Pearl [40], or Moxie [41] might benefit from the use of PA to establish better social connections with their human users. In addition, robots with PA capabilities in mixed human-robot teams could lead to better collaboration by promoting more communication [42]. Future work could explore these

	Latency			PA Strategy		
	Low Latency (LL) v Medium Latency (ML)	**LL v High Latency (HL)**	**ML v HL**	**Low PA (LPA) v Medium PA (MPA)**	**LPA v High PA (HPA)**	**MPA v HPA**
Task Performance Time	**7.45e+12**	**3.09+27**	**4.61e+4**	2.109	**1.78e+9**	**1.21e+4**
Perceived Robot Dependency	0.214	0.342	0.232	**10.092**	**3.4e+6**	**20.867**
Perceived Teaming Quality	0.401	**1709.705**	**20.199**	0.711	**57.908**	1.256

TABLE III: Inclusion Bayes Factors (BF_{10}) for Latency and PA Strategy. Results with conclusively positive evidence are bolded; results with conclusively negative evidence are grayed out.

	Low Latency			Medium Latency			High Latency		
	Low PA (LPA) vs Medium PA (MPA)	**LPA vs High PA (HPA)**	**MPA vs HPA**	**LPA vs MPA**	**LPA vs HPA**	**MPA vs HPA**	**LPA vs MPA**	**LPA vs HPA**	**MPA vs HPA**
Task Performance Time	**5.857**	**1630.703**	0.608	0.561	**5.79e+9**	**1.6e+6**	**3286**	**9.46e+13**	**9.82e+13**
Perceived Robot Dependency	0.386	**17.279**	2.024	**763.634**	**101.040**	0.310	0.311	**67.945**	**25.535**
Perceived Teaming Quality	0.318	0.370	0.332	0.361	0.982	0.561	1.237	**110.344**	1.80

TABLE IV: Inclusion Bayes Factors (BF_{10}) for Interaction Effects. Results with conclusively positive evidence are bolded; results with conclusively negative evidence are grayed out.

opportunities within and beyond space contexts.

2) Task Performance: Performing lower levels of autonomy led to longer task completion times, as participants needed additional time to process and respond to robots' queries. As such, although providing clear benefits to SA [5], PA as an autonomy design strategy increases the time needed to complete tasks. This suggests that while PA can provide key SA benefits, it nevertheless may need to be avoided in HRI contexts with serious time constraints. Future work should investigate the suitability of PA in such scenarios.

3) Perceived Dependency: Robots performing lower levels of autonomy were perceived as more dependent on their human interactants. These results bolster the results of [5] by confirming that levels of autonomy can be "performed" through dialogue moves, and that these moves are interpreted by humans as intended. Moreover, these results suggest that humans might be more willing to collaborate with robots that are perceived as less autonomous, even if this leads to longer task completion times.

4) Perceived Cognitive Load: Finally, our results confirm [5]'s finding that performing lower levels of autonomy does not increase cognitive load. These results provide continued motivation for the use of PA as an autonomy design strategy. However, future work should explore whether a task imposing a greater baseline level of cognitive load, or consideration of a broader range of performed levels of autonomy, could ultimately lead to effects of PA on cognitive load.

B. Hypothesis Two

Our second hypothesis, **(H2)**, was that negative consequences associated with low PA strategies would be exacerbated by increased latency. This hypothesis was partially supported: increases to task performance times were indeed exacerbated by latency, but cognitive load increases did not begin to manifest as latency increased. Moreover, the way that perceptions of performance degraded under latency were quite different from what we expected. We will now step through each of these findings.

1) Perceived Teaming Quality: While we expected the degradation of perceived teaming quality under increased latency to be exacerbated under lower levels of PA, we observed that while perceptions of teaming quality did degrade under latency, these degradations happened for *higher* levels of PA. That is, while the *Low PA* strategy led to only a slight fall in teaming quality ratings as latency increased, the *Medium* and *High PA* strategies led to steeper drops. In other words, the PA design strategy unexpectedly provided resilience in the face of latency.

Reflecting on these results, we suspect that the way that higher levels of autonomy strip control from human teammates may have been acutely felt under latency. These results suggest that PA may be a beneficial strategy for encouraging cohesive teams under conditions of latency, but also suggest that lower levels of autonomy in general may be advantageous in such condition.

2) Task Performance: As we observed, the ways that PA increased task completion time (and thus negatively impacted task performance) were exacerbated as latency increased, with the gaps in completion time observed under low latency growing more pronounced as latency increased. These results reinforce our suggestion that despite the subjective benefits observed above and the benefits to SA observed by [5], PA may be ill-suited to contexts with significant time constraints. Ultimately, the choice of whether to adaptively perform lower levels of autonomy – and moreover, *which* lower levels should be performed – must be grounded in analysis of

a given task and the need for expedience versus need for subjective and awareness benefits specific to that task.

3) Perceived Cognitive Load: Finally, we observed that even as latency increased, cognitive load levels stayed constant, regardless of PA strategy. These results thus confirm the results of [5] while extending our understanding of those results with respect to the effects of Latency. These results provide further motivation for this autonomy design strategy. However, as we mentioned in the previous section, it is possible that other collaborative tasks imposing higher baseline levels of cognitive load might ultimately result in differential changes in cognitive load as latency increases. This represents a natural direction for future work.

C. Limitations and Future Work

While the small latency intervals used in this study are sufficient for the purposes of this paper, remote human-robot interactions between Earth and Mars are substantially longer, on the order of 6-40 minutes for round-trip communication [43]. While these types of delays could not feasibly be explored in our proposed lab experiment context, these longer intervals are likely to manifest entirely different consequences than those observed in this work. Thus, as humanity moves to the moon, Mars, and beyond, it will be critical to understand the impact of these larger latencies on communication dynamics. In addition, our simulations do not capture the risk that is intrinsic to collaborative tasks in space. Thus, tasks that have a more critical nature can be explored in future work. Furthermore, although both the *Low Latency* condition in this experiment and the general experimental paradigm used by [5] accurately represent the *Latency* dimension of proximal human-robot interactions, neither investigate latency-free *face-to-face* interactions. As such, future work is needed to understand the effects of PA in truly proximal, face-to-face interactions.

While [5]'s work evaluated PA in terms of SA and Cognitive Load, and this work investigated PA through task performance, Cognitive Load, and perceptions of dependency and teaming quality, future work should explore the impacts of this autonomy design strategy on other key measures of HRI quality. One such measure worth exploring is Trust, a key human factors construct that plays a central role in human-robot interactions [44], [45], [46], [47], and has been suggested to depend on both autonomy [48], [49], [50] and on perceptions of robots as being good teammates [51]. It would be interesting to determine whether the resilience to latency that PA provided for perceived teaming quality would translate to trust resilience [52].

Future work could also explore a wider range of PA levels and dynamics. In this work, our experimental conditions dictated two specific levels of PA (below the fully autonomous baseline). However, the ideal use of this autonomy design strategy would be for robots to *adaptively* switch PA strategies based on their interaction context. Future work should explore dynamically adaptive PA, as well as the effects of performing lower levels of autonomy relative to baseline levels that are not fully autonomous.

Finally, as discussed previously in this paper, future work should (a) explore the effects of PA in time constrained contexts and in contexts with higher baseline demands on cognitive load, (b) explore the opportunities afforded by PA in other contexts within and beyond space exploration domains, and (c) explore whether the effects of latency observed in this work would mediate the impacts of PA on Situation Awareness that had been observed by [5].

VI. CONCLUSION

This paper presented the first study investigating the role of latency in mediating the benefits of the *Performative Autonomy* design strategy. Our results suggest that as latency increases, lower performed autonomy does indeed exacerbate the longer task completion times experienced by interactants under high latency. However, lower performed autonomy did not exacerbate the negative impacts of latency on subjective perceptions of robots' teaming quality. Furthermore, while lower performed autonomy increased subjective perceptions of robot dependency, we argued that this is not necessarily a bad thing, as lower performed autonomy simultaneously led to perceptions that robots were better teammates. Finally, we found that despite our concerns regarding cognitive load, participants' cognitive load did not degrade with increasing latency, nor was it negatively effected by lower performed autonomy, even under high latency. These results support [5]'s conclusion that *Performative Autonomy* does not meaningfully increase cognitive load, despite its inversion of *Adaptive Autonomy*, in which higher levels of autonomy are performed explicitly to reduce cognitive load. Moreover, these results generally support the use of *Performative Autonomy* even when high communication latency is imposed, except perhaps in contexts with high time pressure. These results thus motivate further work within the new *Performative Autonomy* paradigm to understand how it affects other key constructs central to HRI, such as human-robot trust, and to understand whether and how latency might impact these constructs.

VII. ACKNOWLEDGEMENTS

This work was funded in part by NASA Early Career Faculty award 80NSSC20K0070.

REFERENCES

[1] M. Bualat, T. Smith, E. Smith, T. Fong, and D. Wheeler, "Astrobee: A new tool for iss operations," in *SpaceOps Conference*, 2018.

[2] M. Bualat, J. Barlow, T. Fong, C. Provencher, and T. Smith, "Astrobee: Developing a free-flying robot for the international space station," in *AIAA SPACE 2015 Conference and Exposition*, 2015.

[3] T. Fong and I. Nourbakhsh, "Interaction challenges in human-robot space exploration," *Interactions*, 2005.

[4] P. Thuot, "International space station (iss) eva suit water intrusion, high visibility close call," NASA, Tech. Rep. IRIS Case Number: S-2013-199-00005, 2013.

[5] S. Roy, T. Smith, B. Coltin, and T. Williams, "I need your help... or do i? maintaining situation awareness through performative autonomy," in *Int'l Conf. on Human-Robot Interaction (HRI)*, 2023.

[6] D. Kornuta, A. Abbud-Madrid, J. Atkinson, J. Barr, G. Barnhard, D. Bienhoff, B. Blair, V. Clark, J. Cyrus, B. DeWitt *et al.*, "Commercial lunar propellant architecture: A collaborative study of lunar propellant production," *Reach*, 2019.

[7] D. Chiappe, K.-P. Vu, C. Rorie, and C. Morgan, "A situated approach to shared situation awareness," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 2012.

[8] T. Sheridan and W. Verplank, "Human and computer control of undersea teleoperators," Massachusetts Inst of Tech Cambridge Man-Machine Systems Lab, Tech. Rep., 1978.

[9] M. Endsley and D. Kaber, "Level of automation effects on performance, situation awareness and workload in a dynamic control task," *Ergonomics*, 1999.

[10] M. Dorneich, S. Whitlow, S. Mathan, P. Ververs, D. Erdogmus, A. Adami, M. Pavel, and T. Lan, "Supporting real-time cognitive state classification on a mobile individual," *Journal of Cognitive Engineering and Decision Making*, 2007.

[11] M. Dorneich, W. Rogers, S. Whitlow, and R. DeMers, "Human performance risks and benefits of adaptive systems on the flight deck," *The International Journal of Aviation Psychology*, 2016.

[12] D. Kaber, E. Onal, and M. Endsley, "Design of automation for telerobots and the effect on performance, operator situation awareness, and subjective workload," *Human factors and ergonomics in manufacturing & service industries*, 2000.

[13] L. Onnasch, C. Wickens, H. Li, and D. Manzey, "Human performance consequences of stages and levels of automation: An integrated meta-analysis," *Human factors*, 2014.

[14] M. Endsley and E. Kiris, "The out-of-the-loop performance problem and level of control in automation," *Human factors*, 1995.

[15] M. Endsley, "From here to autonomy: lessons learned from human–automation research," *Human factors*, 2017.

[16] C. Wickens, "Automation lessons from other domains," *Handbook of Hum. Fac. for Automated, Connected, and Intelligent Vehicles*, 2020.

[17] K. Petersen and O. Von Stryk, "Towards a general communication concept for human supervision of autonomous robot teams," in *Proceedings of the fourth international conference on advances in computer-human interactions (ACHI)*, 2011.

[18] M. Goodrich, A. Schultz *et al.*, "Human–robot interaction: a survey," *Foundations and Trends® in Human–Computer Interaction*, 2008.

[19] B. Wilcox, "Robotic vehicles for planetary exploration," *Applied Intelligence*, 1992.

[20] R. McGregor and L. Oshinowo, "Flight 6a: deployment and checkout of the space station remote manipulator system (ssrms)," in *Int'l Symp. on AI, Robotics and Automation in Space (i-SAIRAS)*, 2001.

[21] M. Diftler, J. Mehling, M. Abdallah, N. Radford, L. Bridgwater, A. Sanders, R. Askew, D. Linn, J. Yamokoski, F. Permenter *et al.*, "Robonaut 2-the first humanoid robot in space," in *IEEE international conference on robotics and automation*, 2011.

[22] A. Khasawneh, H. Rogers, J. Bertrand, K. Madathil, and A. Gramopadhye, "Human adaptation to latency in teleoperated multi-robot human-agent search and rescue teams," *Automation in Construction*, 2019.

[23] S. Rumble, D. Ongaro, R. Stutsman, M. Rosenblum, and J. Ousterhout, "It's time for low latency," in *13th Workshop on Hot Topics in Operating Systems (HotOS XIII)*, 2011.

[24] B. Briscoe, A. Brunstrom, A. Petlund, D. Hayes, D. Ros, J. Tsang, S. Gjessing, G. Fairhurst, C. Griwodz, and M. Welzl, "Reducing internet latency: A survey of techniques and their merits," *IEEE Communications Surveys & Tutorials*, 2014.

[25] L. Zhu and T. Williams, "Effects of proactive explanations by robots on human-robot trust," in *Int'l Conference on Social Robotics*, 2020.

[26] S. Hart and L. Staveland, "Development of nasa-tlx (task load index): Results of empirical and theoretical research," in *Advances in psychology*, 1988.

[27] D. Makowski, M. Ben-Shachar, and D. Lüdecke, "bayestestr: Describing effects and their uncertainty, existence and significance within the bayesian framework," *Journal of Open Source Software*, 2019.

[28] R. Morey, J. Rouder, T. Jamil, and M. R. Morey, "Package 'bayesfactor'," *URL http://cran/r-projectorg/web/packages/BayesFactor/BayesFactor*, 2015.

[29] M. Hinne, Q. Gronau, D. van den Bergh, and E.-J. Wagenmakers, "A conceptual introduction to bayesian model averaging," *Advances in Methods and Practices in Psychological Science*, 2020.

[30] A. Jarosz and J. Wiley, "What are the odds? a practical guide to computing and reporting bayes factors," *Problem Solving*, 2014.

[31] J. Verhagen and E.-J. Wagenmakers, "Bayesian tests to quantify the result of a replication attempt." *Journal of Experimental Psychology: General*, 2014.

[32] A. Ly, A. Etz, M. Marsman, and E.-J. Wagenmakers, "Replication bayes factors from evidence updating," *Behavior research meth.*, 2019.

[33] J. Berger and T. Sellke, "Testing a point null hypothesis: The irreconcilability of p values and evidence," *Journal of the American statistical Association*, 1987.

[34] J. Simmons, L. Nelson, and U. Simonsohn, "False-positive psychology: undisclosed flexibility in data collection and analysis allows presenting anything as significant." in *Methodological issues and strategies in clinical research*, 2016.

[35] J. Sterne and G. D. Smith, "Sifting the evidence—what's wrong with significance tests?" *Physical therapy*, 2001.

[36] E.-J. Wagenmakers, "A practical solution to the pervasive problems of p values," *Psychonomic bulletin & review*, 2007.

[37] P. Brown and S. Levinson, *Politeness: Some universals in language usage*. Cambridge university press, 1987.

[38] R. Jackson and T. Williams, "A theory of social agency for human-robot interaction," *Frontiers in Robotics and AI*, 2021.

[39] A. Pandey and R. Gelin, "A mass-produced sociable humanoid robot: Pepper: The first machine of its kind," *IEEE Robotics & Automation Magazine*, 2018.

[40] M. Pollack, L. Brown, D. Colbry, C. Orosz, B. Peintner, S. Ramakrishnan, S. Engberg, J. Matthews, J. Dunbar-Jacob, C. McCarthy *et al.*, "Pearl: A mobile robotic assistant for the elderly," in *AAAI workshop on automation as eldercare*, 2002.

[41] N. Hurst, C. Clabaugh, R. Baynes, J. Cohn, D. Mitroff, and S. Scherer, "Social and emotional skills training with embodied moxie," *arXiv preprint arXiv:2004.12962*, 2020.

[42] G. Hoffman and C. Breazeal, "Collaboration in human-robot teams," in *AIAA 1st intelligent systems technical conference*, 2004.

[43] A. Mishkin, Y. Lee, D. Korth, and T. LeBlanc, "Human-robotic missions to the moon and mars: operations design implications," in *2007 IEEE Aerospace Conference*, 2007.

[44] P. Hancock, D. Billings, K. Schaefer, J. Chen, E. De Visser, and R. Parasuraman, "A meta-analysis of factors affecting trust in human-robot interaction," *Human factors*, 2011.

[45] A. Freedy, E. DeVisser, G. Weltman, and N. Coeyman, "Measurement of trust in human-robot collaboration," in *International symposium on collaborative technologies and systems*, 2007.

[46] T. Law and M. Scheutz, "Trust: Recent concepts and evaluations in human-robot interaction," *Trust in human-robot interaction*, 2021.

[47] E. De Visser, M. Peeters, M. Jung, S. Kohn, T. Shaw, R. Pak, and M. Neerincx, "Towards a theory of longitudinal trust calibration in human–robot teams," *International journal of social robotics*, 2020.

[48] C. Furlough, T. Stokes, and D. Gillan, "Attributing blame to robots: I. the influence of robot autonomy," *Human factors*, 2021.

[49] J. Beer, A. Fisk, and W. Rogers, "Toward a framework for levels of robot autonomy in human-robot interaction," *Journal of human-robot interaction*, 2014.

[50] M. Desai, M. Medvedev, M. Vázquez, S. McSheehy, S. Gadea-Omelchenko, C. Bruggeman, A. Steinfeld, and H. Yanco, "Effects of changing reliability on trust of robot systems," in *7th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 2012.

[51] R. Wen, Z. Han, and T. Williams, "Teacher, teammate, subordinate, friend: Generating norm violation responses grounded in role-based relational norms." in *Proceedings of the ACM/IEEE International Conference on Human-RObot Interaction*, 2022.

[52] E. De Visser, S. Monfort, R. McKendrick, M. Smith, P. McKnight, F. Krueger, and R. Parasuraman, "Almost human: Anthropomorphism increases trust resilience in cognitive agents." *Journal of Experimental Psychology: Applied*, 2016.