

# Analyzing Autonomy and Its Relationship to Interdependence in Human-Agent-Robot Teams

---

## Introduction

It is commonly believed that increasing the autonomy of certain classes of systems will improve their performance. For example, the United States Department of Defense Unmanned Systems Roadmap [1] states “The Department will pursue greater autonomy in order to improve the ability of unmanned systems to operate independently, either individually or collaboratively, to execute complex missions in a dynamic environment (pg. 1).” In the context of a report on a Gulf oil spill, a recent IEEE article suggested “Automation techniques will improve not only the time that it takes to do these tasks but also the quality of the results [2].”

General conclusions of this sort can be misleading for a variety of reasons. In this article, we highlight one of these reasons: namely, that in complex joint activity involving mixed teams of humans, software agents, and robots, increases in autonomy may eventually lead to degradations in performance when the conditions that enable effective management of interdependence among the team members are neglected.

More effective management of interdependence in joint activity will become increasingly important in the coming years. The sophisticated robots envisioned for the future will be increasingly collaborative in nature, not merely doing things *for* people, but also working together *with* people and intelligent systems. Though continuing research is needed to make agents and robots more *independent* during times when unsupervised activity is desirable or necessary (i.e., *autonomy*), they must also be more capable of sophisticated *interdependent* joint activity when such is required (i.e., *coactivity*). The mention of *joint* activity highlights the need for coactive human-agent-robot systems to support not only fluid orchestration of task handoffs among different people and machines, but also combined participation on shared tasks requiring continuous and close interaction. Because the capabilities for coactivity interact with autonomy algorithms at a deep level, they must be embedded in system design from the beginning, not layered on with a thin veneer after the fact, as is sometimes attempted.

Based on this premise, our long-range goal is to develop a prescriptive methodology to guide the design and analysis of human-agent-robot systems. This methodology will be formulated in light of an appreciation for the essential role of interdependence in joint human-agent-robot activity. In this article, we explore how changes in autonomy can affect various dimensions of performance when interdependence is neglected. Although our experimental results stem from a simple task domain performed in a simulation environment, both our findings in the literature on human teamwork and our experience in a variety of human-agent-robot teamwork experiments and field exercises give us reason to believe that these results eventually can be generalized.

## Background

In a previous article [3], we described the problems of what we refer to as "autonomy-centered approaches." We conclude that:

*Even when self-directedness and self-sufficiency are reliable, matched appropriately to each other, and sufficient for the performance of the robot's individual tasks, human-robot teams engaged in consequential joint activity frequently encounter the potentially debilitating problem of opacity, meaning the inability for team members to maintain sufficient awareness of the state and actions of others to maintain effective team performance [3].*

Many examples supporting this conclusion can be found in the literature. For example, Stubbs recently noted lack of transparency as a problem in human-robot interaction [4]. More generally, this issue was identified more than two decades ago by Norman as "silent automation" [5], and subsequently by Woods as "automation surprises" [6]. In this article we will use the term "opacity" to highlight similar problems stemming from a lack of transparency in human-automation interaction. However, it is important to recognize that the challenges go far beyond simply not being able to see needed information. They can also involve predictability, directability or other challenges that must be addressed in order to turn autonomous systems into team players [7].

## The Experiment

Our goal was to demonstrate that in human-agent-robot systems engaged in joint activity, increasing autonomy without addressing interdependence may lead to suboptimal performance. We attempted to rule out over-trust in automation as a failure factor by ensuring that the agent players never made mistakes and that they exhibited reasonably intelligent behavior. We also attempted to ensure that the interaction between the human and the agent could be at a relatively high level of abstraction—i.e., that the agent's capabilities for autonomy were not under-utilized. We did not want an agent capable of completing the mission autonomously managed at a low level akin to teleoperation. To this end, we provided an interface appropriate to agents' capabilities. These elements of our experimental design are illustrated in Figure 1 (A).

Figure 1 (B) illustrates the general trends we expected to find in our results. We anticipated that the management burden the agent player imposed on the human player would decrease as agent autonomy increased. Such a finding would be no surprise, since reduction in human workload is both the common expectation and the major motivation for automation. However, we also anticipated that, without support for managing interdependence issues, the opacity of the work system to task participants would grow with increasing autonomy. Due to these competing factors of burden and opacity, we expected an inflection point in team performance, where the benefits of increasing autonomy eventually would be completely offset by the negative side effects of opacity. In other words, we predicted that the highest level of autonomy would not demonstrate the highest level of team performance, consistent with the general shape of the notional bar graph shown in Figure 1 (C).

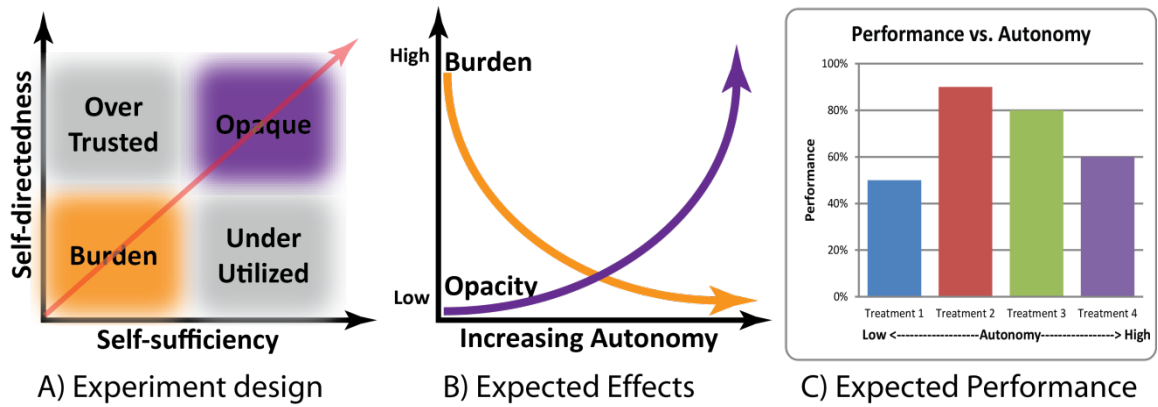


Figure 1 A) Illustration of our experimental design approach. B) Expected effects of increasing autonomy on the burden of managing the agent and the opacity of the agent to other task participants. C) Expected performance under treatment conditions of increasing autonomy, due to the competing factors of agent management burden and agent opacity.

## The Experimental Domain

Our domain for this experiment is Blocks World for Teams (BW4T) [8]. Similar in spirit to Winograd’s classic AI planning problem of Blocks World, the goal of BW4T is to “stack” colored blocks in a particular order. The task environment (Figure 2) is composed of nine rooms containing a random assortment of blocks and a drop off area for the goal. Each player controls an avatar in the game. This avatar can be moved between rooms to pick up and drop off blocks. For this experiment, teams were composed of two players—a human and a software agent. The two players work toward the shared team goal, which is to deliver the colored blocks to the drop zone in a specified order. Players are limited in their awareness of the situation: they cannot see each other and they can only see blocks that are in their current room. Human players control their own avatar in order to find and deliver blocks. They also command their agent partner through an appropriate interface. Variations on the basic game, and different experimental manipulations, can be easily programmed into the environment.

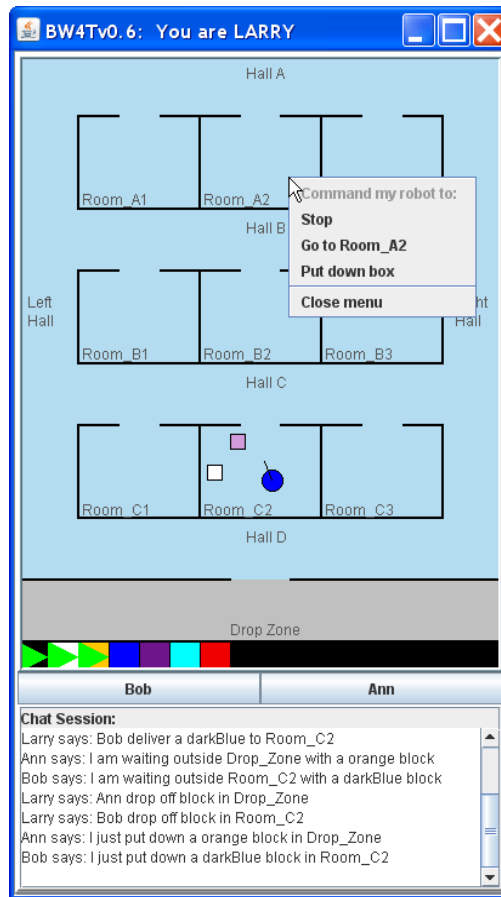


Figure 2 Example Blocks World for Teams (BW4T) interface

## Defining the Agent Teammate

The algorithm chosen as the basis for the agent behavior reflects the most common approach we observed for human players of the game. This algorithm was chosen because we felt it would be easily understandable and predictable for most human players. The algorithmic solution is shown on the left side of Figure 3. The main goal (a color sequence) is composed of several subgoals (individual colors). To achieve any given subgoal, one simply finds the block of the appropriate color and delivers it. Note that these tasks need not be performed in sequence or by the same player. For example, a player could first find all the blocks and then deliver them. Alternatively, one player could find a block and another could deliver it. The overall task can be thought of as being composed of several *find* tasks and several *deliver* tasks, which are themselves composed of some decision and action primitives. The action primitives include going to a room, entering the room, going to a block, picking up a block, and putting down a block. The two main decisions are: 1) whether to look for a block or to deliver a block, and 2) which room to go to in order to look for a block. The agent player is designed to perform its task “perfectly,” meaning it will perform any assigned task efficiently and will make rational decisions based on a complete and accurate recollection of where it has been and what it has seen in the past. It will also report when a task is completed. To be consistent, it *only* reports the completion status when an assigned task is completed, and does not provide any additional information.

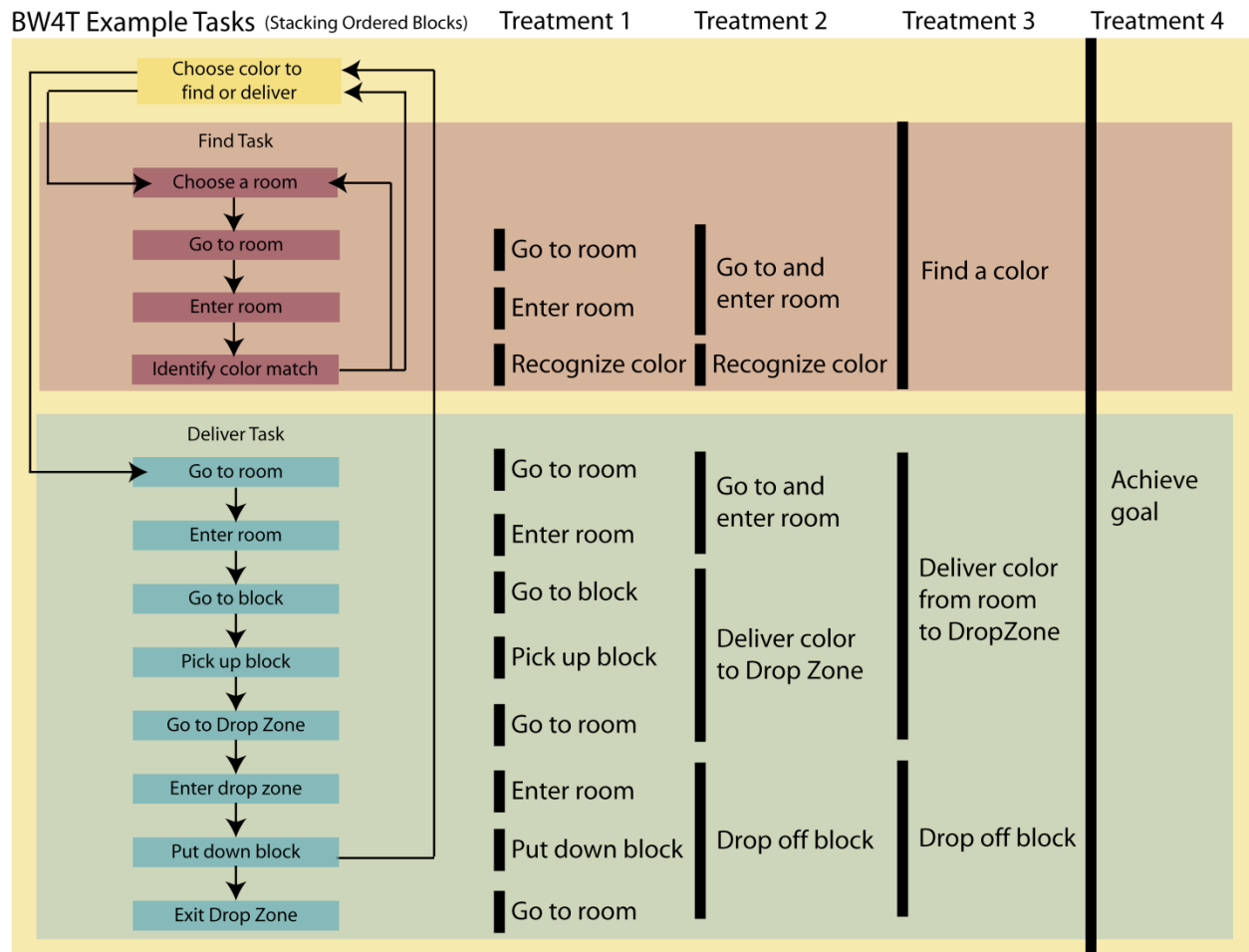


Figure 3 Defining Autonomy Treatments for BW4T

## Defining the Autonomy Treatments

In order to compare the effects of changing autonomy, we defined different experimental conditions or “autonomy treatments.” Additionally, we needed some way to rank the treatments ordinally in terms of their relative degree of autonomy. For this purpose we applied the concepts of levels of autonomy [9], proposed by Sheridan and Verplank, and the neglect tolerance metric [10], proposed by Olsen and Goodrich. Neglect Tolerance is a metric based on the amount of time a human can ignore a given robot performing a given task before the robot becomes unproductive.

Treatment 1 requires the human player to direct the agent player using only the action primitives. The vertical black lines or bands in Figure 3 are used to indicate the portion of the algorithm that is performed autonomously by the agent player. During the time it spends in the black band, the agent can be considered as functioning at Sheridan’s highest level of autonomy, since the agent will perform on its own everything necessary to complete the task specified by the band. Outside the band, the agent is at the lowest level of autonomy and is completely reliant on the human for all decisions and actions. The behavior associated with each band is always initiated by the human teammate. The neglect tolerance correlates to the length of the band, though the band covers a portion of the algorithm and does not directly correspond to length of time, since some tasks take longer than others. However, longer bands

cover more sections of the algorithm; thus, in general they entail more autonomy. The bands in treatment 1 are the shortest, requiring more direction from their human teammate and therefore have the lowest neglect tolerance.

In treatment 2, we combine several action primitives into a single action. For example, with a single command the agent can now be ordered to go to and enter a room. To inhibit under-utilization, the command set available to the human player was restricted to the new “higher-level” commands listed under treatment 2 in Figure 3. We are only combining action primitives, so Sheridan’s scale does not provide much guidance, but it is clear that agent neglect tolerance increases and, thus, this treatment has more autonomy for the agent than the first.

Treatment 3 is identical to Treatment 2, except that it also provides the ability to command the agent to find a color. This new command delegates the decision on where to search to the agent, who is now required to provide its own search algorithm and only reports when a color is found. This was implemented as a nearest-unsearched room algorithm, which was the most common approach for human players. Again, the human player was restricted to the commands that are listed under treatment 3 in Figure 3. Consistent with Sheridan’s specification for levels of autonomy, this is a higher level of autonomy than the previous treatment, since the agent can now make its own decision on how to achieve the find task. The level of neglect tolerance is also higher.

Treatment 4 is identical to Treatment 3, except that it also allows the agent to choose whether to look for a block or deliver a block. This enabled the agent player to be able to complete the entire task without any assistance from the human player. This competence level equates to Sheridan’s highest level of autonomy and an infinite tolerance for neglect. The only required command by the human player is to tell the agent to achieve the goal. As in Sheridan’s level ten, the agent “decides everything, acts autonomously, ignoring the human” [11].

We have intentionally left out any support for managing interdependence, except for communicating task completion status. There is neither communication about world state nor coordination of task activity. While this may seem extreme in this simple domain with obvious coordination needs, we believe it is not unrealistic given the prevalence of similarly opaque systems [4-6]. By this means, we hoped to explore the relationship between autonomy and interdependence.

## Experimental Design

24 participants (17 male and 7 female) were selected from a student population at TU Delft, with an age range of 19-39. We employed a complete randomized block design based on the autonomy treatment, with each participant performing each treatment once. The data are cross-classified by  $k = 4$  autonomy treatments and  $b = 24$  blocks, consisting of the individual participants. All participants received a demographic survey. They were trained on the game until they demonstrated proficiency by completing a simplified version of the task. Next they performed a series of trials, one for each treatment. The participant filled out a brief survey at the end of the experiment, evaluating team burden, opacity, performance, and preference in each treatment.

## Results

Our results include quantitative numeric data as well as subjective ranking data. For the former, we use standard approaches for normal data. For the ranked data, we used the nonparametric Friedman test. Based on our design, and using the  $\alpha = 0.05$  level of significance, the critical value is  $\chi^2_{.95,3} = 7.815$ .

### Assessing Burden

Our hypothesis predicted a decrease in agent management burden as autonomy increased from treatments 1 to 4. We asked the participants to rank how demanding it was to work with the agent in each condition, on a scale of 1 (least demanding) to 4 (most demanding). The results, shown in Figure 4 (A), indicate a very clear ( $\chi^2_{.95,3} = 34.225$ ) decrease in burden as autonomy increased. As a second, independent measure of burden, we also counted the number of commands the human player had to give to the agent teammate in each condition. Figure 4 (B) shows the results, which correlate with the subjective assessment.

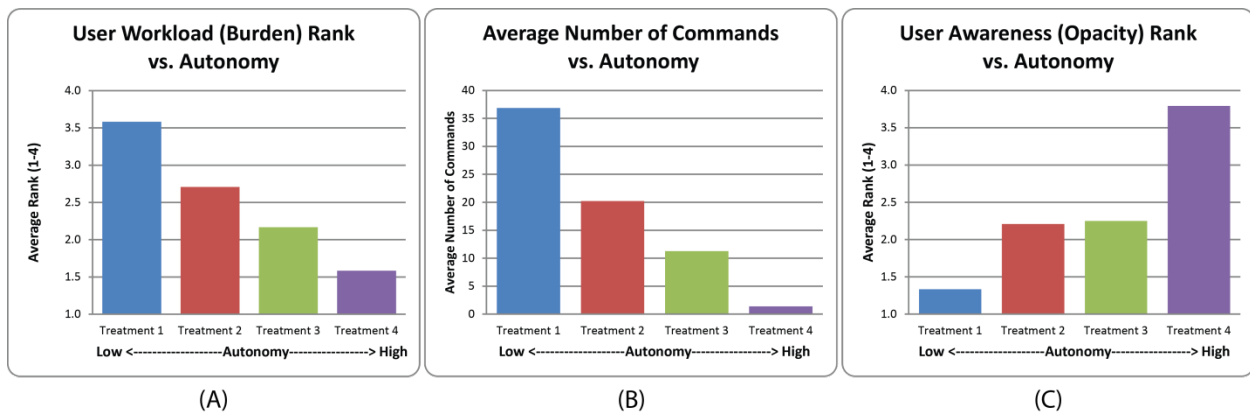


Figure 4 (A) Subject ranking of agent management workload (burden) as autonomy increases across experimental treatments. (B) Average number of commands (Burden) as autonomy increases. (C) Average subjective rankings of awareness (Opacity) as autonomy increases.

### Assessing Opacity

Our hypothesis predicted an increased subject perception of opacity with increasing autonomy across the experimental conditions. We expected this to be reflected in reports of subjects having more difficulty in understanding what was happening and in anticipating the agent's behavior as autonomy increased. An exit survey was used where subject were asked to rank their ongoing sense of awareness of current and future agent actions in the different conditions on a scale of 1 (most aware) to 4 (least aware). The results in Figure 4 (C) show opacity increasing with increasing autonomy as predicted ( $\chi^2_{.95,3} = 49.700$ ). This confirms our prediction about opacity in this experimental setting, and validates the general expectation illustrated in Figure 1 (B).

## Quantitative Performance Assessment

We performed three different quantitative performance assessments: time to complete task, idle time, and error rate.

### Time to complete task

The simplest performance metric is time to completion—i.e., delivering all the required blocks in the requested order. Figure 5 shows the results. At first glance, the results appear promising. We can clearly see the inflection point where performance begins to degrade rather than improve under conditions of increasing autonomy, consistent with the prediction of Figure 1 (C). The differences, however, were not statistically significant ( $p = 0.20$ ). We believe that this is best explained by the fact that the task itself has a large amount of variance from run to run, and the penalty incurred by errors is less than the variance between runs. We note, however, that in 83% of the participants, the highest-autonomy condition (Treatment 4) was not the highest-performing condition by the time-to-completion criterion.

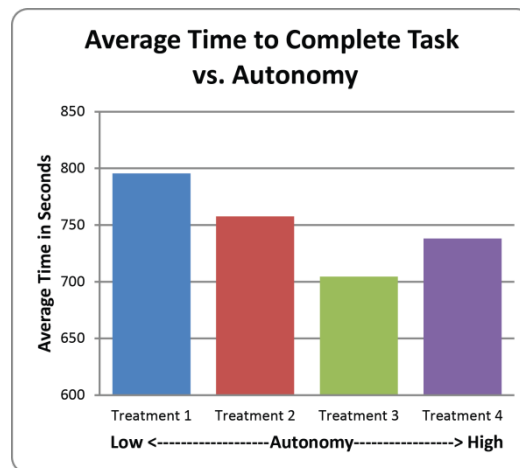


Figure 5 Time-to-completion as autonomy increases across treatments.

### Idle Time

Another important performance measure is idle time (or wait time [12]). In the BW4T task, the agent player will be in near constant motion once a task has been assigned to it by its human teammate. Any idle time is indicative of inefficient use of the agent player (e.g., while it awaits the next command). Figure 6 (A) shows the results of average idle time for the agent player. There is a clear and significant decrease in idle time from treatment 1 to 4. On the surface, this could be taken as indicating more effective use of the agent player by the human, and thus suggesting improved performance. However, this is not borne out by the time-to-completion results (Figure 5). Additionally, we note that the amount of work done is fairly consistent across treatments. For example, the number of rooms entered and the number of boxes delivered does not change much across treatments. This also makes sense when one looks at the human player's idle time, shown in Figure 6 (B). There is a slight decrease in idle time as the burden is reduced, but not much, and certainly not on the order of the change seen in the agent player. This indicates that the interaction efficiency [12] is not that significant. This could be due to an effective

interface, but it also can be due to the ability to multi-task and complete interactions concurrent with motion. The interesting takeaway lesson from this result is that “keeping your agent busy” does not equate to improved performance.

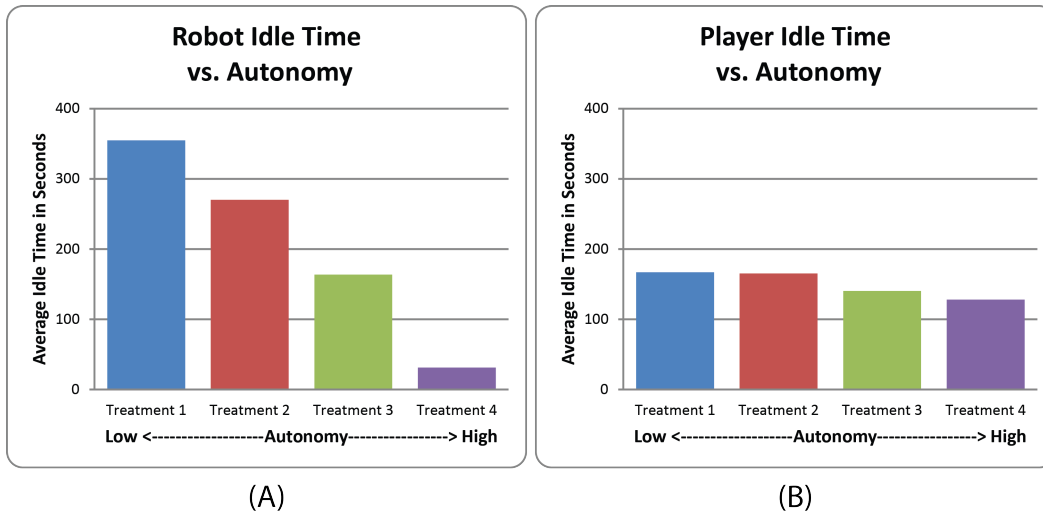


Figure 6 (A) Average agent player idle time across treatment conditions. (B) Average human player idle time.

### Error Rate

For some kinds of tasks, error rate can be a good way to compare performance. We measured this in three ways. Our first was the amount of time that both players spent holding the same color block (Figure 7 (A)). Since, for this experiment, the goals were composed of unique colors (no repeats), this represented a measure of some fraction of overall redundant activity or inefficiency in task performance. This type of error, for the most part, only occurred in treatment 4 and is a side effect of the high opacity of the highest-autonomy condition. These results are no surprise, since this is the only treatment in which the agent player can make its own decision about which block to pick up. However, this does emphasize that functional differences matter when automating tasks [13].

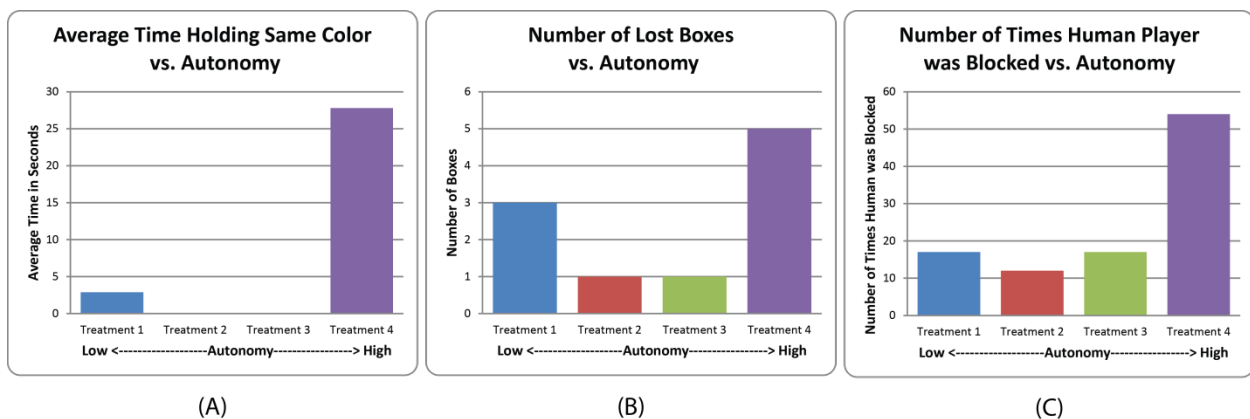


Figure 7 (A) Average time holding the same color (inefficiency) (B) Number of lost boxes (C) Number of times a human player was blocked by their agent partner while trying to enter a room

A second measure of error is the number boxes lost—i.e., dropped in the hallway or placed in the drop zone erroneously. Since BW4T is very simple, there were not many mistakes made by the human players, but of the ten lost boxes, 50% of them occurred in treatment 4 and 30% occurred in treatment 1, as shown in Figure 7 (B). The boxes lost in treatment 1 were most likely due to the high workload imposed by the minimal amount of autonomy. However, treatment 4 does not have the obvious workload challenges of treatment 1. In fact, it was clearly ranked as the least burdensome, so why would it have the highest occurrences of errors? We believe the high error rate is a side effect of the high opacity of the highest-autonomy condition.

Our third measure of error was the number of times a player was blocked while entering a room. This measure is indirect because it is possible that the most efficient act would be to wait outside a blocked door, but in general it indicates poor coordination. As shown in Figure 7 (C), the human player was blocked in treatment 4 much more often, indicating significantly more coordination breakdowns than any other treatment.

## Subjective Performance Assessment

### User Performance Assessment

We asked the subjects to identify which team they felt performed best. Treatment 3 was the clear winner, with 63% of the participants selecting it as the best performing treatment (Figure 8(A)). Only 17% of the subjects choose treatment 4 as the best performing.

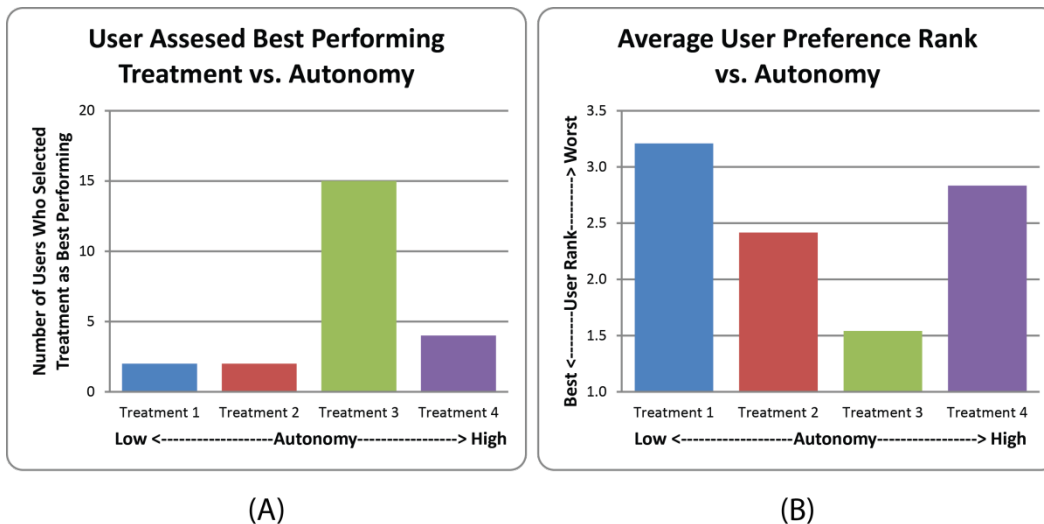


Figure 8 (A) User Assessment of Performance vs. Autonomy (B) User Preference vs. Autonomy

### User Treatment Preference

Human acceptance is an important component of overall system performance in tasks like ours. We asked the participants to rank the agents in each experimental condition with respect to their preference as to which one they would like to play with again, on a scale of 1 (most like to play with again) to 4 (least like to play with again).

Figure 8 (B) shows the results. Treatment 3 was preferred with statistical significance ( $\chi^2_{95,3} = 22.150$ ). This result also demonstrates the inflection point anticipated by the increasing opacity in the system from Figure 1 (C). We suspect this is because in treatment 3 the human holds the overall plan, most of the context, and exercises the greatest degree of creativity. In this context, transparency and control (directability) may be more important than autonomy (independent operation), especially in light of the particulars of the autonomous task.

We asked participants about the reasons for their rankings, and the responses were enlightening. Reasons for preferring Treatment 3 included:

- Shared information
- Able to anticipate
- Predictable
- Low burden
- Cleverest
- Automatic, but still have control

The first three reasons correlate with our predictions about opacity. The comment about low burden is interesting, because treatment 4 was objectively less burdensome. This comment suggests that there may be other types of burden besides the manual workload of tasking the agent. The comment about treatment 3 being cleverest is also interesting, because treatment 4 is objectively the most capable (clever) based on what the agent can do on its own (Figure 3). Perhaps this suggests that sometimes being more independent may not necessarily lead to being viewed as more clever. The final reason is also important because it relates to the broader issue. We focused on opacity in order to keep the experiment simple, but predictability, directability and other challenges in making automation a team player [7] are no doubt also affected by increased autonomy.

## Conclusions

The results of our initial limited evaluation support our claim that increasing autonomy does not always improve performance of the human-machine system. In the BW4T domain, this was principally due to opacity in the system, derived from increasing autonomy without accounting for the interdependence of the actions and decisions of the players and the coordination challenges this creates. Additionally, we showed how keeping an agent busy does not equate to improved performance, how human error rates are not only due to workload but can also be affected by opacity, and how user preference is not necessarily driven by reduced burden when other factors such as transparency, predictability and directability are relevant to the task. A key point to take away is that the ability to work *with* others becomes increasingly important as interdependence in the joint activity grows. It is possible that in complex and uncertain domains, this may be more valuable than the ability to work independently.

It is obvious why opacity has such an effect on the system in the BW4T domain. The greater the autonomy of players, the greater the opacity, and hence the more room for coordination breakdowns. The independent activity in treatment 4 inhibited the team's ability to engage in what most people

would consider “natural” coordination, resulting in a breakdown of common ground [14] and reduction in each player’s individual situation awareness. This then caused suboptimal decisions and errors. While obvious in this simple, abstract domain, the problem remains prevalent in many systems today, as noted by several researchers [4-6]. Understanding the relationship of autonomy to interdependence is one step toward addressing the challenges facing future systems. We believe that consideration for interdependence while designing the autonomous capabilities of an agent can mitigate the effects demonstrated and will enable future systems to achieve greater potential.

## References

1. *Unmanned Systems Roadmap, 2007-2032.*
2. Bleicher, A., *The Gulf Spill's Lessons for Robotics*, in *ieee spectrum special report* 2010. p. 9-11.
3. Johnson, M., et al., *The Fundamental Principle of Coactive Design: Interdependence Must Shape Autonomy*, in *Coordination, Organizations, Institutions, and Norms in Agent Systems VI*, M. De Vos, et al., Editors. 2011, Springer Berlin / Heidelberg. p. 172-191.
4. Stubbs, K., P. Hinds, and D. Wettergreen, *Autonomy and common ground in human-robot interaction: A field study*. IEEE Intelligent Systems, 2007(Special Issue on Interacting with Autonomy): p. 42-50.
5. Norman, D.A., *The "problem" of automation: Inappropriate feedback and interaction, not "over-automation"*, in *Human factors in hazardous situations*, D.E. Broadbent, A. Baddeley, and J.T. Reason, Editors. 1990, Oxford University Press. p. 585-593.
6. Woods, D.D. and N.B. Sarter, *Automation Surprises*, in *Handbook of Human Factors & Ergonomics*, G. Salvendy, Editor 1997, Wiley.
7. Klein, G., et al., *Ten Challenges for Making Automation a "Team Player" in Joint Human-Agent Activity*. IEEE Intelligent Systems, 2004. **19**(6): p. 91-95.
8. Johnson, M., et al., *Joint Activity Testbed: Blocks World for Teams (BW4T) in Engineering Societies in the Agents World X2009*.
9. Sheridan, T.B. and W. Verplank, *Human and Computer Control of Undersea Teleoperators* 1978, Man-Machine Systems Laboratory, Department of Mechanical Engineering, MIT: Cambridge, MA.
10. Olsen, D.R. and M. Goodrich. *"Metrics for Evaluating Human-Robot Interaction*. in *Proceedings of PERMIS*. 2003.
11. Parasuraman, R., T. Sheridan, and C. Wickens, *A model for types and levels of human interaction with automation*. Systems, Man and Cybernetics, Part A, IEEE Transactions on, 2000. **30**(3): p. 286-297.
12. Crandall, J.W. and M.L. Cummings, *Developing performance metrics for the supervisory control of multiple robots*, in *Proceedings of the ACM/IEEE international conference on Human-robot interaction* 2007, ACM: Arlington, Virginia, USA.
13. Johnson, M., et al., *Beyond Cooperative Robotics: The Central Role of Interdependence in Coactive Design*. IEEE Intelligent Systems, 2011. **26**: p. 81-88.
14. Klein, G., et al., *Common Ground and Coordination in Joint Activity*, in *Organizational Simulation*, K.R.B. William B. Rouse, Editor 2005. p. 139-184.