Improving Transparency in Human-Collective Visualizations

Joshua Bhagat Smith, Prakash Baskaran and Julie A. Adams, *Senior Member, IEEE*

***Abstract*— Quantifying transparency requires evaluating the transparency embedded in the various system design elements to determine how they impact one another and influence human-collective interactions. Prior work demonstrated limitations of an abstract collective interaction. Interface designs to address these limitations and improve human-collective interaction transparency were evaluated for a sequential best-of-N decision-making task with four collectives, each consisting of 200 individual entities. The Informed and Simple visualizations' predictive progress bars improved transparency and the overall human-collective team performance.**

I. INTRODUCTION

Biologically inspired discrete robotic collective decision-making (e.g., termites [1], and honeybees [2]) requires choosing the best option from of a set [3] and applies to several problem domains (e.g., [4], [5], [6]). Robotic collectives are more resilient [7] to individual agent failure and can scale to very large groups [8]. These promising behaviors, particularly in inhospitable or remote environments must respond to a human's influence to ensure their safe operation. The human-collective interface must leverage fundamental visualization principles that aid accurate information perception and comprehension to inform operator actions.

Transparency, the principle of providing easily exchangeable information to enhance comprehension [9], can support meaningful and insightful information exchanges between an operator and a collective. Enhancing transparency can mitigate poor operator behaviors and improve the system's overall performance, but achieving transparency is challenging as collectives become larger, tasks become more complex, and the amount of information provided increases. Providing too much transparency may overload the operator and negatively affect performance. Prior work demonstrated that an abstract interface can enable a human to effectively control four collectives of 200 agents each [10], and the interface provided better overall transparency than traditional agent-based visualizations [11]. However, three design limitations must be addressed to improve transparency: (1) provide perceivable collective state information, (2) provide predictive information about operator actions, and (3) provide the operator with the collective's predicted state [11].

Two abstract collective interfaces differentiated by the presented predictive information address these design changes. The *Simple* interface focuses on perceivable collective states and simple communications of the operators' potential actions. The *Informed* interface also incorporates a collective's predicted long-term behavior. The within-subjects evaluation focused on understanding the transparency improvements and the benefits of predictive information.

II. BACKGROUND

Robotic collective control is often inspired by social insects [12]–[14] (e.g. consensus Best-of-N decision algorithms that mimic the nest selection [2], [15]). The best-of-N algorithm models a collective as a colony, where agents randomly explore an environment to identify potential new nest (i.e., site) locations. Once a new site is identified, agents return to the hub and disseminate that information to others. Agents have one of four states: Uncommitted, Favoring, Committed, and Executing. Uncommitted agents do not prefer any site and are either exploring the environment, or communicating with other agents in the hub. The favoring agents prefer a particular site and are either assessing that site's value or communicating that site's information to others. Agents transition from the favoring state to the committed state when a sufficient number support the same site. The committed agent's notify the others of the decision. Agents transition from the committed state to the executing state once enough agents are notified of the decision, causing them to actively move to the new site [12].

The best-of-N algorithm can identify the locally best resource (e.g., [16], [17]), or shortest path (e.g., [18], [19]). Prior work demonstrated that a simple state machine algorithm enables selection of a higher valued site [14], [20]; however, larger collectives can make suboptimal decisions when the cost (e.g., distance) of the sites is asymmetric (e.g., a significant delay exists between site discoveries) [21], which is known as environmental bias. This bias can heavily impact the best-of-N algorithm's performance, as sites located closer to the hub are easier to discover and require less travel time to assess, which causes agents that favor closer, potentially lower valued, sites and to recruit and inhibit other agents more frequently. Cody et al.'s [10] state-machine algorithm introduced mechanisms to compensate for environmental bias: interaction delay and interaction frequency modulation, which enable selecting the highest valued site, even if it is the most distant.

Transparency seeks to provide an easy to use information exchange between a human operator and collectives to promote comprehension, intent, performance, and reasoning [11]. Many factors influence transparency directly (e.g., explainability, usability, performance, trust), along with many indirect factors (e.g., workload, situational awareness,

This effort was supported by an Office of Naval Research award N00014-21-1-2190. The contents are those of the authors and do not represent the official views of, nor an endorsement, by the Office of Naval Research.

The authors are with the Collaborative Robotics and Intelligent Systems Institute, Oregon State University, Corvallis, Oregon USA julie.a.adams@oregonstate.edu

quantity of information) [22]. Two of the most influential transparency factors are performance and usability [11]. Performance represents a human's ability to produce a desired output when executing a task [23]. Usability is a multifaceted quality that enables operators to achieve desired goals that can be anticipated easily and do not cause confusion [24]. Usability interacts with fifteen transparency factors, making it useful for assessing and implementing transparency within visualizations [22]. More transparent interfaces visualize information (e.g., future collective states) in a cohesive manner that can help an operator draw conclusions [25] and justify actions faster [26], with lower levels of effort. Prior work evaluated a traditional (i.e., individual-agent-based) interface and an abstract interface for a sequential best-of-N decision-making task with four collectives of 200 agents each [11]. Visualization transparency was measured with respect to how the interface impacted the operator's comprehension, the interface's usability, and human-collective's performance. The abstract interface was more transparent, as it enabled higher performance and understanding [11].

The abstract interface's transparency can be improved by addressing usability related concerns. The abstract interface's operators relied heavily on supplementary pop-up windows to access information required to influence the collective. These pop-ups increased the visual clutter, and impacted negatively operators' ability to perceive and comprehend information [22]. The pop-up windows' information resulted in executing redundant commands, indicating a lack of understanding collectives' behavior. Design recommendations were provided, but were not evaluated [11].

III. Methods

Three major recommendations for addressing the limitations of the collective best-of-N site selection transparency results [11] were to provide: perceivable collective state information, predicted outcomes of operators' actions, and projected collective states [22]. Four design changes to the (Prior) abstract collective interface [10] were made to address the first two recommendations. An updated hub icon, an abandoned site overlay, and a new executed commands display were developed. New information provides the expected impact of the operator's commands on the collective's decision-making process. These changes constitute the *Simple* interface. Information modifications that indicate a collective's predicted long-term behavior in response to a specific command address the third design recommendation, resulting in the *Informed* interface.

A. Interface Design

The Prior abstract collective interface permitted displaying additional information via a pop-up window for a given entity (i.e., hub or site) using a right-click [10], (see Figure 1), which created significant visual clutter [11]. Visual clutter can increase operator frustration [27] and decrease usability [11]. The redesigned hub icons retain the Roman numeral at the top identifying the hum (i.e., I, II, III, IV) and incorporate information directly about the collective's current state.

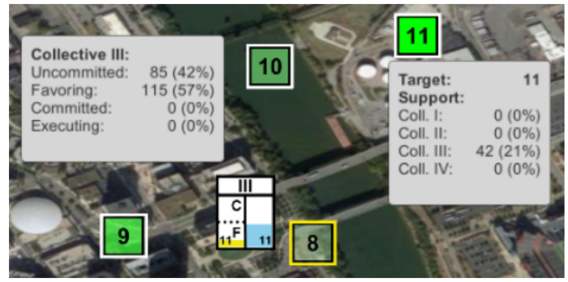
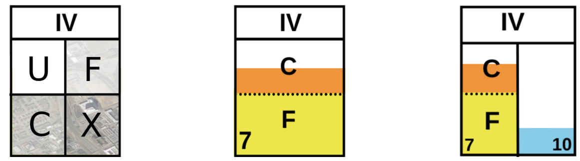


Fig. 1: Example pop-up information for a hub and site.



(a) Abstract (b) Simple (c) Informed

Fig. 2: Hub icon comparisons: (a) Prior abstract hub [28], (b) Simple interface hub, and (c) Informed interface hub.

The new hub icons indicate the collective's most favored site and its progression towards a decision via a vertical bar, shown in Figures 2b and 2c (yellow and orange). The Prior icon, shown in Figure 2a, represented the relative progress for each state (i.e., Uncommitted, Favoring, Committed, and Executing) using opacity, with the associated percentages provided via the Collective's hub pop-up window. The Simple hub's icon (Figure 2b) is like a progress bar, with the yellow portion representing the percentage of agents favoring the most supported site, whose ID is in the bottom left corner. As agents begin to favor the same site, the yellow bar approaches the dotted line representing the 50% decision-making threshold (100 of 200 agents), at which point agents begin transitioning to the committed state. The top, orange bar increases as the number of committed agents approaches the 50% threshold, at which point these agents transition to the execute state and begin moving to the selected site. The Informed interface, shown in Figure 2c, also includes predictive information. The bottom right number is the collective's perceived highest valued site, and the associated blue bar represents the likelihood the collective will choose that site, where a larger bar equals higher likelihood.

The Prior interface permitted three commands, each required selecting a hub and a site. The *Investigate* command randomly selected 20 agents in the hub to investigate a selected site. The *Abandon* command removed the selected site from consideration. The *Decide* command, only available once 30% of the selected collective supported the same site, forced an immediate commitment to that site. The operator was only permitted to cancel an Abandon command request.

The Prior evaluation's operators issued repeated abandon commands for the same site and often abandoned the highest value site, resulting in negative outcomes [11]. Design changes were implemented to enhance understanding and reduce the unnecessary abandon command frequency. A red

X was placed over the entire abandoned site, see Figure 3, rather than the Prior interface's simple red outline [10].

Fig. 3: The Abandoned site red **X** overlay.

A design change indicated how a command will impact the collective's state over time (e.g., $>$ 1 minute). The Simple interface's expected impact is a visual symbol: check mark (net positive), dash (net neutral), and not symbol (net negative) (see Figure 4a). The Informed interface shows this information by indicating a command's affect via directional arrows associated with the hub icon (see Figure 4b). The left arrow indicates an increase or decrease in the number of agents favoring the displayed site, and the right arrow indicates a change in the likelihood of choosing the indicated highest valued site. The magnitude of the impact is presented via a lightly shaded region above the existing progress bar.

A simple heuristic predicts a commands' impact and the collective's likelihood of choosing the highest valued site based on the collective's current state (i.e., agents supporting each site), and the command's affect on that state. For instance, abandoning the lowest-valued site reduces the number of sites and can increase the number of agents supporting the remaining sites. The Simple interface's naive algorithm uses the highest and lowest value sites to determine whether a command will have a positive, neutral, or negative impact on the collective (see Figure 4a).

The Informed interface's algorithm calculates a command's impact on the collective as well as its affect on the probability of selecting the highest valued target (i.e., blue bar) and the most favored target (i.e., yellow/orange bars). An investigate command's expected outcome is that 20 agents will support that site. If the selected site is the highest valued site, the blue bar increases. If the selected site is the most favored site, the yellow bar increases. Abandon commands force all agents currently favoring a site into the uncommitted interactive state and those agents' support is distributed to the other sites proportional to each site's current support level. This interface displays the maximum support for the most favored site (i.e., yellow bar) if the selected site is the most favored. Likewise, the maximum probability of selecting the highest valued site (i.e., blue bar) is visible if the highest valued site is selected. If the selected site is not the most favored or highest valued, zero is displayed for the most favored site or zero is displayed for the probability of selecting the highest valued site, respectively.

B. Experimental Design

A single operator supervised four robotic collectives of 200 agents each that performed the best-of-N decision-making task. A single decision-making task required a collective to explore its environment and select the highest valued site within a 500-meter range.

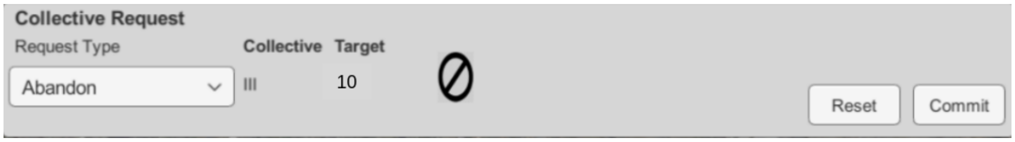
The within-subjects independent variables were the Simple and Informed interfaces and problem difficulty. Trials with the highest valued sites close to a collective's hub were easy, while hard trials placed the highest valued sites far away. The interface conditions were randomized, with problem difficulty randomized within the interface condition. A trial ended when each collective solved two best-of-N problems, or when a trial was over 10-minutes with six decisions.

The primary hypothesis was that the redesigned hub visualizations, combined with the predictive information, improve transparency. Three hypotheses investigated how the visualizations affected the performance, understanding and usability transparency factors [22]. Objective data and the operators' actions were recorded via the interface.

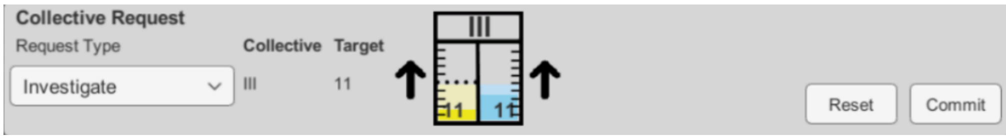
Hypothesis $\mathbf{H_1}$ predicted the redesigned interfaces allow operators to efficiently and positively influence collectives' long-term behavior. $\mathbf{H_2}$ stated the redesigned visualizations improve the operator's understanding of the collectives over the Prior interface ($\mathbf{H_2^a}$), with the Informed interface providing the best understanding ($\mathbf{H_2^b}$). $\mathbf{H_3}$ predicted both redesigned interfaces' usability will be better than the Prior interface ($\mathbf{H_3^a}$), with the Informed interface having the best usability ($\mathbf{H_3^b}$). ($\mathbf{H_4}$) predicted the redesigned visualizations enable improved operator performance ($\mathbf{H_4^a}$), and the Informed interface improves performance the most ($\mathbf{H_4^b}$).

Selection accuracy is the ratio of correct decisions (i.e., selecting the highest valued site) relative to the total decisions. *Decision time* is the total time (minutes) required to reach a decision. The *Selected site value* is the average site value across all selected sites in a trial. The *highest valued site abandoned (%)* is the percentage of times the highest valued site was abandoned relative to the total number of sites abandoned during a trial. The *cancel abandon command* is the number of times an operator canceled the issued abandon commands. The *abandoned command exceeds* metric represents the difference between the number of abandoned commands issued and the number of sites actually abandoned. The *decide success rate (%)* represents selecting the highest valued sites when issuing the decide command. The collective and site left-clicks are the number of collective and site selections, respectively for issuing commands, while the corresponding right-clicks are the number collectives and sites selected to view the pop-up windows.

Twelve *Situational Awareness* (SA) probes were asked verbally once per minute for the first six minutes of each trial to assess the operator's perception (SA_1), comprehension (SA_2), and projection capabilities (SA_3) [29] (i.e., four per SA level). The *SA probe accuracy* was the percentage of correct responses. Overall SA (SA_O) represents the percent of the correctly answered SA probes out of twelve. The Prior evaluation's SA probe questions were unevenly distributed across the three SA levels (i.e., 5 SA_1, 4 SA_2, and 3 SA_3 probes). The current evaluation distributed the probes evenly across all levels. Additionally, some SA_2 and SA_3 questions were altered to evaluate interface design changes, which



(a) Simple interface with the negative impact icon.



(b) The Informed interface's directional arrows indicating the level of support and selection of highest value site changes.

Fig. 4: Visual representation of the potential command's impacts on collective outcomes.

made some SA probe questions more difficult. For instance, Yes or No questions like "Is Site **16** likely to be selected by Swarm **III**?" were replaced by "Which Swarm can select the optimal site faster by investigating Site **16**?".

Age, education, and graphical user interface proficiency was gathered from twenty-five participants (7 females, 15 males, and 2 non-binary). The mean age was 36.75 (standard deviation-std. = 14.71). Six had a high school degree, eight held an undergraduate degree, nine master's degree, and one a doctorate. The majority (20) indicated they use computers more than eight hours per week. Participants video game skill level was 4.33 (std. = 2.10) on a Likert scale (1-little to 9-expert). Participant's correctly answered 11.33 (std. = 5.95) of the 24 mental rotation assessment questions [30].

IV. Results

The Wilcoxon signed-rank test was used to compare the interfaces. Non-parametric statistics ensure that the outcomes were unaffected by the distribution across operators. Bolded results in the tables represent the best results.

The Informed interface had the highest overall selection accuracy (see Table I), due increased hard problem accuracies; however, it performed worse for easy problems. The Simple interface performed the best for the easy problems. No significant differences existed. The Prior interface's selection accuracies across all difficulty levels were comparable to the Simple interface results, but with higher variability.

The objective was to select the highest valued site for each decision. The Informed interface's selected site values were consistently higher than the Simple interface, regardless of the decision difficulty. Significant effects ($p < 0.01$) existed for the hard decision problems, but no other differences existed. The Prior interface's selected site values were the lowest for all problem difficulties.

The Informed interface's decision times, while similar with no significant differences to the Simple interface, were faster for the easy problems and slightly slower for hard problems. The Prior interface had the fastest overall decision times, but with substantial disparity across difficulties. This interface was the fastest for easy problems and the slowest for the hard problems. The Informed interface had lower selection accuracy with faster decisions for easy problems, but higher accuracy with longer decision times for hard problems.

TABLE I: Mean (std.) performance and understandability metrics by interface (UI) and difficulty. The cancel abandon command metric is not available by difficulty for the Prior interface. Selection Accuracy, Highest Valued Site Abandoned and Abandoned Command Exceeded are percentages.

Metric	UI	Difficulty		
		Easy	Hard	Overall
Selection Accuracy	Inform	91.8 (12.9)	**87.5 (12.4)**	**89.7 (12.4)**
	Simple	**96.9 (7.0)**	81.0 (17.0)	89.0 (15.3)
	Prior	94.4 (23.0)	81.4 (39.0)	88.4 (32.1)
Decision Time	Inform	4.2 (1.3)	4.5 (1.4)	4.4 (1.4)
	Simple	4.3 (1.5)	**4.5 (1.1)**	4.4 (1.3)
	Prior	**3.4 (1.2)**	4.7 (1.2)	**4.0 (1.4)**
Selected Site Value	Inform	**94.9 (7.3)**	**97.0 (4.2)**	**96.0 (6.0)**
	Simple	94.8 (7.8)	95.9 (5.2)	95.4 (6.6)
	Prior	92.1 (5.5)	92.0 (4.5)	92.1 (5.1)
Highest Site Abandon	Inform	12.0 (18.1)	14.3 (18.3)	13.2 (18.2)
	Simple	**7.3 (17.1)**	**6.5 (11.7)**	**6.9 (14.7)**
	Prior	33.3 (36.0)	48.7 (36.9)	43.6 (31.9)
Abandon Command Exceeded	Inform	**0.0 (0.0)**	0.2 (0.7)	0.1 (0.5)
	Simple	**0.0 (0.0)**	**0.0 (0.0)**	**0.0 (0.0)**
	Prior	2.1 (5.1)	3.1 (7.7)	2.7 (6.3)
Cancel Abandon Command	Inform	**0.0 (0.0)**	**0.0 (0.2)**	**0.02 (0.2)**
	Simple	0.1 (0.3)	**0.0 (0.2)**	0.1 (0.3)
	Prior	-	-	0.4 (1.0)

The redesigned interfaces resulted in approximately twice as many commands per decision as compared to the Prior interface (see Table II), indicating that operators were more involved in the decision-making process. Abandoning lower valued sites can improve the selection accuracy and decision times, particularly for the hard problems. Table II also shows operators abandoned more sites with the redesigned interfaces. Generally, the Informed interface had more commands issued than the Simple interface across all difficulty levels, with no significant differences. More commands were issued for the hard problems across all interfaces.

The Informed interface operators abandoned the highest valued site more frequently than the Simple interface for all difficulty levels (Table I), which was significant ($p = 0.04$) for hard problems. Both redesigned interfaces performed

TABLE II: The number of commands issued descriptive statistics by command type and interface. Decisions for which the human provided no commands are excluded.

	UI	Difficulty		
		Easy	Hard	Overall
Investigate Command	Informed	**3.2 (2.8)**	**4.0 (3.0)**	**3.5 (2.7)**
	Simple	2.6 (2.5)	3.2 (2.3)	3.0 (2.2)
	Prior	1.7 (1.1)	2.9 (1.3)	1.9 (1.1)
Abandon Command	Informed	**1.3 (1.1)**	**1.2 (0.8)**	**1.3 (1.0)**
	Simple	0.8 (1.2)	1.0 (1.2)	1.0 (1.2)
	Prior	0.1 (0.2)	0.2 (0.2)	0.1 (0.1)
Decide Command	Informed	**0.7 (0.3)**	**0.7 (0.3)**	**0.7 (0.3)**
	Simple	0.6 (0.4)	**0.7 (0.3)**	**0.7 (0.3)**
	Prior	0.6 (0.4)	0.5 (0.3)	0.5 (0.3)
Total Commands	Informed	**5.1 (3.5)**	**5.7 (3.0)**	**5.4 (3.1)**
	Simple	4.0 (3.6)	4.9 (3.2)	4.7 (3.1)
	Prior	2.5 (1.1)	2.8 (1.3)	2.6 (1.1)

substantially better than the Prior interface that had a much higher (3-6 times) highest valued site abandoned rate.

The operator may accidentally issue the abandoned command repeatedly for the same site; hence, the percentage of times *abandon commands exceeded abandoned sites* was examined (see Table I). The Simple interface had no instances of abandoning the same site multiple times, while the Informed interface had $< 1\%$ on average. No significant differences existed, but both interfaces reduced the excessive abandon commands compared to the Prior interface.

TABLE III: The Operator Influenced Decisions (%) and Decision Success descriptive statistics by interface.

	UI	Difficulty		
		Easy	Hard	Overall
Investigate Influenced Decision	Informed	85.5 (30.8)	**99.0 (3.9)**	88.0 (25.2)
	Simple	78.1 (40.5)	94.1 (23.4)	83.5 (27.9)
	Prior	**89.9 (17.6)**	92.9 (15.1)	**91.4 (14.3)**
Decision Success	Informed	95.1 (8.2)	93.5 (7.8)	93.7 (8.2)
	Simple	**98.8 (3.1)**	91.9 (10.3)	94.7 (9.0)
	Prior	98.6 (3.2)	**94.8 (7.6)**	**96.9 (4.2)**

The *cancel abandon command* nullifies an issued abandon command for a particular collective and site. The Informed interface generally had fewer cancel abandon commands compared to the Simple interface across all difficulty levels; however, no significant effects existed. Overall, operators using the Prior interface canceled the abandoned command more frequently than the redesigned interfaces.

The redesigned interfaces' operators influenced (issued $\geq$ 1 command) fewer easy and overall decisions, but influenced more hard decisions than the Prior interface (see Table III). This observation was especially true for the Informed interface, where operators influenced over 99% of the decisions with less variability. No significant differences were found.

Decide success was 100% when the operator decided the highest valued site, 66% when deciding the second highest, 33% for the third highest, and 0% for the lowest (shown in Table III). Overall, the Prior interface had the best decide success rate. Among the redesigned interfaces, Simple had higher decide success overall and for easy problems, while the Informed interface's decide success was higher for hard problems. No significant differences existed. While the Informed interface had a slightly lower overall decide success rate, it had the best selection accuracy overall and for hard decisions, with the highest overall selected site values.

TABLE IV: Collective and site left- and right-clicks mean (std.) by interface and difficulty.

Clicks	UI	Difficulty		
		Easy	Hard	Overall
Collective Left	Informed	78.8 (32.8)	83.3 (22.5)	81.0 (28.2)
	Simple	**64.6 (43.4)**	**68.5 (36.8)**	**66.6 (40.3)**
	Prior	-	-	122.0 (47.4)
Collective Right	Informed	**26.5 (22.6)**	**28.5 (21.4)**	**27.5 (22.0)**
	Simple	30.3 (40.7)	45.0 (53.1)	37.7 (47.9)
	Prior	-	-	30.6 (32.0)
Site Left	Informed	70.6 (26.3)	72.2 (21.3)	71.4 (23.9)
	Simple	**68.5 (31.1)**	**65.0 (30.6)**	**66.7 (30.9)**
	Prior	-	-	185.6 (64.3)
Site Right	Informed	**10.9 (10.5)**	**11.5 (15.6)**	**11.2 (13.3)**
	Simple	13.1 (16.8)	11.9 (11.0)	12.5 (14.2)
	Prior	-	-	82.4 (60.2)

Left- and right-clicks were examined for both the sites, and the collectives' hubs. The collective and site clicks descriptive statistics are provided in Table IV. The Simple interface had fewer left-clicks, but more right-clicks across all problem difficulties. The results across difficulty levels were significant ($p = 0.04$) for the collective left-clicks between the redesigned interfaces. These differences indicate that design changes (e.g., the blue likelihood bar and the expected impact directional arrows) permitted examining a command's expected impact on the hub without relying on the supplemental pop-up window information (figure 1. The number of clicks were not analyzed by difficulty for the Prior study, but the overall number of left- and right-clicks were 1.5 to 7.3 times higher than the redesigned interfaces.

TABLE V: SA probe accuracy (%) mean (std.) by interface and difficulty aggregated across operators.

Level	UI	Difficulty		
		Easy	Hard	Overall
Overall SA	Informed	65.8 (24.2)	62.0 (19.2)	64.4 (16.6)
	Simple	66.7 (22.0)	60.9 (15.2)	63.8 (15.3)
	Prior	**91.9 (11.6)**	**87.6 (15.4)**	**89.9 (13.7)**
SA Level 1	Informed	67.4 (31.7)	73.9 (32.5)	71.4 (22.5)
	Simple	78.3 (24.8)	56.5 (37.0)	67.4 (24.9)
	Prior	**93.4 (14.8)**	**89.6 (17.9)**	**91.7 (16.5)**
SA Level 2	Informed	65.2 (37.4)	52.2 (37.5)	58.0 (29.0)
	Simple	65.2 (31.0)	63.0 (36.8)	64.1 (28.4)
	Prior	**88.2 (20.4)**	**88.0 (19.9)**	**88.1 (20.3)**
SA Level 3	Informed	65.2 (37.4)	60.9 (36/0)	63.8 (26.3)
	Simple	56.5 (39.9)	63.0 (30.3)	59.8 (29.3)
	Prior	**95.3 (20.9)**	**84.3 (35.4)**	**90.2 (29.0)**

Overall, SA probe accuracies (Table V) were similar for the redesigned interfaces with no significant differences.

The Simple interface's SA_1 accuracy was 11% higher than the Informed interface for the easy problems, while the corresponding SA_1 accuracy for hard problems was 17% higher. The Informed interface's SA_3 accuracy for the easy problem was 9% higher than the Simple interface. The redesigned interfaces consistently led to lower SA accuracy (up to 30%) than the Prior interface across all SA levels.

V. Discussion

Hypothesis $\mathbf{H_1}$ predicted that the redesigned interfaces support operators' efficient and positive influence on the colony's long-term behavior. Operators with the redesigned interfaces better understood the collectives and how to influence the algorithm's outcome, even though the decision times and SA results did not improve. Each hub always chose from four sites, which was harder than the Prior interface's evaluation. The Prior evaluation removed sites once they were selected, which is known to a easier decision problem.

Hypothesis $\mathbf{H_2^a}$ predicted that the redesigned interfaces provide a better understanding of the collectives, which was supported. Operators were expected to abandon lower valued sites to expedite the decision-making process. However, operator's with the Prior interface abandoned the highest valued sites more frequently, leading to more frequently canceling the abandon commands. The redesigned interfaces' ability to predict the results of a command mitigated this issue, enabling operators to better understand how their actions may affect decision-making. However, the SA probe accuracies for the redesigned interfaces were lower, on average.

Overall, the collectives faced more difficult decision problems. The Prior interface's evaluation had exactly sixteen sites (i.e., four per collective), and the number of sites decreased after collectives made their first decisions resulting in fewer sites for the second decisions, where the differences between these sites' values (i.e., decision difficulty) was not enforced. The current evaluation required collectives to choose between at least four sites and maintained the problem difficulty for all decisions. These differences made the decision problems harder and made it harder for operators to attain good SA. As well, the SA probe questions were substantially more difficult than the Prior evaluation, which also negatively influenced SA probe accuracy.

Hypothesis $\mathbf{H_2^b}$ predicted the Informed interface will improve understandability over the Simple interface, which was not supported. The Informed interface had a higher number of cancel abandon commands and the highest valued site abandoned percentage. The Informed interface's directional arrows may have been confusing, while the Simple interface's discrete symbol presentation (i.e., checkmark, dash, and not symbol) was easier to perceive and comprehend.

Hypothesis $\mathbf{H_3^a}$ predicted that the redesigned visualizations will enhance usability, which was supported. Operators of the Prior interface repeatedly issued additional abandoned commands for sites they had already abandoned, meaning these additional commands had no additional impact on the hub's decision-making process. Overlaying a $\mathbf{X}$ on the site in addition to the red outline, and separating the request monitoring area into two panes reduced the *abandon command exceeds* drastically, thereby enhancing usability. The new Hub icon design and the expected command impact design changes also enabled the operators to rely less on the supplemental pop-up window information.

Hypothesis $\mathbf{H_3^b}$ predicted that the Informed interface will enhance the usability significantly over the Simple interface. This hypothesis was only partially supported, as the Informed interface results indicated that operators better understood how the issued commands affected the decision-making, and aided in answering the SA probe questions without relying on the pop-up windows, which resulted in less visual clutter.

Hypothesis $\mathbf{H_4^a}$ predicted that operators using the redesigned interfaces will have higher performance than those using the Prior interface, which was partially supported. The redesigned interfaces' operators more accurately selected similar or higher valued sites with less variability; however, the Prior interface's decision times were faster due to the decision problems becoming easier as the trials progressed.

Favoring and *Committed* are the most important states informing how close the collective is to a decision. Displaying these states' progress bar, and the expected impact of potential commands encouraged operators to issue positive decision-making commands (e.g., investigating highest valued and abandoning lower-valued sites) more frequently, and prevented issuing commands with a negative impact. The redesigned interface's improved performance over the Prior interface. The redesigned interfaces' slower decision times can be attributed to the increased problem difficulty, while the Prior interface's evaluation resulted in some hubs selecting from three sites or easier decision problems as trials progressed. Thus, the redesigned interfaces allowed the operators to efficiently and positively influence collective decision-making, even with more challenging problems.

Hypothesis $\mathbf{H_4^b}$ predicted that the Informed interface will result in higher performance than the Simple interface, which was not supported. The Informed interface had lower selection accuracy with faster decision times for easy problems, higher selection accuracy with longer decision times for hard problems, and overall slightly better selected site values. It appears the operators preferred the simpler interface's discrete symbol information for easy problems, while the directional arrows with the predictive blue bar were beneficial for harder problems. Generally, longer decision times led to a higher selection accuracy, as devoting more time to ensure high task performance is a commonly observed trade-off.

The redesigned interfaces provided better transparency, as the human-collective team performed better with high understandability and usability. Presenting the most favored site and the collective's progress throughout the decision-making process via the progress bars enabled understanding the collective's state and helped identify necessary interventions. The reliance on collective and site information pop-up windows was common with the Prior interface. An ideal visualization must provide necessary supplementary information, but require minimal reliance on on the pop-up windows. The progress bar representation reduced the operators' reliance on pop-up windows, and improved the collective decision-making performance.

Providing the expected impact information incentivized operators to take positive actions more frequently and avoid negative actions, leading to better operator comprehension and performance. Both the Informed and Simple interfaces had similar transparency, which may be attributed to using the same heuristic-based algorithm for deriving the expected impact information. A collective is a highly probabilistic system with complex inter-agent dynamics; therefore, predicting future behaviors using only the current state's heuristics may have been inaccurate and unuseful.

VI. Conclusion

Quantifying transparency requires evaluating the transparency embedded in the design elements to determine their impact on one another and their influence on human-collective interactions. Visualizations with Informed and Simple predictive information were compared to a Prior abstract collective visualization for a best-of-N decision task with four collectives, each consisting of 200 individual collective entities. The Prior abstract hub visualization's opacity challenged operators' ability to perceive and comprehend information, as some key information was hidden. The redesigned visualizations' progress bar improved transparency and the overall human-collective team performance. Providing the operator with information related to a command's expected impact had a net positive transparency outcome, which supports visualizations that incorporate projections into human-collective interface designs.

References

[1] E. Bonabeau, M. Dorigo, G. Theraulaz, and G. Theraulaz, *Swarm Intelligence: From Natural to Artificial Systems*. Oxford university press, 1999, no. 1.

[2] T. D. Seeley, *Honeybee Democracy*. Princeton, NJ: Princeton University Press, 2010.

[3] G. Valentini, *Achieving Consensus in Robot Swarms: Design and Analysis of Strategies for the Best-of-N Problem, Studies in Computational Intelligence Series, 706*. Springer, 2017.

[4] G. Coppin and F. Legras, "Autonomy spectrum and performance perception issues in swarm supervisory control," *Proceedings of the IEEE*, vol. 100, no. 3, pp. 590–603, 2011.

[5] M. Deijfen and M. Lindholm, "Growing networks with preferential deletion and addition of edges," *Physica A: Statistical Mechanics and its Applications*, vol. 388, no. 19, pp. 4297–4303, 2009.

[6] D. J. Sumpter, *Collective Animal Behavior*. Princeton University Press, 2010.

[7] J. Leaf, J. A. Adams, M. Scheutz, and M. A. Goodrich, "Resilience for goal-based agents: Formalism, metrics, and case studies," *IEEE Access*, vol. 11, pp. 121 999–122 015, 2023.

[8] E. Şahin, "Swarm robotics: From sources of inspiration to domains of application," in *International Workshop on Swarm Robotics*. Springer, 2004, pp. 10–20.

[9] K. A. Roundtree, M. A. Goodrich, and J. A. Adams, "Transparency: Transitioning from human–machine systems to human-swarm systems," *Journal of Cognitive Engineering and Decision Making*, vol. 13, no. 3, pp. 171–195, 2019.

[10] J. R. Cody, K. A. Roundtree, and J. A. Adams, "Human-collective collaborative target selection," *ACM Transactions on Human-Robot Interaction*, vol. 10, no. 2, pp. 1–29, 2021.

[11] K. A. Roundtree, J. R. Cody, J. Leaf, H. O. Demirel, and J. A. Adams, "Human-collective visualization transparency," *Swarm Intelligence*, vol. 15, no. 3, pp. 237–286, 2021.

[12] J. R. Cody and J. A. Adams, "An evaluation of quorum sensing mechanisms in collective value-sensitive site selection," in *International Symposium on Multi-Robot and Multi-Agent Systems*, 2017, pp. 40–47.

[13] S. C. Pratt and D. J. Sumpter, "A tunable algorithm for collective decision-making," *National Academy of Sciences*, vol. 103, no. 43, pp. 15 906–15 910, 2006.

[14] A. Reina, G. Valentini, C. Fernández-Oto, M. Dorigo, and V. Trianni, "A design pattern for decentralised decision making," *PLoS One*, vol. 10, no. 10, p. e0140950, 2015.

[15] N. R. Franks, K. A. Hardcastle, S. Collins, F. D. Smith, K. M. Sullivan, E. J. Robinson, and A. B. Sendova-Franks, "Can ant colonies choose a far-and-away better nest over an in-the-way poor one?" *Animal Behaviour*, vol. 76, no. 2, pp. 323–334, 2008.

[16] A. Campo, S. Garnier, O. Dédriche, M. Zekkri, and M. Dorigo, "Self-organized discrimination of resources," *PLoS One*, vol. 6, no. 5, p. e19888, 2011.

[17] M. S. Talamali, J. A. Marshall, T. Bose, and A. Reina, "Improving collective decision accuracy via time-varying cross-inhibition," in *IEEE International Conference on Robotics and Automation*, 2019, pp. 9652–9659.

[18] A. Reina, M. Dorigo, and V. Trianni, "Towards a cognitive design pattern for collective decision-making," in *International Conference on Swarm Intelligence*. Springer, 2014, pp. 194–205.

[19] A. Scheidler, A. Brutschy, E. Ferrante, and M. Dorigo, "The k-unanimity rule for self-organized decision-making in swarms of robots," *IEEE Transactions on Cybernetics*, vol. 46, no. 5, pp. 1175–1188, 2015.

[20] D. Pais, P. M. Hogan, T. Schlegel, N. R. Franks, N. E. Leonard, and J. A. Marshall, "A mechanism for value-sensitive decision-making," *PLoS One*, vol. 8, no. 9, p. e73216, 2013.

[21] T. Laomettachit, T. Termsaithong, A. Sae-Tang, and O. Duangphakdee, "Decision-making in honeybee swarms based on quality and distance information of candidate nest sites," *Journal of Theoretical Biology*, vol. 364, pp. 21–30, 2015.

[22] K. A. Roundtree, "Achieving transparency in human-collective systems," Ph.D. dissertation, School of Mechanical, Industrial and Manufacturing Engineering, Oregon State University, 2020.

[23] G. M. Alarcon, R. Gamble, S. A. Jessup, C. Walter, T. J. Ryan, D. W. Wood, and C. S. Calhoun, "Application of the heuristic-systematic model to computer code trustworthiness: The influence of reputation and transparency," *Cogent Psychology*, vol. 4, no. 1, p. 1389640, 2017.

[24] J. Rubin and D. Chisnell, "How to plan, design, and conduct effective tests," *Handbook of Usability Testing*, vol. 17, no. 2, p. 348, 2008.

[25] M. John, H. S. Smallman, and D. I. Manes, "Recovery from interruptions to a dynamic monitoring task: The beguiling utility of instant replay," in *Human Factors and Ergonomics Society Annual Meeting*, vol. 49, no. 3, 2005, pp. 473–477.

[26] M. Fox, D. Long, and D. Magazzeni, "Explainable planning," in *Explainable AI Workshop, International Joint Conference on Artificial Intelligence*, 2017.

[27] J. Preece, H. Sharp, and Y. Rogers, *Interaction Design: Beyond Human-computer Interaction*. John Wiley & Sons, 2015.

[28] J. R. Cody, "Discrete consensus decisions in human-collective teams," Ph.D. dissertation, School of Computer Science, Vanderbilt University, 2018.

[29] M. R. Endsley, "Toward a theory of situation awareness in dynamic systems," *Human Factors*, vol. 37, no. 1, pp. 32–64, 1995.

[30] S. G. Vandenberg and A. R. Kuse, "Mental rotations, a group test of three-dimensional spatial visualization," *Perceptual and Motor Skills*, vol. 47, no. 2, pp. 599–604, 1978.